

Speech Recognition System based on Discrete Wave Atoms Transform Partial Noisy Environment

Mohamed Walid¹, Bousselmi Souha², Cherif Adnen³

Laboratory Analysis and processing of electrical and energy systems
Faculty of Sciences of Tunis, FST, Tunis, Tunisia

Abstract—Automatic speech recognition is one of the most active research areas as it offers a dynamic platform for human-machine interaction. The robustness of speech recognition systems is often degraded in real time applications, which are often accompanied by environmental noises. In this work, we have investigated the efficiency of combining wave atoms transform (WAT) with Mel-Frequency Cepstral Coefficients (MFCC) using Support Vector Machine (SVM) as classifier in different noisy conditions. A full experimental evaluation of the proposed model has been conducted using Arabic speech database (ARADIGIT) and corrupted with “NOISEUS database” noises at different levels of SNR ranging from -5 to 15dB. The results of Simulation have indicated that the proposed algorithm has improved the recognition rate (99.9%) at 15 dB of SNR. A comparative study was conducted by applying the proposed WAT-MFCC features to multilayer perceptron (MLP) and hidden Markov model (HMM) in order to prove the efficiency and the robustness of the proposed system.

Keywords—WAT; SVM; HMM; thresholding; noise; MFCC; MLP

I. INTRODUCTION

Automatic speech recognition allows the machine to understand and process information provided orally by a human user. It consists of using matching techniques to compare a sound wave to a set of samples, usually composed of words but also, more recently, phonemes (minimum sound unit). Speech recognition is based on the knowledge of several sciences: anatomy (The functions of the phonatory apparatus and the ear), the signals emitted by speech, phonetics, signal processing, linguistics, computer science, artificial intelligence and statistics. Automatic speech recognition opens new perspectives, given the considerable difference between manual and voice control. The use of natural language in the human-machine dialogue puts technology within the reach of all and leads to its popularization, reducing the constraints of the use of keyboards, mice and command codes to control. By simplifying the human-machine dialogue protocol, the automatic speech processing also aims to gain productivity since it is the machine that adapts to humans to communicate, not the other way around. In addition, it makes possible the simultaneous use of the eyes or hands to another task. It helps to humanize information management systems by focusing their design on users. A good speech recognition rates have been mostly reached using small vocabularies. In fact, this result is considered to be sufficient for the implementation of the most voice control devices. Thus, the error rate and learning time are steadily decreased. Also, this rate is

obviously variable and depends on vocabulary and language dialect. Sometimes, the ASR system has some troubles making it enables to avoid some linguistic traps. However, the development of sophisticated ASR systems with an improved speech recognition rates has become an interesting research subject for all scientific researchers in speech recognition domain. Indeed, several classification and parameterization methods have been emerged to accomplish this task. Among the most commonly used classification methods, we can cite HMM, SVM, Artificial neural networks (ANN), and Gaussian Mixture Model (GMM) [1]-[4]. Nevertheless, MFCC, Linear Predictive Coding (LPC), and Perceptual Linear Prediction (PLP) [5] constitute the well-known parameterization approaches. Among the most recent researches in the field of automatic speech recognition, attention may be attracted to [6] in which a deep Belief Networks (DBNs) has been presented at the aim to extract discriminative information using frames with a larger size in a speech signal. The desired objective in that work was to explore efficiency the DBNs in learning features which are more invariant to deep fluctuation in speech signal. Relying on other recognition methods, the author has tried to prove the reliability of his adopted technique which has significantly contributed to reduce the recognition error rate. However, in [7] an analysis of the impact of database size has been performed and that one of the impact of dialect in the context of independent-speaker text using SVM alone and then hybridized with GMM has been carried out too.

In this paper, an efficient isolated-words recognition system has been developed using SVM as classifier, and Mel-frequency Cepstral coefficient (MFCC) combined with discrete wave atoms transform (WAT) as feature extraction methods. Indeed, the adopted approach has been tested on Arabic language database in both clean and noisy conditions.

This manuscript is structured as follows: In Section II, a brief literature review of ASR Systems is presented. In Section III, the proposed speech recognition system is exhibited. However, an analysis and experimental results are given in Section IV. Concluding remarks and perspectives are presented in Section V.

II. RELATED WORKS

Several works have been applied for isolated-words recognition systems using different classification algorithms, such as machine learning via ANN, GMM, Dynamic Time Warping (DTW), K-Nearest Neighbor (KNN), SVM, and HMM. A summarize of the performances of some ASR systems are given in Table I.

TABLE I. LITERATURE SURVEY OF SPEECH RECOGNITION

REF	parameterization	classifier	database	Accuracy (%)
[8]	MFCC	ANN	Language Indian	99.84
		SVM		94.25
[9]	MFCC	DTW KNN	Language Quechua	91.1
[10]	MFCC	GMM KNN	Language Hindi	94.31
[11]	MFCC	DTW KNN	Language English	98.4
[12]	CMUSphinx toolkit	HMM	Language Indonesian	80
[13]	MFCC	HMM	Language Maly	80
[14]	MFCC	HMM	Language Maithili	95

III. THE PROPOSED SPEECH RECOGNITION SYSTEM

A new speech recognition system based on WAT-MFCC and SVM was developed in this paper to improve the accuracy of recognition. The general scheme of the proposed speech recognizer is depicted in Fig. 1. It contains three parts: the preprocessing step, the features extraction step and SVM classification. Each part is explained in the following sections of this manuscript.

A. Feature Extraction Stage

1) Pre-processing speech stage

a) *Pre-emphasize*: This step consists of signal filtering and the determination of its first order finite impulse response such as.

$$H(z) = 1 - \alpha Z^{-1} \quad 0.9 \leq \alpha \leq 1 \quad (1)$$

The signal is sampled at 16 kHz and then pre-emphasized to determine high frequencies which are less energetic than low ones.

b) *Windowing*: In this step, the signal is split into frames by multiplying the frame samples by a Hamming window of length 20 ms. This latter is given by the following expression:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N-1 \quad (2)$$

2) *Discrete wave atoms transform*: The wave atoms transform is an orthogonal transformation that was developed by Demanet. This transformation offers a parsimonious representation of a signal. In [15], a signal is assumed as an oscillatory model that can be described by the following function:

$$f(x) = \sin(Ng(x))h(x) \quad (3)$$

Where x represents the coordinate, g and h are C^∞ scale function. h Has a compact support in the closed set $[0,1]^2$ and N is a constant. In order to get sparse solution of signal f , wave atoms were proposed by Demanet and Ying in [15] and Demanet in [16].

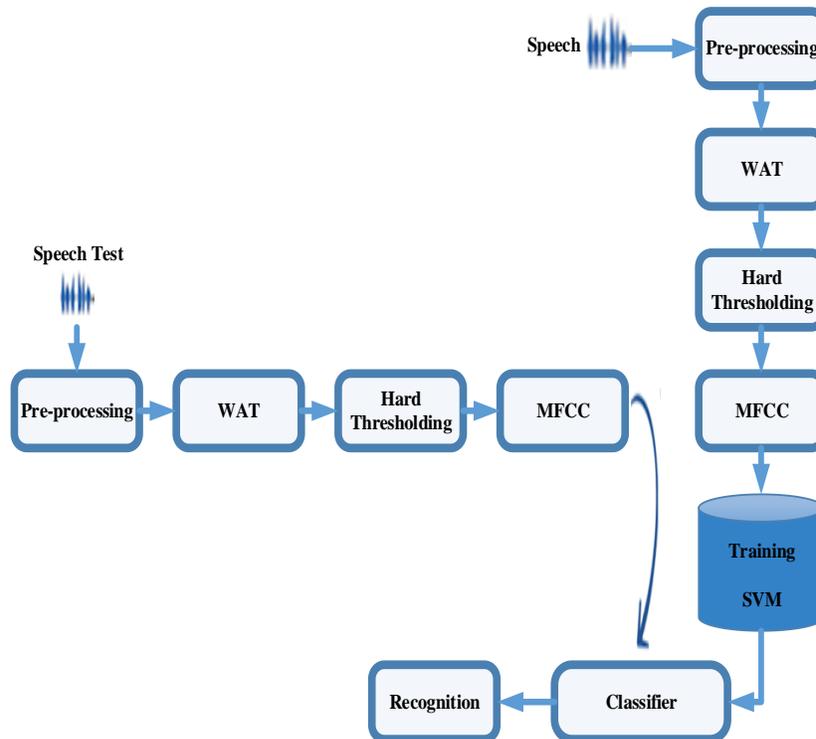


Fig. 1. Block Diagram of the Proposed Speech Recognition System.

The wave transform is a variant of wavelet packets. However, it allows a high-frequency localization that cannot be obtained using a filter bank as the case of the wavelet packets transform. To situate the wave atoms transform among the existing geometrical transforms, according to the analysis given by Demanet, two parameters are involved. The parameter, α represents the multi-scale aspect of a transformation and the parameter β , represents its directionality aspect. A transformation is uniform for $\alpha=0$, dyadic for ($\alpha=1$).

Fig. 2 shows the distribution of various transforms according to their multi-scale and multidirectional aspects.

When $\alpha=0$, the transform is not multi-scale, such as the case of the Gabor transform. When $\alpha=1$ the transform is multi-scale, this is the case of wavelets, ridgelets and curvelets. When $\beta=0$ the transform has a maximum selectivity as the Gabor transform where the decomposition can be done in all possible directions. Whereas $\beta=1$ represents a minimal selectivity, this is the case of wavelets where there is very little directional information.

In this context, the transformation into wave atoms is applicable in the case where $\alpha=\beta=1/2$. This transformation is as multi-scale as multidirectional. Therefore, it represents the compromise between the two properties.

Wave atoms 1D family function is defined as $\varphi_\mu(x)$, with subscript $\mu=(j, m, n)$. The indexed point (x_μ, ω_μ) in phase-space is defined as follows.

$$x_\mu = 2^{-j}m, \omega_\mu = \pi 2^j n$$

$$C_1 2^j \leq \max |m_i| \leq C_2 2^j \quad (4)$$

The elements of frame $\varphi_U(x)$ are named wave atoms when:

$$|\hat{\varphi}(\omega)| \leq C_M 2^{-j} (1+2^{-j}|\omega-\omega_\mu|)^{-M} + C_M 2^{-j} (1+2^{-j}|\omega+\omega_\mu|)^{-M}$$

$$|\varphi_\mu(x)| \leq C_M 2^j (1+2^j|x-x_\mu|)^{-M}$$

$$M > 0 \quad (5)$$

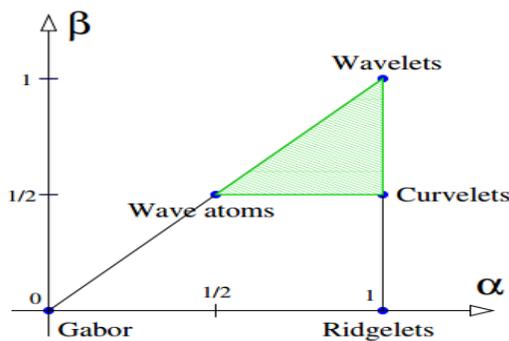


Fig. 2. Identification of Various Transforms as (α, β) Families of Wave Packets [16].

Practically, wave atoms are constructed from tensor products of a wavelet packet obeying the parabolic scaling law that is performed using decomposition as incomplete wavelet packets, as shown in Fig. 3. This figure describes the decomposition tree corresponding to the transformation into wave atoms defining the partitioning of the frequency axis in 1D. LH and RH respectively designate the left and right windows.

3) Thresholding: Thresholding is considered as a crucial step in speech recognition in tumultuous environments. It allows the rejection of the coefficients in which the WAT transformation is lower than a given threshold. There are several Thresholding methods, like the hard threshold and the soft threshold, which are the most used methods. In this article, we will apply the hard threshold given by the following equation:

$$C_{Re} = \begin{cases} C_{Re} & \text{if } |C_{Re}| \geq T \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

4) Mel-Scaled Frequency Cepstral Coefficients (MFCCs): Owing to their low complexity estimation and their good performance, MFCCs are commonly integrated in automatic speech recognition systems [17]. Thus, the MFCC representation appears to better approximate the structure of human auditory system than the traditional linear and predictive features. Although, MFCC coefficients are easily affected by common perturbations of random localized-frequency, to which human perception is largely insensitive. For each speech frame, a MFCC vector is computed as follows:

- The spectrum power of the windowed signal block is mapped onto the Mel scale using triangular filters.
- The Discrete Cosine Transform (DCT) is then applied to filter bank output logarithm.

The relationship between scale Mel and frequency is given by the following expression

$$f_{mel} = 2595 \times \log_{10} \left(1 + \frac{f}{700} \right) \quad (7)$$

An illustration of the computing of MFCCs coefficients is shown in Fig. 4.

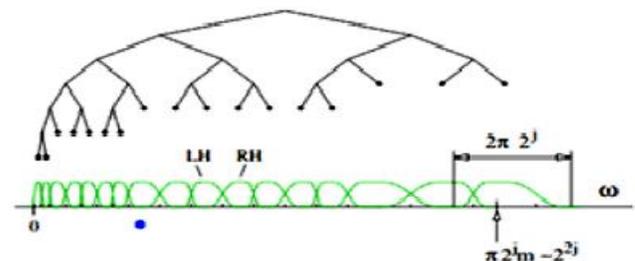


Fig. 3. Strategy of Wave Atoms and Corresponding Set of Sub-Bands [16].

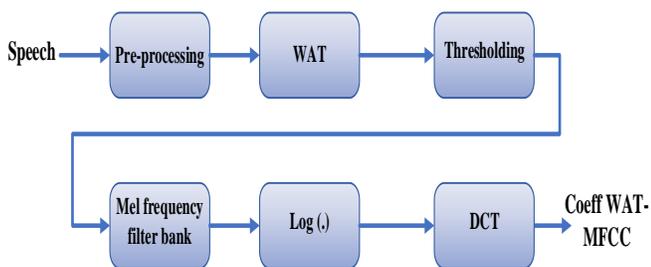


Fig. 4. Process to Create WAT- MFCC.

B. Support Vector Machine (SVM) Classification

SVM is considered as a segregating classifier associated with a few real-world applications. By maximizing the margin between boundary points of the classes and the separating hyper plane, SVM represents the “best” separating hyper plane (Fig. 5). These end-focuses are known as support vectors. SVMs apply linear and nonlinear separation hyper plans to classify data.

By specifying how SVM finds an ideal hyper plan that ranks the new models. The feature vector is noted by $x_i \in \mathbb{R}^n$, $i \in \{1, \dots, M\}$, where M is a number of training samples, n is the number of speech signal characteristics. The aim is to classify a speech unit into two classes $y_i = -1$ and $y_i = +1$, which design respectively the out vocabulary units and the hyper plane vocabulary unit.

The hyper plan form is described by the following equation:

$$\omega^t x + b = 0 \tag{8}$$

Where ω is a M dimensional vector normal to the plane and b is a scalar.

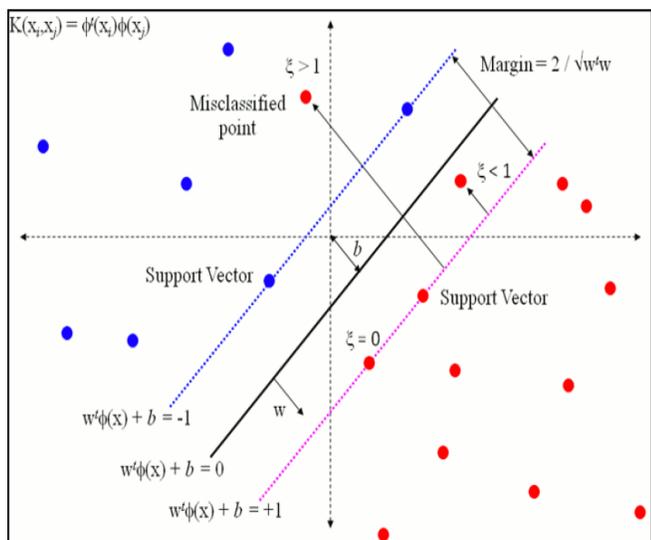


Fig. 5. Binary SVM Classification [18].

SVMs construct this hyper plan (ω and b) to be far from the boundary of both classes.

$$\omega^t x_i + b \geq 1 - \xi_i \text{ if } y_i = +1 \tag{9}$$

$$\omega^t x_i + b \leq -1 + \xi_i \text{ if } y_i = -1 \tag{10}$$

Where $\xi = (\xi_1, \dots, \xi_M)$ is slack variable, it controls the further processing of outliers, called “Soft-margin SVM”.

In order to find best hyper plan we have to maximize the margin with constraints combining in constraints (9) and (10).

In such case, the margin is equal to $\frac{1}{\|\omega\|}$; hence, this problem is equivalent to minimizing $\frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^M \xi_i$. This optimization problem can be solved by Lagrange Multipliers method [18].

$$L_p \equiv \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^M \xi_i - \sum_{i=1}^M \alpha_i y_i (x_i \cdot \omega + b) + \sum_{i=1}^M \alpha_i \tag{11}$$

It is a convex optimization problem and can be solved by Quadratic Programming method which returns the optimal α and it permits us to find ω by applying the formula given by (12).

In the next step, we have to describe the set of Support Vectors S which indices hold $\alpha_i > 0$ condition and calculate b:

$$\frac{\partial L_p}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^M \alpha_i y_i x_i \tag{12}$$

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{i=1}^M \alpha_i y_i = 0 \tag{13}$$

Substituting (12) and (13) into (11) gives Dual form L_D of the Primary L_p :

$$L_D \equiv \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i H_{ij} \alpha_j \tag{14}$$

Where

$$H_{ij} = y_i y_j x_i \cdot x_j$$

Minimizing L_p equivalent to maximizing L_D :

$$\sum_{i=1}^M \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, \dots, M$$

It is a convex optimization problem and can be solved by Quadratic Programming method which returns "best" α and it is allow us to calculate w by the formula (12).

In the next step we need to define the set of Support Vectors S which indices hold $\alpha_i > 0$ condition and calculate b :

$$b = \frac{1}{N_s} \sum_{s \in S} \left(y_s - \sum_{m \in S} \alpha_m y_m x_m \cdot x_s \right) \quad (15)$$

Each new sample x^* is classified by $y^* = \text{sgn}(\omega \cdot x^* + b)$ in the test process.

SVMs transform data (feature vectors) into a high dimensional space, whose training data become linearly separable, i.e. SVMs transform n -dimensional feature vector x into a N -dimensional feature vectors [19]:

$$\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^N$$

This transformation is realized via Kernel functions. There are different types of Kernel functions in SVM method such as:

Linear Kernel:
$$\mathbf{K}(x_i, x_j) = x_i^T x_j \quad (16)$$

Sigmoid Kernel:
$$\mathbf{K}(x_i, x_j) = \tanh(ax_i \cdot x_j - b) \quad (17)$$

Radial Basis Kernel:
$$\mathbf{K}(x_i, x_j) = e^{-\left(\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)} \quad (18)$$

Polynomial Kernel:
$$\mathbf{K}(x_i, x_j) = (x_i \cdot x_j + a)^b \quad (19)$$

Where a and b are kernel's parameters.

Each new sample x^* is classified by the following formula for the non-linear Kernel based SVM

$$y^* = \text{sgn} \left(\sum_{s \in S} y_s a_s \mathbf{K}(x^*, x_s) + b \right) \quad (20)$$

SVM classification is a binary classification machine-learning algorithm. For this reason, it is required to modify decision-making part of the algorithm for the multiclass tasks in real world situations. Two most popular methods are used for multiclass classification, which are "one against one" (OAO) and "one against all" (OAA) techniques. The OAA SVM multiclass approach involves the division of N class dataset into N two-class cases, whereas OAO approach involves training separate classifier for each pair of classes. This leads to $\frac{N(N-1)}{2}$ classifiers. This method is less sensitive to the imbalanced datasets but it is more computationally expensive. In this paper, we use "one-against-all" SVM multiclass technique.

IV. ANALYSIS AND RESULTS

A. Speech Database

The database used in this work is the ARADIGIT database [20], which is a local Arabic speech database, collected from 110 Algerian speakers who live in different regions of their country and speaks different dialects. The individuals chosen to record the words used in the database are aged between 18 and 50 years old. The database was recorded in a very quiet environment, at sample frequency 22.050 kHz and down-sampled at 16 kHz. In the training phase, they have used 1800 utterances pronounced by 60 speakers of both genders (30 male speakers and 30 female speakers) where each speaker repeats the same digit 3 times. In the testing phase, they have used 1000 utterances pronounced by 50 other speakers of both gender (25 males and 25 females) where each speaker repeats the same digit 2 times.

B. Experimental Results

Performances of the proposed system were tested on both clean and noisy signals. For noisy signals, we have corrupted data with three noises (white, car and babble) extracted from the NOISEX92 database (Varga et al. 1992). Noises are added to the speech signal in a SNR range from -5 to 15 dB with step size 5 dB. For training phase, clean speech database are explored. Indeed, noises are only added for testing the recognition performance. The robustness of the system is evaluated according to the recognition rate defined below.

$$Accur = \left[\frac{\text{Correctly recognised samples}}{\text{Total number of test samples}} * 100 \right] \quad (21)$$

Using three classifiers (SVM, HMM and MLP) in clean condition, the obtained recognition rates with MFCC only and WAT-MFCC parameterizations are presented in Fig. 6. Obviously, we can note that SVM combining MFCC soft and WAT-MFCC classifier has reached best performances in terms of recognition accuracy in comparison with other classifier as HMM and MLP. Thus, SVM based MFCC only reached 94.6% whereas using WAT-MFCC; we obtained a rate of 100%. In addition, it is remarkable that MFCC soft has contributed to reach competitive performances compared to WAT-MFCC. Despite worst performances have been obtained using MLP based MFCC soft with an achieved rate of 84.2%; the use of WAT-MFCC has registered an acceptable accuracy 92.4%.

Recognition results obtained from comparison between MFCC and WAT-MFCC using three speech recognition algorithms are illustrated in the following tables. Based on the displayed numbers in Tables II to IV, it has to note that WAT-MFCC has the ability to achieve the best recognition accuracy with respect to MFCC for all SNR levels and in all noisy conditions using SVM classifier. These results were achieved in comparison with HMM and MLP algorithms (see Tables II to zizv). Indeed, the best recognition rate (99.9%) was recorded at 15 dB of SNR. As shown, HMM

and MLP algorithms have contributed to give the second and third best recognition accuracies, respectively, at different levels of SNR under different noisy conditions. Indeed, these algorithms have recorded respectively 84.8% and 81.3% rates in "white" noise condition at 15 dB of SNR. Moreover, it can be seen that more the level of SNR increases, more the recognition accuracy increases too. In all tests, worse results were recorded with "babbel" noise compared to those obtained with "white" and "car" noises.

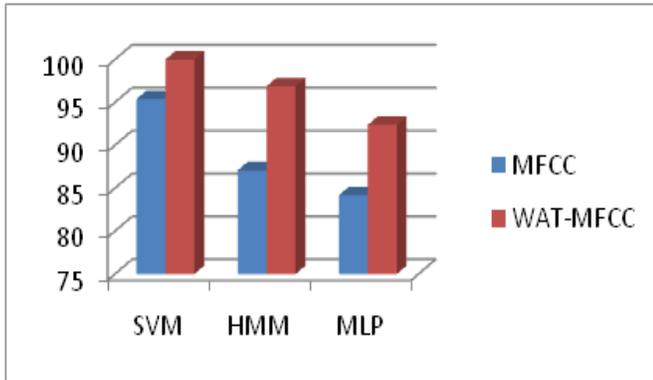


Fig. 6. Recognition Accuracy and Training Time using Three Classifiers with MFCC only and WAT-MFCC in Clean Condition.

TABLE II. COMPARISON OF THE OBTAINED RECOGNITION ACCURACY FOR MFCC ONLY AND WAT-MFCC USING DIFFERENT CLASSIFIERS IN CAR NOISE CONDITION

Noise	Speech recognition algorithm		SNR				
			-5db	0db	5db	10db	15db
Car	SVM	WAT-MFCC	67.4	72.7	84.8	95.4	96.7
		MFCC	60.60	65.4	70.4	75.7	90.90
	HMM	WAT-MFCC	63.6	65.2	71.2	75.1	83.8
		MFCC	57.5	60.60	62.1	74.2	80.30
	MLP	WAT-MFCC	60.60	62.3	66.6	73.4	81.2
		MFCC	54.5	54.5	59.6	64.4	71.9

TABLE III. COMPARISON OF THE OBTAINED RECOGNITION ACCURACY FOR MFCC ONLY AND WAT-MFCC USING DIFFERENT CLASSIFIERS IN WHITE NOISE CONDITION

Noise	Speech recognition algorithm		SNR				
			-5db	0db	5db	10db	15db
White	SVM	WAT-MFCC	68.7	86.3	90.90	92.4	99.9
		MFCC	64.3	68.9	75.3	84.8	90.90
	HMM	WAT-MFCC	60.60	65.9	68.9	73.4	84.8
		MFCC	56.8	59.1	63.6	67.4	73.4
	MLP	WAT-MFCC	57.5	60.60	65.1	69.6	81.3
		MFCC	53.6	58.3	62.8	65.90	76.9

TABLE IV. COMPARISON OF THE OBTAINED RECOGNITION ACCURACY FOR MFCC ONLY AND WAT-MFCC USING DIFFERENT CLASSIFIERS IN BABBLE NOISE CONDITION

Noise	Speech recognition algorithm		SNR				
			-5db	0db	5db	10db	15db
Babble	SVM	WAT-MFCC	66.6	71.8	84.6	91.90	96.9
		MFCC	62.8	64.4	66.8	77.2	80.30
	HMM	WAT-MFCC	63.6	65.90	73.4	75.7	81.5
		MFCC	59.6	61.1	65.90	70.6	73.4
	MLP	WAT-MFCC	60.60	63.6	71.2	73.4	77.9
		MFCC	56.8	60.60	65.8	71.2	71.6

From the obtained results, we can observe that SVM classifier with WAT-MFCC extraction features has proved its efficiency compared to HMM and MLP classifiers. Also, the adopted speech recognition method seems to be promising as it can greatly contribute to reach the best performances in noisy conditions.

V. CONCLUSION

In this paper, a new model for Arabic speech recognition using a combination of WAT with MFCC feature has been presented. The obtained results of the proposed method have shown that WAT is a powerful and effective technique in optimizing MFCC parameters. In fact, it has enhanced the learning ability of SVM which has given us a high recognition rate (100%) in clean environment. Furthermore, the assessment of the proposed method has been performed in noisy conditions without any speech enhancement algorithm. Also, it has been compared to other speech recognition approaches such as HMM and MLP. Moreover, the evaluation has shown that the proposed technique have reached the best performance (99.9%) in term of recognition rate in noisy condition at 15dB of SNR.

As a further work, we would like to test the proposed model on electronic architectures such as raspberry pi3, FPGA, and STM32 in order to more follow its evolution in real time applications.

REFERENCES

- [1] J. Ming, D. Crookes, Speech Enhancement Based on Full-Sentence Correlation and Clean Speech Recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25 (2017) 531–543.
- [2] Z-Q. Wang, D. Wang, A Joint Training Framework for Robust Automatic Speech Recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24 (2016) 796–806.
- [3] S. Chandrakala, N. Rajeswari, Representation Learning Based Speech Assistive System for Persons With Dysarthria, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25 (2017) 1510–1517.
- [4] Z.Tang, L. Li, D. Wang, R.Vipperla, Collaborative Joint Training With Multitask Recurrent Model for Speech and Speaker Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25 (2017) 493–504.
- [5] D. Namrata. Feature extraction methods LPC, PLP and MFCC in speech recognition. International journal for advance research in engineering and technology, 2013, vol. 1, no 6, p. 1-4.

- [6] M. Farahat, "Noise Robust Speech Recognition Using Deep Belief Networks", International Journal of Computational Intelligence and Applications, Vol. 15, No. 1 (2016) 1650005 (17 pages).
- [7] K. Yasmine, Z. and A. Amrouche, "SVM against GMM/SVM for Dialect Influence on Automatic Speaker Recognition Task", International Journal of Computational Intelligence and Applications, Vol. 13, No. 2 (2014) 1450012 (10 pages).
- [8] Londhe, N. D., & Kshirsagar, G. B. (2017). Speaker independent isolated words recognition system for Chhattisgarhi dialect. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).
- [9] Hernan Faustino Chacca Chuctaya, Rolfy Nixon Montufar Mercado and Jeyson Jesus Gonzales Gaona, "Isolated Automatic Speech Recognition of Quechua Numbers using MFCC, DTW and KNN" International Journal of Advanced Computer Science and Applications (ijacsa), 9(10), 2018.
- [10] P. Wani, U. G. Patil, D. S. Bormane, and S. D. Shirbahadurkar, "Automatic speech recognition of isolated words in Hindi language," in Proceedings - 2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016, 2017.
- [11] I., Muhammad, Atif and R. Gulistan. Isolated Word Automatic Speech Recognition (ASR) System using MFCC, DTW & KNN. In : 2016 Asia Pacific Conference on Multimedia and Broadcasting (APMediaCast). IEEE, 2016. p. 106-110.
- [12] P. Hamdan, F. Ridi, and H. Rudy. Indonesian Automatic Speech Recognition system using CMUSphinx toolkit and limited dataset. In : 2016 International Symposium on Electronics and Smart Devices (ISESD). IEEE, 2016. p. 283-286.
- [13] ANAND, Anu V., DEVI, P. Shobana, STEPHEN, Jose, et al. Malayalam Speech Recognition system and its application for visually impaired people. In : 2012 Annual IEEE India Conference (INDICON). IEEE, 2012. p. 619-624.
- [14] R. Rajeev, and D. R. Kumar. Isolated word recognition using HMM for Maithili dialect. In : 2016 International Conference on Signal Processing and Communication (ICSC). IEEE, 2016. p. 323-327.
- [15] L. Demanet, L. Ying, Wave atoms and sparsity of oscillatory patterns, Applied and Computational Harmonic Analysis. 23 (2007) 368-387.
- [16] L. Demanet, "Curvelets, wave atoms, and wave equations," PHD Thesis in California Institute of Technology, 2006.
- [17] S. Mada, and S. Zabidin. Implementasi Pengenalan Pola Suara Menggunakan," Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inferense System (ANFIS), " sebagai Kontrol Lampu Otomatis. ALHAZEN, 2014, vol. 1, no 1, p. 43-54.
- [18] A. Ganapathiraju, J. Hamaker and J. Picone, Applications of Support Vector Machines to Speech Recognition. IEEE Transactions on Signal Processing. 52 (2004) 2348-2355.
- [19] M. Elleuch, R. Mokni, M. Kherallah, Offline Arabic Handwritten Recognition System with Dropout applied in Deep Networks based-SVMs. International Joint Conference on Neural Networks, Vancouver, BC, Canada 2016, pp.3241-3248.
- [20] A. Amrouche "Reconnaissance automatique de la parole par les modèles connexionnistes" .Thèse de doctorat, faculté d' électronique et d' informatique, USTHB. 2007.