



## Recombined and Hybridized Speaker Recognition System (SRS)

LAMKADAM ABDELMAJID<sup>1</sup>, KARIM MOHAMED<sup>2</sup>

<sup>1</sup>LISTA, Department of Computer Science and Mathematics, FSDM, USMB University,  
Fez, Morocco, majidxy@yahoo.fr

<sup>2</sup>LISTA, Department of Physics, FSDM, USMB University,  
Fez, Morocco, karim\_lessi@yahoo.fr

### ABSTRACT

We started this report with a description of the organs constituting the speaker recognition system (SRS), so we immediately went on to explore the types of speaker recognition and the difficulties cited at this level. Next, we carried out a review of the main approaches to combinations and hybridizations developed over the past ten years in the field of the automatic speech recognition (ASR), by carrying out an experimental and comparative study of such an approach.

Thus, we summarized the advantages and disadvantages presented by these methods of extraction and classification of acoustic characteristics, and their contributions to ASR. We have developed an approach to estimate a relevant speaker verification system, using a combined extractor, and a hybrid classifier, which allowed us to build our system of recognition.

**Key words:** Speaker Recognition, ASR, LPC, MFCC, PLP, RASTA, DTW, HMM, ANN, RPF, Recombination, hybridization, Threshold.

### 1. INTRODUCTION

Among the applications of voice technology, we find the system [22] of recognition or verification of the speaker [20], [26], which multiple fundamental concepts coincide, namely:

The signal processing which analyzes the sound signal for a better interpretation; thus the different extractors [4] of the acoustic characteristics which extract the maximum information from this signal; and the speech recognition which uses a set of algorithms which allow the classification [25] of the speech corpus.

To build a speaker verification system [21], [26] progress in various areas is used. So to arrive at an ideal and complete speaker verification system [30], we have:

- Optimized the learning parameters of the test corpus,
- Combined the extractors: LPC-MFCC-PLP-RASTA,
- Hybridized the classifiers: DTW / HMM / ANN.

This shows that the main objective to be achieved through this work is to carry out a speaker verification system, which allows us access to information in an efficient and optimal manner, based on the voice biometric feature "voice", as an authentication password.

### 2. AUTOMATIC SPEAKER RECOGNITION

#### 2.1. Presentation

Automatic Speaker Recognition (ASR) [21], [12] is a technology that identifies a person from their voice, after capturing a digital model of a person's voice, to determine the unique way of talk about a human being. The captured voice is transformed into a unique model which is stored in a database.

The applications [22] are applied to secure access to data, or to protected sites, and uses in audio indexing or in forensic expertise, or even are applied in telephone surveillance.

#### 2.2. Types of speaker recognition

In recognition of the speaker [20], [21] we make the difference between identification and verification [23] of the speaker, depending on whether the problem is to verify that the analyzed voice corresponds to the person who is supposed to produce it, or that it is a question of determining who, among a predetermined number of speakers, produced the analyzed signal.

We thus find recognition of the speaker [24]:

- Dependent on the text: the sentence to be recognized is fixed when the system is designed.
- With dictated text: the sentence to be pronounced is fixed during the test.
- Independent of the text: the sentence to be pronounced is not specified.

We also distinguish speech recognition [4], [22]:

- Mono speaker: the system is trained to recognize a person's voice.
- Multi speaker: the system is trained by a finite group of people.
- Speaker independent: the system can recognize anyone.

Finally, we cite recognition:

- Isolated words: the speaker separates each word with silence.
- Connected words: the speaker continuously speaks a series of predefined words.
- Continuous speech: the speaker pronounces any sequence of words continuously.

### 2.3. Developments

Advanced and recent studies are described in [28], [29], have shown encouraging and considerable results.

Also, several methods [14], [25] applying statistical and connectionist approaches are used regarding speaker identification and verification [30], such as statistical methods, and connectionist methods which are improved and applied to the speaker recognition, and that have resulted in satisfactory and convincing results.

## 3. FUNCTIONING OF THE SYSTEM

### 3.1. Principle

Speaker recognition [21] or verification [23] consists of assigning an identity to a segment (or group of segments) of speech. This comparison is made using a system called automatic speaker recognition (ASR) [23], [24] whose function is to determine whether these voice signals have been produced by the same person.

When the user presents him for identification, he just needs to say his password orally. The system analyzes the user's voice to determine his identity, and then compares the password said to that stored in the sound database.

In the case of an independent text access control system, a password will not be required. The system will analyze the voice and compare the unique characteristics of the voice with the stored data.

### 3.2. Global schema of the system

The speaker recognition system [22] is essentially consists of the following phases as shown in figure 1:

- Acquisition: Edition, collection of the voice.
- Parameterization: Acoustic analysis, processing, coding and extraction of characteristics.
- Learning: Identification, classification and storage.
- Recognition: Comparison [28] and verification of the speaker [26], [30].
- Decision: Comparison and verification of the threshold [27], [29].

Each of the system phases are themselves made up of elementary steps detailing the body of such a system.

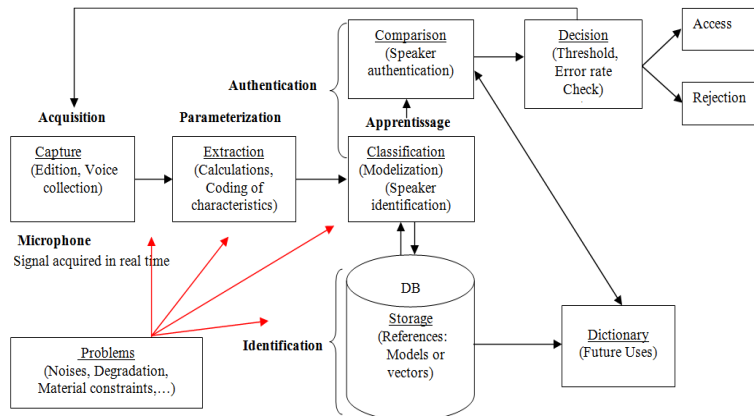


Figure 1: The global diagram of the test process for the speaker recognition system (SRS)

The speaker recognition procedure is done through an offline and online test process.

## 4. PROBLEMES

These speaker verification systems [6], [21], and [30] have seen more improvements in recent years [24], [29] and [30], but their performance remains insufficient due to several factors such as the recording conditions which are degraded (ambient noise, material quality, State of the speaker ...).

Also, this speaker verification operation [26] can be difficult if:

- The speaker change voluntarily his voice during the recording,
- The acquisition materials are very different,
- The initial registration is short-lived.
- Poor recognition of these voice prints.
- Precision errors on the decision threshold.

Under these conditions, the best current speech recognition systems [23], [26] present recognition difficulties. This reflects the interest represented by research in this domain.

## 5. METHODOLOGY

### 5.1. Preparation of the reference corpus

The following Table 1 collecting the acoustic characteristics, to realize the basic references containing Arabic numbers (0 to 9). The corpus record [1] is done for one Moroccan speaker.

Table 1: Acoustic characteristics of speech corpus

Parameter	Value
Format	Mono (.wav)
Sampling	8 Khz
Codage	16 Bits
Frames number	50
Recording time	5 Secondes/digit
Windowing	Hamming
Corpus	10 Arabic numerals

Optimization and improvement concerned the three phases that make up the body of such a system, as realized in our work: Analysis [4] phase, extraction [14] phase and recognition [25] phase.

## 5.2. Optimization of the parameterization phase

This step consists of configuring the speaker's voice signal, by selecting the relevant acoustic parameters, and by using a set of signal processing tools [1].

The work carried out in [5] and our work [1] to select the relevant and effective parameters, in order to improve our study corpus, the results obtained are collected in the Table 2 below, which contains the best values for the parameters used.

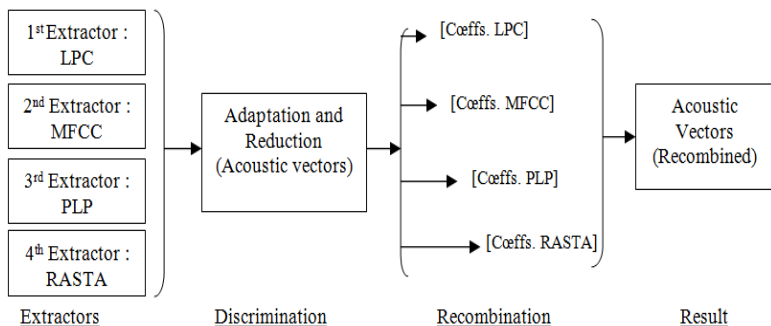
**Table 2:** Best values of the parameters applied to the speech signal

Parameter type	Best values	Performances (RPF (%))
Quantification	all values	90
Sampling	32 KHz	70
Formatting	single	100
Segmentation	64 frames	90
Windowing	Hamming	80

During the learning and acquisition process we must to set the acoustic parameters indicated above.

## 5.3. Optimization of the extraction phase

This task was done via the recombination [4] of the extraction methods; in fact, we carried out the extraction approach: multi-band recombination [14] which combines the extractors: LPC [13], MFCC [13], PLP [7], and RASTA [9], [14]. Figure 2 illustrates the multi-extractors approach (proposed solution) already achieved in our work [14], figure 2 below illustrates this recombination:



**Figure 2:** The multi-extractors recombination approach

### A. Recombination strategy

The proposed recombination as already realized in our work in [14], we considered for extractors: LPC, MFCC, PLP, and RASTA, this strategy takes the form:

## LPC-MFCC-PLP-RASTA

We adopted this recombination in order to condense and concatenate [9] the acoustic vectors, by ensuring a maximum of elementary resources, and by using the acoustic and by using the adapted and reduced [8], [10] acoustic elements.

For reasons of optimality and adaptability, a study carried out in [21] led to the transfer [18] and transcription [11] parameters; this task is ensured by the transition from a redundant representation to a reduced and appropriate representation.

For the classification phase, we used the HMM classifier, which was applied to the proposed recombination.

### B. Advantages of such recombination

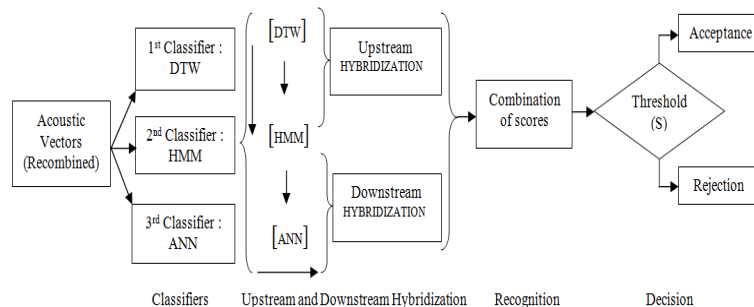
This extraction approach developed takes in common and at the same time the following aspects:

- Acoustic vectors adapted and discriminated,
- Reduced space of presentation,
- Benefits of the HMM advantages,
- Temporal and frequency aspect,
- Linear and non-linear aspect,
- Perceptual aspect.

## 5.4. Optimization of the classification phase

This an essential phase in speaker recognition systems [22] consists in classifying an unknown form in one of the possible categories.

This operation was carried out via the hybridization of classification methods. Figure 3 shows the hybridization [25] or the multi-classifiers approach (proposed solution), by merging the classifiers concerned:



**Figure 3:** The multi-classifiers hybridization approach

### A. Hybridization strategy

As already realized in our work in [25] we considered the three classifiers: DTW [15], [16], HMM [6], [17], and ANN [18], [19].

The proposed hybridization [25] takes the following form:

## DTW/HMM/ANN

This hybridization is spread over two parts:

1. **Upstream:** HMM hybridized with DTW, which is used in HMM to calculate the closest distance.

2. **Downstream:** HMM combined with ANN, applying HMM at the exit of ANN.

#### B. Advantages of such hybridization

The hybridization method created takes into account all three and at the same time the following aspects:

- Dynamic aspect,
- Statistical, linear aspect,
- Connectionist, non-linear aspect.

## 6. EXPERIMENTAL RESULTS

### 6.1. Evaluation rates

To assess the empirical results obtained, we used a factor called a RPF [3] (Recognition Performance Factor) defined as follows:

$$RPF = \frac{((Total\ number\ of\ trials - Number\ of\ recognized\ trials) / Total\ number\ of\ trials) * 100}{100} \quad (1)$$

### 6.2. Results of multi-extractors and multi-classifiers approaches

Table 3 shows the detailed results for the tests of a digit from 0 to 9 with several digits in the database {0, 1, 2... 9}, by the technique of multi-extractors recombination: LPC-MFCC - PLP-RASTA, with the multi-classifiers hybridization method: DTW/HMM/ANN.

**Table 3:** Detailed result of the Recognition of Arabic numerals

Digits	Recognition (RPF (%))
0 (Siffr) صفر	80
1 (wahed) واحد	95
2 (Ithnan) إثنان	85
3 (thalatha) ثلاثة	90
4 (Arbaa) أربعة	90
5 (Khamsa) خمسة	75
6 (Sitta) ستة	90
7 (Sabaa) سبعة	80
8 (Thamania) ثمانية	95
9 (Tisaa) تسعة	100
<b>Average rate</b>	<b>88</b>

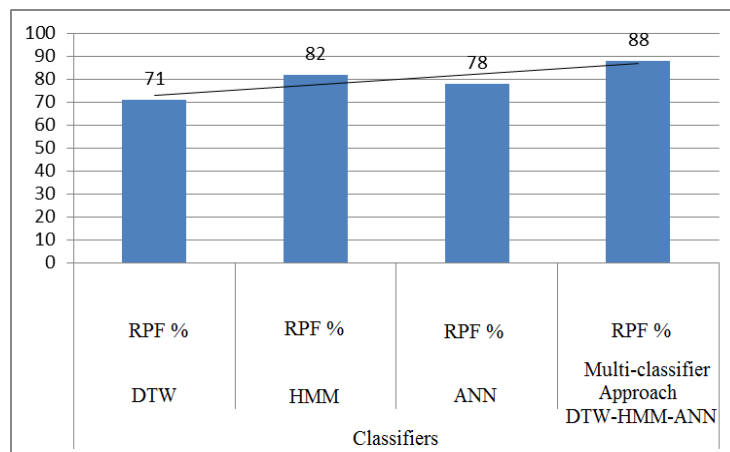
### 6.3. Comparison of the hybridized classification method with each individual classifier

We have collected in Table 4, the average results of the scores for the method of extracting (multi-stream), using each individual classifier: DTW, HMM, ANN, and the classifier resulting from the hybridization approach: DTW/HMM/ANN.

**Table 4:** Result of hybridization recognition with each classifier

Extractors	Classifiers			
	DTW RPF (%)	HMM RPF (%)	ANN RPF (%)	Multi-classifiers Approach DTW/HMM/ANN RPF (%)
LPC-MFCC-PLP-RASTA	71	82	78	88

Figure 4 shows the graphical representation of the table 4 above:



**Figure 4:** The result of recognition by the recombination and hybridization strategy

## 7. DISCUSSIONS

It clearly appears that the rate of recognition by the assembly of the two proposed recombination and hybridization techniques is increased by 6% with a recognition rate of 88%, compared to the best result (that of HMM 82%) with other classifiers each alone.

With the combination carried out along the study of extractors and the hybridization proposed at this level, we have achieved encouraging results, imagining other combinations and hybridizations in future work.

Despite the problems and constraints cited in the field of speech recognition which degrades the quality of recognition, and which makes the level of authentication lower, we have overcome some problems encountered at the start such as:

- Environmental noise,
- Intra speaker variability,
- Microphone position,
- Processing devices.

## 8. DECISION THRESHOLD

According to the work documented in [27], in any speaker verification application, it is necessary to determine the decision threshold before performing the verification, to accept

or reject the speaker to be verified. The threshold is given by the following formula:

$$S = \frac{p(o|locutor)}{p(o|no-locutor)} \quad (2)$$

The decision threshold is shown in the following figure 5:

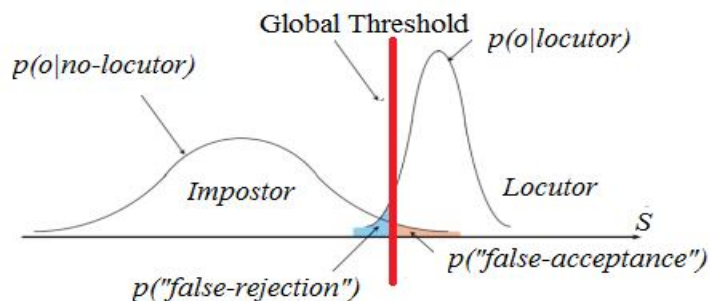


Figure 5: The decision threshold for speaker recognition.

## 9. CONCLUSION

Our work was concentrated and oriented towards research and studies to optimize combinations and hybridizations which respond to and solve certain problems of speaker recognition in general.

Finally, the idea of optimizing the speech signal and of having an optimized corpus, and the use of the recombination of extraction methods: LPC-MFCC-PLP-RASTA, to have fairly representative acoustic vectors, the incorporated training has been extended to classifiers: DTW / HMM / ANN with multiple flows.

These techniques bring a significant improvement to the speaker recognition performance, making it more robust and effective.

Without doubt, the possibilities of speaker recognition are numerous and are not limited to access control. We will be able to integrate such a speaker system into the telephone system, which will allow it to identify calling customers.

Hoping to make a compact tool containing all the information we need, in order to assess the impact of performance ... And really reduce a recognition system to a simple smart card.

## REFERENCES

1. A.LAMKADAM *et al.* **Impact of Acoustic Parameters on Speech Signal Processing: Optimization of the Recognition for Arabic Numerals**, IJSRD -International Journal for Scientific Research & Development- Vol. 5, Issue 05, p.p 1702-1706, July 2017.
2. M. Bergounioux. **Méthodes mathématiques pour le Traitement du Signal**, Université orléans, France, 5 September 2009.
3. Olivier Le Blouch. **Décodage acoustico-phonétique et applications à l'indexation audio automatique**,

- Université Toulouse III - Paul Sabatier, France, 12 June 2009.
4. A.LAMKADAM *et al.* **Comparative study and improvement of acoustic vectors extractors: Multiple streams applied to the recognition of Arabic numerals**, ISCV 2015, Fez, Morocco, 25-26 March 2015. <https://doi.org/10.1109/ISACV.2015.7106179>
5. Abdenour Hacine-Gharbi. **Sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole**, Université d'Orléans. France, 09 December 2012.
6. Mohamed CHETOUANI. **Codage neuro- prédictif pour l'extraction de caractéristiques de signaux de parole**, Université Pierre & Marie Curie, France, 14 December 2004.
7. Hemant. MISRA. **Multi-stream Processing for Noise Robust Speech Recognition**, Ecole polytechnique fédérale de Lausanne, 12 May 2006.
8. C.LEVY, G.LINARES, P.NOCERA, J-F.BONASTRE. **Reconnaissance de chiffres isolés embarquée dans un téléphone portable**, Journées d'études de la Parole, JEP Fez, Maroc, ACTN 2004.
9. A. RABAOUI, Z. LACHIRI *et* N. ELLOUZE. **Classification Robuste des Sons Environnementaux : Amélioration des Paramètres et Adaptation des Modèles**, TAIMA'07, Hammamet, Tunisie, 21-26 May 2007.
10. Loïc BARRAULT. **Diagnostic pour la combinaison de systèmes de reconnaissance automatique de la parole**, Laboratoire d'Informatique d'Avignon, 18 July 2008.
11. Benjamin LECOUTEUX. **Reconnaissance automatique de la parole guidée par des transcriptions a priori**, Université d'Avignon et des Pays de Vaucluse (Laboratoire d'Informatique EA 4128), 5 December 2008.
12. Didier MEUWLY. **Reconnaissance de locuteurs en sciences forensiques, l'apport d'une approche automatique**, Institut de police scientifique et de criminologie, Université de lausanne, 2001.
13. Urmila SHRAWANKAR. **Techniques for feature extraction in speech recognition system: A comparative study**, Dept. of Computer Science, SGB Amravati University, Amravati. Publié en Cornell University Library, 6 May 2013.
14. Abdelmajid LAMKADAM and Mohamed KARIM. **Extracting Multi Band Approach of Acoustic Vectors Extractors: Using HMM Classifier**, IJSTE - International Journal of Science Technology & Engineering, Volume 3, Issue 02/048, p.p 305-310 August 2016.
15. B. KUMAR PADHY & S. KUMAR SAHU. **Analysis of speech recognition techniques**, National Institute of Technology Rourkela - 769008, India, 10 October 2009.
16. C.LEVY, G.LINARES, P.NOCERA. **Comparaison of several acoustic modeling techniques and decoding algorithms for embedded speech recognition systems**, LIA / CERI Univ. Avignon France, December 2008.
17. C.LEVY, G.LINARES, P.NOCERA, Jean-François BONASTRE. **Reconnaissance de chiffres isolés embarquée dans un téléphone portable**, LIA, Avignon,

- & Stepmind SA, Cannet. *Journées d'études de la Parole, JEP*, Fez, Maroc, ACTN, 2004.
18. Michel WEINFELD. *Réseaux de neurones*, CNRS, Labo. D'Informatique de l'École polytechnique, Palaiseau, H 1990.
  19. TRANI & RAKHA. **MATLAB Toolboxes, Computer Applications in CEE**, Virginia Polytechnic Institute and State University, Spring, 2000.
  20. Bilal TOUNSI. **Inférence d'identité dans le domaine forensique en utilisant un système de reconnaissance automatique du locuteur adapté au dialecte Algérien**, Institut National de Formation en Informatique (I.N.I), Oued Smar, Alger, Algérie, 2008.
  21. Hassan EZAIDI. **Discrimination Parole/Musique et étude de nouveaux paramètres et modèles pour un système d'identification du locuteur dans le contexte de conférences téléphoniques**, Université du Québec à Chicoutimi, 10 October 2002.  
<https://doi.org/10.1522/17603685>
  22. Abdelmajid LAMKADAM *et al.* **Comparative and selective study of systems and libraries of speech recognition**, IJNTSE (International Journal of New Technologies in Science and Engineering), Vol. 5, Issue. 3, p.p 210-220, May 2018.
  23. Eduardo SANCHEZ-SOTO, Raphael BLOUET, Gerard CHOLLET and Marc SIGELLE. **Speaker Verification with Bayesian Networks**, Workshop on Multimodal User Authentication, p.p 61-65, MMUA, 2003.
  24. Waad BEN KHEDER. **Reconnaissance du locuteur en milieux difficiles**, Laboratoire d'Informatique d'Avignon, Université d'Avignon et des pays de vaucluse, 18 july 2017.
  25. Abdelmajid\* *et al.* **Comparative study and new approach multi classifiers: Application to the recognition of Arabic numerals**, IJESRT-International Journal of Engineering Sciences & Research Technology, ISSN: 2277-9655, 5(6), p.p 29-33, June 2016.
  26. Riadh AJGOU. **Reconnaissance Automatique du Locuteur à Travers les Canaux Digitaux**, Faculté des Sciences et de la Technologie de Biskra, Algérie, 14 February 2016.
  27. Dominique GENOUD, « **Reconnaissance et transformation de locuteurs** », Ecole Polytechnique Fédérale de Lausanne (EPFL), January 1999.
  28. Reda JOURANI. **Reconnaissance automatique du locuteur par des à grande marge**, Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier) - Université Mohammed V-Agdal, Rabat, 6 September 2012.
  29. François GRONDIN, « **Reconnaissance de locuteurs pour robot mobile** », Faculté de génie, Université de Sherbrooke, Québec, Canada, January 2012.
  30. Halim SAYOUD. **Reconnaissance Automatique du Locuteur-Approche connexionniste-**, Faculté d'Electronique et d'Informatique, USTHB. September 2003.