Evaluating an Amazigh Speech Recognition System in a Noisy Environment Using MFDWC Features

Fadwa Abakarim[1(✉)] and Abdenbi Abenaou[2]

[1] ENSEIRB-MATMECA, Bordeaux INP, University of Bordeaux, 33400 Talence, France
fadwa.abakarim@bordeaux-inp.fr

[2] National School of Applied Sciences, Ibn Zohr University, 80000 Agadir, Morocco

Abstract. This work proposes an Automatic Speech Recognition (ASR) system for the Amazigh language, designed to operate in both calm and noisy environments. The system preprocesses speech using Zero-Crossing Rate (ZCR) and Short-Time Energy (STE) to remove silence and compress the signals. Features are then extracted using Mel-Frequency Discrete Wavelet Coefficients (MFDWC). Classification is performed using a Support Vector Machine (SVM) with a One-versus-One (OvO) approach for multiclass problems. A database of 63600 recordings was created, including 53 isolated units (33 letters, 10 digits, and 10 words) from 120 Moroccan speakers, 60 native Tachelhit speakers and 60 non-native speakers. Recordings were made in a calm environment (personal office) and a noisy environment (café). Experimental results show that the proposed MFDWC based system outperforms traditional MFCC and DWT approaches, achieving recognition rates of 90.37% in calm conditions and 73.55% in noisy conditions. These results demonstrate the effectiveness and robustness of the system, providing a valuable tool for supporting Amazigh language preservation and applications such as language learning, voice-based interfaces, and automatic transcription.

Keywords: Automatic Speech Recognition · Amazigh language · Zero-Crossing Rate · Short-Time Energy · Mel-Frequency Cepstral Coefficients · Discrete Wavelet Transform · Mel-Frequency Discrete Wavelet Coefficients · Support Vector Machine · Calm environment · Noisy environment

1 Introduction

The development of Automatic Speech Recognition (ASR) systems for the Amazigh language is important, particularly in response to the growing demand for e-learning tools in this language. Implementing speech-based interfaces opens new perspectives for the development of applications and voice-controlled systems in Amazigh, thereby addressing the expectations of Amazigh-speaking communities seeking to benefit from speech technologies.

However, the development and deployment of ASR systems in real-world conditions face several challenges. The primary difficulty lies in the presence of various types

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16769, pp. 209–227, 2027.
https://doi.org/10.1007/978-3-032-30494-0_14

of noise stemming from environmental characteristics, communication channels, and recording devices. Moreover, the system may be used by speakers who differ considerably in terms of gender, age, and native or non-native status, making it necessary to account for speech variability within a diverse user population. All these factors directly affect the performance and reliability of speech recognition systems.

The objective of this work is to develop an Amazigh ASR system capable of operating in noisy conditions. Our previous study [1] focused on designing an Amazigh ASR system under a normal environment. The proposed system relies on the Mel-Frequency Discrete Wavelet Coefficients (MFDWC) [1, 2] feature extraction method, which offers effective time-frequency localization and improved robustness in noisy acoustic conditions.

A common challenge in signal processing and machine learning is the lack of sufficient input data. Training and evaluating intelligent systems typically require large and representative datasets to achieve accurate and reliable results.

In this study, we enhanced the previous database used in our earlier work. The initial version contained 33 letters and 10 digits of the Amazigh language recorded in a normal environment [1]. The updated version now includes 53 isolated Amazigh units (33 letters, 10 digits, and 10 words), collected from 120 speakers, among whom 60 are native speakers from the Agadir region and 60 are non-native speakers. The complete dataset contains 63600 audio samples, recorded in both a noisy environment (a café) and a calm environment (personal office).

The remainder of this article is organized as follows: Sect. 2 describes the methods employed to develop the system, Sect. 3 presents the experimental results, and Sect. 4 concludes the study.

2 Applied Methods

In recent years, there has been a growing interest in the Amazigh language, which has led to research efforts aimed at developing speech recognition systems specifically tailored to this language [3–11]. These initiatives seek to improve the accessibility of speech recognition technologies for Amazigh speakers, while also contributing to the preservation and promotion of their cultural heritage.

Previous studies on the development of automatic speech recognition systems for Amazigh have primarily focused on Tarifit [3–8]. However, in the present research, our focus is on the Amazigh language spoken in the city of Agadir and its surroundings, namely Tachelhit.

In the field of Amazigh speech recognition, Mel-Frequency Cepstral Coefficients (MFCC) are the most used features [3–11]. However, they are sensitive to noise in adverse acoustic environments [12–15]. A key step in their extraction involves the use of the Discrete Cosine Transform (DCT).

The DCT basis vectors have similar temporal and frequency resolutions, meaning that all coefficients are affected by noise present within a specific frequency band [16]. This limitation reduces the effectiveness of the MFCC method in noisy environments. To address this issue, the use of the Discrete Wavelet Transform (DWT) can be advantageous due to its improved time-frequency localization [12–16].

Furthermore, available memory space is an important criterion when choosing a feature extraction method that achieves good accuracy with a limited feature vector size. The DWT method is a suitable option when storage capacity is constrained, whereas the MFCC method requires more storage space [13, 15].

Consequently, in many cases, the MFCC method is combined with other feature selection techniques aimed at reducing the dimensionality of the extracted feature vectors while maintaining good accuracy. Principal Component Analysis (PCA) is one such technique [15].

As part of the development of a new methodology for creating a noise-robust Amazigh automatic speech recognition system, we propose an approach combining MFCC and DWT methods. This approach, called Mel-Frequency Discrete Wavelet Coefficients (MFDWC) [1], provides good time-frequency localization and enhanced robustness in noisy environments.

This article proposes an Amazigh ASR system based on three main stages: preprocessing, feature extraction, and classification.

2.1 Preprocessing

The combination of Zero Crossing Rate (ZCR) and Short-Time Energy (STE) [17, 18] provides an effective method for detecting both speech activity and silence in a vocal signal.

Low ZCR values combined with high STE indicate the presence of speech, i.e., pronounced words. Using this criterion, it is possible to identify segments containing active speech.

Segments where ZCR is high and STE is low, corresponding to silence, are removed, resulting in purified speech signals that retain only the pronounced words while effectively eliminating silent periods.

Figure 1 illustrates the operation of the ZCR-STE combination.

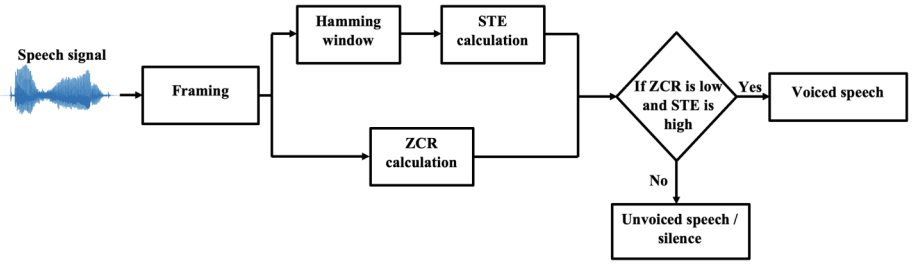

Fig. 1. The ZCR-STE process.

Figure 2 shows the speech signal of the Amazigh word “Azul” before and after applying the ZCR-STE combination.

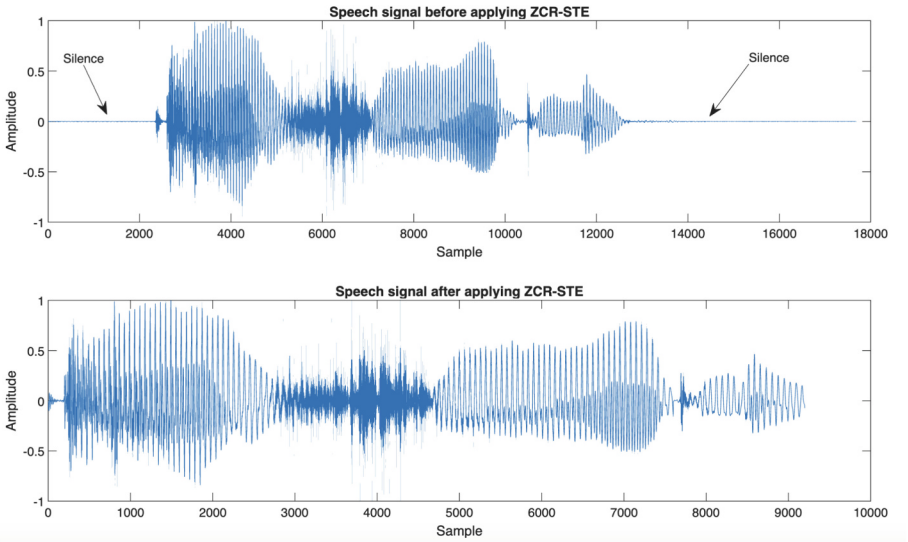

Fig. 2. Silence removal in the speech signal of the Amazigh word "Azul" using the ZCR-STE method.

2.2 Feature Extraction

After removing silent periods using the ZCR-STE combination, the speech signals are ready for feature extraction. In our previous work, we demonstrated the effectiveness of Mel-Frequency Discrete Wavelet Coefficients (MFDWC) for Amazigh speech recognition in a normal environment [1]. MFDWC combines the advantages of Mel-frequency analysis with DWT [14, 16], providing good time-frequency localization and improved robustness of extracted features.

In the present study, we extend the application of MFDWC to a noisy environment using a newly recorded dataset, where all speech samples were captured in both a café setting and a quiet environment, allowing for a comparison of results between clean and noisy conditions. This adaptation enables us to evaluate the robustness of MFDWC under real-world noise scenarios. While the overall feature extraction process follows the same principles as in our previous study [1], the performance under noisy conditions highlights the method's ability to maintain discriminative power despite challenging acoustic conditions.

Each speech frame is first processed through a high-pass pre-emphasis filter to enhance high-frequency components and reduce undesired low-frequency energy. The speech signal is then segmented into overlapping frames of 25 ms with a 10 ms overlap, and each frame is multiplied by a Hamming window to reduce spectral leakage. The Mel-scale filter bank is applied to the power spectrum of each frame, and the discrete wavelet transform is then performed on the logarithm of the filter bank energies to generate the MFDWC features [1].

The choice of the mother wavelet, Daubechies4, is consistent with our previous work [1]; however, unlike the previous study where 16 MFDWC features were obtained

in a normal environment, 11 coefficients are extracted in this study to optimize the representation for noisy conditions.

Figure 3 illustrates the complete process of MFDWC feature extraction.

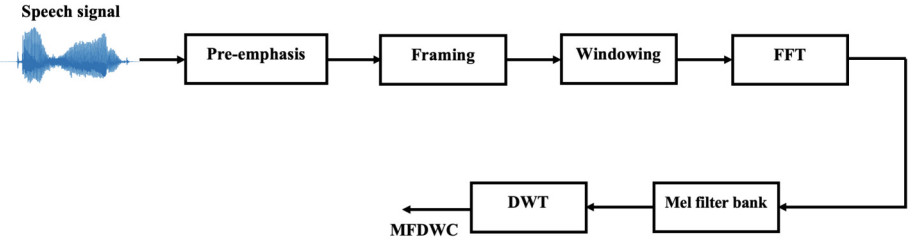

Fig. 3. The MFDWC feature extraction process [1].

2.3 Classification

After feature extraction, the next step is classification, which is crucial for achieving high recognition performance. Various models are commonly used in speech recognition, including supervised approaches such as Multilayer Perceptron (MLP) [19, 20] and Hidden Markov Models (HMM) [14, 21], as well as unsupervised methods such as Dynamic Time Warping (DTW) [22] and K-means [23]. Each method has specific limitations: MLPs may struggle with large vocabularies, HMMs rely on the Markov assumption that ignores long-term dependencies, DTW requires extensive computation and memory, and K-means may perform poorly when clusters are not well separated [13, 14].

Considering these limitations, we adopted Support Vector Machine (SVM) for classification in this study. SVMs are widely recognized in speech recognition due to their ability to handle non-linearly separable data through kernel functions, capturing complex relationships between acoustic features and speech classes [14, 24–26]. They also offer high accuracy, robustness to noisy data, and efficient decision-making, making them particularly suitable for real-world noisy environments.

As SVM is inherently a binary classifier, multiclass problems must be decomposed into multiple binary classification tasks. Two common approaches are One-versus-All (OvA) and One-versus-One (OvO). In OvA, each class is treated as positive against all other classes as negative, requiring K binary classifiers for K classes. In OvO, a binary classifier is trained for each pair of classes, resulting in $K(K-1)/2$ classifiers. While OvO involves more classifiers, each is trained on fewer samples and solves a simpler two-class problem, often leading to faster and more robust multiclass classification.

In this study, we employed the OvO approach with a Radial Basis Function (RBF) kernel. This combination balances computational efficiency and accuracy while effectively handling noisy speech signals.

3 Experimental Results and Discussion

3.1 Amazigh Speech Database

For training and testing our system, we created a dedicated database. In our previous work [1], we developed an Amazigh speech database containing 33 letters and 10 digits recorded in a normal environment. In this study, our goal is to evaluate the MFDWC method using both noisy and calm data. The newly created database consists of 53 isolated Amazigh units: 33 letters, 10 digits, and 10 isolated words.

A total of 120 Moroccan speakers participated in this study, including 60 native speakers from the Agadir region and 60 non-native speakers who spoke Amazigh-Tachelhit for the first time. The complete database contains 63600 speech recordings (53 units x 10 repetitions x 120 speakers).

Table 1 presents the 10 isolated words of the Amazigh-Tachelhit language along with their correspondences in English and Arabic. The 33 letters and 10 digits are presented in our previous work [1].

Table 1. The 10 isolated words of the Amazigh-Tachelhit language with their English and Arabic correspondences.

Tifinaghe-IRCAM [1]	English transcription	English correspondence	Arabic transcription	Arabic correspondence
ⵄⴰⵣⵓⵍ	Azul	Hello	أزول	مرحبًا
ⵜⴰⵏⵎⵎⵉⵔⵜ	Tanmmirt	Thank you	تانميرت	شكراً
ⵜⴼⵓⵍⴽⵉ	Tfulki	Beautiful	تفولكي	جميلة
ⵉⴷⴳⴰⵎ	Idgam	Yesterday	إضكام	أمس
ⵖⴰⵙⵙⴰⴷ	ɣassad	Today	غَسّاد	اليوم
ⵄⴰⵛⴽⵉⴷ	Achkid	Come	أشكيد	تعال
ⵖⵉⵍⴰⴷ	ɣilad	Now	غيلاد	الآن
ⵉⵎⵏⵙⵉ	Imnsi	Dinner	إمنسي	عشاء
ⵉⵎⴽⵍⵉ	Imkli	Lunch	إمكلي	غداء
ⵄⵖⵓⵊⵊⵔ	Aɣujjr	Breakfast	أغوجّر	فطور

The database used in this study is divided into two distinct parts. The first 31800 recordings were made in a quiet environment, specifically in a personal office, by 60 speakers. These recordings were captured under controlled conditions with minimal background noise, providing a calm acoustic context suitable for speech signal analysis.

The remaining 31800 recordings were made in a noisy environment, specifically in a café, by the other 60 speakers.

The noisy environment consisted of ambient acoustic conditions typical of a café setting, including overlapping background conversations, background music, and general crowd noise. No artificial noise was added or mixed in post-processing; all recordings were made directly in the natural café environment without noise filtering or enhancement. This ensures that the noise conditions reflect real-world deployment scenarios for speech recognition systems.

Figure 4 illustrates the speech signals of the word "Azul" in both environments, quiet and noisy, along with their corresponding spectrograms.

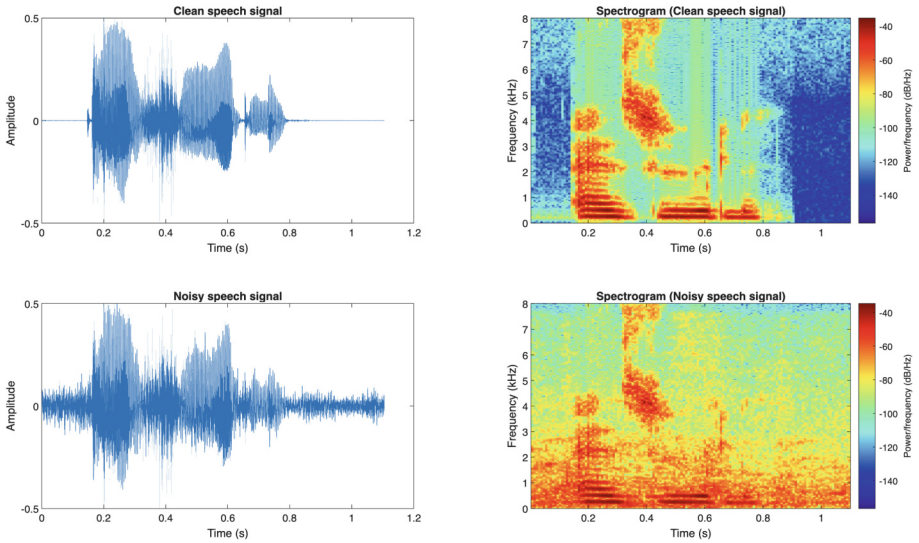

Fig. 4. The vocal signals of the word "Azul" in both quiet and noisy environments, along with their spectrograms.

A total of eight experiments were performed to train and test the system. The first four experiments used data from a quiet environment, while the last four experiments used data from a noisy environment.

For training, 70% of the database was used, while the remaining 30% was reserved for testing. Additionally, the speakers used in the testing phase were different from those used in the training phase.

Tables 2 and 3 summarize the different experiments conducted, along with the corresponding databases used for training and testing.

Table 2. The different experiments conducted in the calm environment.

Experiments	Calm environment			
	Training data		Test data	
	Number of audio files	Number of speakers	Number of audio files	Number of speakers
Experiment 1: 33 letters	13860	42 (30 adults and 12 children)	5940	18 (10 adults and 8 children)
Experiment 2: 10 digits (0–9)	4200	42 (30 adults and 12 children)	1800	18 (10 adults and 8 children)

(*continued*)

Table 2. (*continued*)

Experiments	Calm environment			
	Training data		Test data	
	Number of audio files	Number of speakers	Number of audio files	Number of speakers
Experiment 3: 10 isolated words	4200	42 (30 adults and 12 children)	1800	18 (10 adults and 8 children)
Experiment 4: 33 letters + 10 digits + 10 isolated words	22260	42 (30 adults and 12 children)	9540	18 (10 adults and 8 children)

Table 3. The different experiments conducted in the noisy environment.

Experiments	Noisy environment			
	Training data		Test data	
	Number of audio files	Number of speakers	Number of audio files	Number of speakers
Experiment 5: 33 letters	13860	42 (30 adults and 12 children)	5940	18 (10 adults and 8 children)
Experiment 6: 10 digits (0–9)	4200	42 (30 adults and 12 children)	1800	18 (10 adults and 8 children)
Experiment 7: 10 isolated words	4200	42 (30 adults and 12 children)	1800	18 (10 adults and 8 children)
Experiment 8: 33 letters + 10 digits + 10 isolated words	22260	42 (30 adults and 12 children)	9540	18 (10 adults and 8 children)

3.2 Calm Environment

Experimental Results: First Experiment. Table 4 reports the recognition performance obtained for the 33 Amazigh letters using 5940 speech recordings.

The MFCC based system achieved an overall accuracy of 78.27%, with the best results for YASH (ⵛ) and YAZ (ⵣ) (87.22% and 88.33%), while YA (ⴰ) and YO (ⵓ) showed lower accuracies (72.22% and 70.56%). The DWT based system yielded a higher overall accuracy of 83.01%, notably recognizing YAD (ⴷ) and YAM (ⵎ) with 91.11% and 92.78%, whereas YEY (ⵉⵢ) and YI (ⵉ) remained comparatively low (75.00% and 73.89%). The combined MFDWC approach achieved the best performance, reaching an overall accuracy of 88.52%. It provided high recognition rates for YAD (ⴷ) and YAT (ⵟ)

(93.89% and 95.00%), although YA (ⴰ) and YI (ⵉ) remained among the least accurately recognized (83.33% and 81.67%). Overall, the MFDWC based system outperformed MFCC and DWT by 10.25% and 5.51%, respectively, confirming the effectiveness of the combined feature extraction strategy.

Table 4. Results of the first experiment.

Letters	Recognition rate (%)		
	MFCC	DWT	MFDWC
YA	72.22	77.78	83.33
YAB	75.00	81.11	85.00
YAG	77.78	79.44	90.00
YAGG	73.89	78.33	88.33
YAD	76.11	91.11	93.89
YADD	73.89	87.22	91.11
YEY	76.11	75.00	88.89
YAF	80.00	77.22	90.00
YAK	76.67	83.33	91.11
YAKK	73.33	85.56	87.22
YAH	82.78	88.33	89.44
YAHH	74.44	78.89	90.56
YAAA	76.11	81.11	87.78
YAKH	75.00	83.89	84.44
YAQQ	83.33	88.89	92.78
YI	76.11	73.89	81.67
YAJ	80.00	90.00	88.33
YAL	76.67	87.22	92.78
YAM	83.89	92.78	91.67
YAN	81.67	86.11	90.56
YO	70.56	80.00	84.44
YAR	82.78	84.44	90.00
YARR	83.89	81.67	88.33
YAGH	82.22	86.67	86.11
YAS	80.56	83.33	87.22
YASS	76.67	85.00	83.89
YASH	87.22	82.78	85.56
YAT	75.56	80.00	95.00
YATT	74.44	76.67	92.78

(*continued*)

Table 4. (*continued*)

Letters	Recognition rate (%)		
	MFCC	DWT	MFDWC
YAW	76.11	78.89	90.56
YAY	75.56	86.11	89.44
YAZ	88.33	83.89	85.00
YAZZ	83.89	82.78	83.89
Total recognition rate (%)	**78.27**	**83.01**	**88.52**

Experimental Results: Second Experiment. Table 5 summarizes the results of the second experiment, conducted on 1800 speech recordings to test the recognition performance of the first ten Amazigh digits.

The MFCC based system achieved an overall accuracy of 85.56%, with AMYA (ⴰⵎⵢⴰ) and KRAD (ⴽⵔⴰⴹ) obtaining the highest rates (91.67% and 93.33%), while SA (ⵙⴰ) and TZA (ⵜⵣⵉ) showed the lowest (80.00% and 78.89%). The DWT based system performed better, reaching 88.22% overall accuracy; YEN (ⵢⴻⵏ) and KRAD (ⴽⵔⴰⴹ) achieved the highest recognition rates (94.44% and 95.56%), whereas SA (ⵙⴰ) and TZA (ⵜⵣⵉ) remained the least accurate (81.67% and 82.78%). The combined MFDWC approach yielded the best performance, with an overall accuracy of 93.06%. YEN (ⵢⴻⵏ) and KRAD (ⴽⵔⴰⴹ) reached 96.11% and 97.78%, while KOZ (ⴽⵓⵣ) and SA (ⵙⴰ) recorded the lowest rates (90.56% and 89.44%). Overall, the MFDWC based system outperformed MFCC and DWT by 7.50% and 4.84%, respectively, confirming the effectiveness of combining both feature extraction techniques.

Table 5. Results of the second experiment.

Digits	Recognition rate (%)		
	MFCC	DWT	MFDWC
AMYA	91.67	90.56	93.33
YEN	88.33	94.44	96.11
SIN	83.89	88.33	91.11
KRAD	93.33	95.56	97.78
KOZ	81.67	83.33	90.56
SMMUS	83.89	84.44	91.67
SDES	85.56	90.00	93.89
SA	80.00	81.67	89.44
TAM	88.33	91.11	94.44
TZA	78.89	82.78	92.22

(*continued*)

Table 5. (*continued*)

Digits	Recognition rate (%)		
	MFCC	DWT	MFDWC
Recognition rate (%)	**85.56**	**88.22**	**93.06**

Experimental Results: Third Experiment. Table 6 summarizes the results of the third experiment, conducted on 1800 speech recordings to evaluate the recognition performance of ten isolated Amazigh-Tachelhit words.

The MFCC based system achieved an overall accuracy of 88.56%, with Tfulki (ⵜⴼⵓⵍⴽⵉ) and Idgam (ⵉⴷⴳⴰⵎ) obtaining the highest rates (94.44% and 95.56%), while Achkid (ⴰⵛⴽⵉⴷ) and Imnsi (ⵉⵎⵏⵙⵉ) recorded the lowest (83.33% and 81.67%). The DWT based system performed better, reaching 90.11% overall accuracy; Idgam (ⵉⴷⴳⴰⵎ) and Imkli (ⵉⵎⴽⵍⵉ) achieved the best recognition rates (96.11% and 95.00%), whereas ɣassad (ⵖⴰⵙⵙⴰⴷ) and Imnsi (ⵉⵎⵏⵙⵉ) remained the least accurate (85.56% and 83.89%). The combined MFDWC method yielded the highest overall accuracy of 94.67%. Tfulki (ⵜⴼⵓⵍⴽⵉ) and Idgam (ⵉⴷⴳⴰⵎ) were particularly well recognized (97.78% and 98.89%), while ɣassad (ⵖⴰⵙⵙⴰⴷ) and Imnsi (ⵉⵎⵏⵙⵉ) showed the lowest rates (91.67% and 90.56%). Overall, the MFDWC based system surpassed MFCC and DWT by 6.11% and 4.56%, respectively, confirming the effectiveness of combining both feature extraction techniques for isolated Amazigh word recognition.

Table 6. Results of the third experiment.

Amazigh-Tachelhit words	Recognition rate (%)		
	MFCC	DWT	MFDWC
Azul	92.78	89.44	92.78
Tanmmirt	86.67	87.22	95.00
Tfulki	94.44	91.11	97.78
Idgam	95.56	96.11	98.89
ɣassad	88.33	85.56	91.67
Achkid	83.33	88.33	94.44
ɣilad	86.11	92.78	93.89
Imnsi	81.67	83.89	90.56
Imkli	88.89	95.00	96.11
Aghujjr	87.78	91.67	95.56
Recognition rate (%)	**88.56**	**90.11**	**94.67**

Tables 4, 5, and 6 clearly indicate that the recognition system for the 10 isolated Amazigh-Tachelhit words performs better than the recognition systems for the 33 letters

and 10 digits. This difference in performance can be largely attributed to the length of the units, as letters and digits are shorter compared to the tested Amazigh-Tachelhit words.

Experimental Results: Fourth Experiment. Table 7 presents the overall results of the fourth experiment, which used a total of 9540 speech recordings for testing. These recordings included a mixture of 33 letters, 10 digits, and 10 isolated Amazigh words.

The MFCC based recognition system for the 53 Amazigh units achieved an overall recognition rate of 81.29%. In comparison, the DWT based system reached an overall recognition rate of 85.18%, showing a difference of 3.89% and indicating that the DWT based system provides superior performance.

Furthermore, by combining the MFCC and DWT methods (MFDWC), the recognition system for the 53 Amazigh units achieved an overall recognition rate of 90.37%. This represents a significant improvement of 9.08% over the MFCC based system and a difference of 5.19% over the DWT based system.

These results clearly demonstrate that the combination of the two methods provides better recognition performance for Amazigh words.

Table 7. Results of the fourth experiment.

Overall data	Recognition rate (%)		
	MFCC	DWT	MFDWC
33 letters + 10 digits + 10 words	81.29	85.18	90.37

3.3 Noisy Environment

Experimental Results: Fifth Experiment. Table 8 summarizes the results of the fifth experiment, conducted on 5940 speech recordings to test the recognition performance of the 33 Amazigh letters under noisy conditions.

The MFCC based system obtained an overall accuracy of 49.73%, with YASH (ⵛ) and YAD (ⴷ) achieving the highest recognition rates (54.44% and 55.00%), while YA (ⵄ) and YO (ⵣ) were the least accurately recognized (45.00% and 43.89%). Compared to the calm environment, the accuracy dropped by 28.54%. The DWT based system performed better, reaching 66.84% overall accuracy; YAD (ⴷ) and YAM (ⵎ) achieved the highest rates (74.44% and 75.56%), whereas YA (ⵄ) and YI (ⵉ) recorded the lowest (60.00% and 58.33%). The decrease in accuracy between calm and noisy environments was 16.17%. When combining both techniques, the MFDWC method achieved the best performance, with an overall accuracy of 71.15%. YAD (ⴷ) and YAT (ⵟ) obtained the highest recognition rates (77.78% and 79.44%), while YA (ⵄ) and YI (ⵉ) remained the least recognized (65.56% and 63.89%). The drop in performance compared to calm conditions was 17.37%. Overall, the MFDWC based system outperformed MFCC and DWT by 21.42% and 4.31%, respectively, confirming the robustness and effectiveness of the combined feature extraction approach in noisy environments.

Table 8. Results of the fifth experiment.

Letters	Recognition rate (%)		
	MFCC	DWT	MFDWC
YA	45.00	60.00	65.56
YAB	48.33	68.33	71.67
YAG	51.11	72.78	69.44
YAGG	47.78	70.00	72.78
YAD	55.00	74.44	77.78
YADD	46.67	71.67	75.00
YEY	48.33	62.22	68.33
YAF	50.56	66.67	71.11
YAK	46.11	67.78	73.33
YAKK	47.78	65.00	72.22
YAH	47.22	66.67	68.89
YAHH	46.11	63.89	67.22
YAAA	51.67	71.11	72.78
YAKH	50.00	65.56	67.78
YAQQ	52.22	67.78	73.89
YI	48.33	58.33	63.89
YAJ	46.11	64.44	75.56
YAL	52.22	68.89	72.78
YAM	53.89	75.56	77.22
YAN	50.00	71.67	72.22
YO	43.89	66.67	66.67
YAR	47.78	70.00	72.78
YARR	48.89	63.89	68.89
YAGH	51.11	61.11	71.67
YAS	52.78	67.22	67.78
YASS	48.89	65.56	70.56
YASH	54.44	68.33	66.11
YAT	53.33	70.56	79.44
YATT	51.11	65.56	76.67
YAW	48.89	62.78	73.33
YAY	51.67	66.11	70.56
YAZ	53.33	63.89	67.22

(*continued*)

Table 8. (*continued*)

Letters	Recognition rate (%)		
	MFCC	DWT	MFDWC
YAZZ	50.56	61.11	66.67
Recognition rate (%)	**49.73**	**66.84**	**71.15**

Experimental Results: Sixth Experiment. Table 9 summarizes the results of the sixth experiment, conducted on 1800 speech recordings to evaluate the recognition of the first ten Amazigh digits under noisy conditions.

The MFCC based system achieved an overall accuracy of 56.50%, with AMYA (ⴰⵎⵢⴰ) and KRAD (ⴽⵔⴰⴷ) reaching the highest rates (66.67% and 68.33%). Conversely, TZA (ⵜⵣⴰ) and SA (ⵙⴰ) recorded the lowest accuracies (50.00% and 48.89%). Compared to calm conditions, performance decreased by 29.06%. The DWT based system performed significantly better, with an overall accuracy of 72.11%; YEN (ⵢⴻⵏ) and KRAD (ⴽⵔⴰⴷ) obtained the best results (76.67% and 78.33%), while TZA (ⵜⵣⴰ) and SA (ⵙⴰ) remained less accurately recognized (66.67% and 65.00%). The drop in accuracy relative to calm conditions was 16.11%. The combined MFDWC method further improved performance, achieving an overall accuracy of 78.83%. YEN (ⵢⴻⵏ) and KRAD (ⴽⵔⴰⴷ) showed the highest recognition rates (82.78% and 83.89%), whereas KOZ (ⴽⵓⵣ) and SA (ⵙⴰ) were the least recognized (76.11% and 75.00%). The accuracy difference between calm and noisy environments was reduced to 14.23%. Overall, the MFDWC based system outperformed MFCC and DWT by 22.33% and 6.72%, respectively, confirming the robustness and superiority of the combined feature extraction strategy under noisy conditions.

Table 9. Results of the sixth experiment.

Digits	Recognition rate (%)		
	MFCC	DWT	MFDWC
AMYA	66.67	72.22	79.44
YEN	58.33	76.67	82.78
SIN	53.88	71.11	78.33
KRAD	68.33	78.33	83.89
KOZ	50.56	70.00	76.11
SMMUS	52.22	71.67	77.22
SDES	55.00	73.89	80.00
SA	48.89	65.00	75.00
TAM	61.11	75.56	78.89
TZA	50.00	66.67	76.67

(*continued*)

Table 9. (*continued*)

Digits	Recognition rate (%)		
	MFCC	DWT	MFDWC
Recognition rate (%)	**56.50**	**72.11**	**78.83**

Experimental Results: Seventh Experiment. Table 10 summarizes the results of the seventh experiment, conducted on 1800 speech recordings to evaluate the recognition of ten isolated Amazigh-Tachelhit words in noisy conditions.

The MFCC based system achieved an overall accuracy of 61.94%, with Idgam (ⵉⴷⴳⴰⵎ) and Tfulki (ⵜⴼⵓⵍⴽⵉ) obtaining the highest rates (70.00% and 72.22%), while Achkid (ⴰⵛⴽⵉⴷ) and Imnsi (ⵉⵎⵏⵙⵉ) recorded the lowest (53.89% and 52.78%). The accuracy drop relative to calm conditions reached 26.62%. The DWT based system performed substantially better, achieving 75.33% overall accuracy; Imkli (ⵉⵎⴽⵍⵉ) and Idgam (ⵉⴷⴳⴰⵎ) reached the highest rates (78.89% and 80.56%), whereas Imnsi (ⵉⵎⵏⵙⵉ) and ɣassad (ⵖⴰⵙⵙⴰⴷ) were less accurately recognized (71.67% and 69.44%). The decrease in accuracy compared to calm conditions was 14.78%. The combined MFDWC method achieved the best performance, with an overall accuracy of 77.22%. Idgam (ⵉⴷⴳⴰⵎ) and Tfulki (ⵜⴼⵓⵍⴽⵉ) obtained the highest rates (82.22% and 83.33%), while ɣassad (ⵖⴰⵙⵙⴰⴷ) and Imnsi (ⵉⵎⵏⵙⵉ) showed the lowest (72.78% and 70.00%). The performance drop under noisy conditions was 17.45%. Overall, the MFDWC based system outperformed MFCC by 15.28% and DWT by 1.89%, confirming the robustness of the combined feature extraction strategy for isolated word recognition.

Table 10. Results of the seventh experiment.

Amazigh-Tachelhit words	Recognition rate (%)		
	MFCC	DWT	MFDWC
Azul	65.56	75.56	80.00
Tanmmirt	68.33	74.44	77.22
Tfulki	72.22	76.67	83.33
Idgam	70.00	80.56	82.22
ɣassad	55.56	69.44	72.78
Achkid	53.89	72.78	73.89
ɣilad	58.33	77.22	76.11
Imnsi	52.78	71.67	70.00
Imkli	66.67	78.89	75.00
Aghujjr	56.11	76.11	81.67
Recognition rate (%)	**61.94**	**75.33**	**77.22**

Tables 8, 9 and 10 clearly demonstrate that the system for recognizing the 10 isolated Amazigh-Tachelhit words outperforms the recognition systems for the 33 letters and 10 digits, which is also confirmed by Tables 4, 5, and 6 for the calm environment.

Experimental Results: Eighth Experiment. Table 11 summarizes the results of the eighth experiment, conducted on 9540 speech recordings combining 33 letters, 10 digits, and 10 isolated Amazigh words.

The MFCC based system achieved an overall accuracy of 51.43%, with a substantial drop of 29.86% between calm and noisy conditions. The DWT based system performed better, reaching 69.22% overall accuracy and a difference of 15.96% between environments. The combined MFDWC approach achieved the highest performance, with an overall accuracy of 73.55% and a difference of 16.82% between calm and noisy conditions. The MFDWC based system outperformed MFCC and DWT by 22.12% and 4.33%, respectively. Overall, these results confirm that the proposed MFDWC based system is the most effective and robust for Amazigh ASR across both calm and noisy environments, while the MFCC based system is notably more sensitive to noise.

Table 11. Results of the eighth experiment.

Overall data	Recognition rate (%)		
	MFCC	DWT	MFDWC
33 letters + 10 digits + 10 words	51.43	69.22	73.55

Although different speakers were used during the testing process compared to the training process, the system based on the proposed MFDWC method demonstrated strong performance. In a calm environment, it achieved an accuracy rate above 90.00%, while in a noisy environment, the accuracy remained higher than 73.00%.

Table 12 presents the results obtained by other systems in comparison with the proposed system.

Table 12. Results obtained by other systems and by the proposed system.

State of the art	Proposed system	Amazigh corpus	Number of speakers	Recognition rate (%)
[9]	CMU Sphinx based system	33 letters	34	90.00
[10]	CMU Sphinx based system	33 letters	9	88.00
[27]	MFCC and CNN based system	10 digits	42	91.75

(*continued*)

Table 12. (*continued*)

State of the art	Proposed system	Amazigh corpus	Number of speakers	Recognition rate (%)
[28]	CNN based system	18 isolated words	25	94.77
Current study	MFDWC based system	33 letters + 10 digits + 10 isolated words	120	90.37 (calm environment)
				73.55 (noisy environment)

4 Conclusion

This study proposes an Automatic Speech Recognition (ASR) system for the Amazigh language, built on an expanded database of 63600 recordings, including 33 letters, 10 digits, and 10 isolated words under quiet and noisy conditions. Preprocessing using ZCR and STE removed silent segments and compressed signals, improving efficiency. Features were extracted via a combination of MFCC and DWT (MFDWC), providing enhanced time-frequency representation and robustness to noise. Classification employed a One-versus-One SVM for effective multiclass separation. Experimental results demonstrate the system's superiority, achieving 90.37% accuracy in quiet conditions and 73.55% in noisy conditions, outperforming MFCC or DWT based methods and confirming the effectiveness of the proposed MFDWC architecture.

References

1. Abakarim, F., Abenaou, A.: Enhancing Amazigh speech recognition system with MFDWC-SVM. In: Gervasi, O., et al. (eds.) ICCSA 2023. LNCS, vol. 13956, pp. 471–488. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-36805-9_31
2. Pandit, P., Makwana, P., Bhatt, S.: Automatic speech recognition of continuous speech signal of Gujarati language using machine learning. In: Sahni, M., Merigó, J.M., Jha, B.K., Verma, R. (eds.) Mathematical Modeling, Computational Intelligence Techniques and Renewable Energy. AISC, vol. 1287, pp. 147–159. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-9953-8_13
3. Satori, H., Zealouk, O., Satori, K., ElHaoussi, F.: Voice comparison between smokers and non-smokers using HMM speech recognition system. Int. J. Speech Technol. **20**(4), 771–777 (2017). https://doi.org/10.1007/s10772-017-9442-0
4. El Ouahabi, S., Atounti, M., Bellouki, M.: Toward an automatic speech recognition system for Amazigh-Tarifit language. Int. J. Speech Technol. **22**(2), 421–432 (2019). https://doi.org/10.1007/s10772-019-09617-6
5. El Ouahabi, S., Atounti, M., Bellouki, M.: Optimal parameters selected for automatic recognition of spoken Amazigh digits and letters using Hidden Markov Model Toolkit. Int. J. Speech Technol. **23**(4), 861–871 (2020). https://doi.org/10.1007/s10772-020-09762-3

6. Telmem, M., Ghanou, Y.: The convolutional neural networks for Amazigh speech recognition system. Telkomnika (Telecommun. Comput. Electron. Control) **19**(2), 515–522 (2021). https://doi.org/10.12928/telkomnika.v19i2.16793
7. Elouahabi, S., Atounti, M., Bellouki, M.: Amazigh isolated-word speech recognition system using hidden Markov model toolkit (HTK). In: Proceedings of the 2016 International Conference on Information Technology for Organizations Development (IT4OD), pp. 1–7. IEEE (2016). https://doi.org/10.1109/IT4OD.2016.7479305
8. Satori, H., ElHaoussi, F.: Investigation Amazigh speech recognition using CMU tools. Int. J. Speech Technol. **17**(3), 235–243 (2014). https://doi.org/10.1007/s10772-014-9223-y
9. Telmem, M., Ghanou, Y.: Estimation of the optimal HMM parameters for Amazigh speech recognition system using CMU-Sphinx. Procedia Comput. Sci. **127**, 92–101 (2018). https://doi.org/10.1016/j.procs.2018.01.102
10. Telmem, M., Ghanou, Y.: Amazigh speech recognition system based on CMUSphinx. In: Ben Ahmed, M., Boudhir, A. (eds.) SCAMS 2017. LNNS, vol. 37, pp. 397–410. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74500-8_37
11. El Ghazi, A., Daoui, C., Idrissi, N., Fakir, M., Bouikhalene, B.: Système de reconnaissance automatique de la parole amazigh à base de la transcription en alphabet tifinagh. Revue Méditerranéenne des Télécommunications **1**(2), (2011)
12. Janse, P.V., Magre, S.B., Kurzekar, P.K., Deshmukh, R.R.: A comparative study between MFCC and DWT feature extraction technique. Int. J. Eng. Res. Technol. **3**(1), 3124–3127 (2014)
13. Malik, M., Malik, M.K., Mehmood, K., Makhdoom, I.: Automatic speech recognition: a survey. Multimed. Tools Appl. **80**(6), 9411–9457 (2021). https://doi.org/10.1007/s11042-020-10073-7
14. Cutajar, M., Gatt, E., Grech, I., Casha, O., Micallef, J.: Comparative study of automatic speech recognition techniques. In: Proceedings of the 2013 IET Signal Processing, vol. 7, no. 1, pp. 25–46 (2013). https://doi.org/10.1049/iet-spr.2012.0151
15. Labied, M., Belangour, A.: Automatic speech recognition features extraction techniques: a multi-criteria comparison. Int. J. Adv. Comput. Sci. Appl. **12**(8) (2021)
16. Tufekci, Z., Gowdy, J.N.: Feature extraction using discrete wavelet transform for speech recognition. In: Proceedings of the 2000 IEEE SoutheastCon. 'Preparing for The New Millennium' (Cat. No. 00CH37105), pp. 116–123. IEEE (2000). https://doi.org/10.1109/SECON.2000.845444
17. Shete, D. S., Patil, S. B., Patil, S.: Zero crossing rate and energy of the speech signal of Devanagari script. IOSR J. VLSI Signal Process. (IOSR-JVSP) **4**(1), 1–5 (2014)
18. Bachu, R.G., Kopparthi, S., Adapa, B., Barkana, B.D.: Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal. In: Proceedings of the 2008 American Society for Engineering Education (ASEE), pp. 1–7 (2008)
19. Ahad, A., Fayyaz, A., Mehmood, T.: Speech recognition using multilayer perceptron. In: Proceedings of IEEE Students Conference, pp. 103–109. IEEE (2002). https://doi.org/10.1109/ISCON.2002.1215948
20. Hennebert, J., Hasler, M., Dedieu, H.: Neural networks in speech recognition. Department of Electrical Engineering, Swiss Federal Institute of Technology (1994)
21. Alpaydin, E.: Introduction to Machine Learning. MIT Press, Cambridge (2020)
22. Lama, P., Namburu, M.: Speech recognition with dynamic time warping using MATLAB. CS 525, SPRING 2010-Project report (2010)
23. Majeed, S., Husain, H., Samad, S.A., Hussain, A.: Hierarchical k-means algorithm applied on isolated Malay digit speech recognition. In: Proceedings of International of Computer Science & Information Technology, pp. 33–37 (2012)

24. Padmanabhan, J., Johnson Premkumar, M.J.: Machine learning in automatic speech recognition: a survey. IETE Tech. Rev. **32**(4), 240–251 (2015). https://doi.org/10.1080/02564602.2015.1010611
25. Ghai, W., Singh, N.: Literature review on automatic speech recognition. Int. J. Comput. Appl. **41**(8), 42–50 (2012). https://doi.org/10.5120/5565-7646
26. Ali, H., Jianwei, A., Iqbal, K.: Automatic speech recognition of Urdu digits with optimal classification approach. Int. J. Comput. Appl. **118**(9), 1–5 (2015)
27. Boulal, H., Hamidi, M., Abarkan, M., Barkani, J.: Amazigh spoken digit recognition using a deep learning approach based on MFCC. Int. J. Electr. Comput. Eng. Syst. **14**(7), 791–798 (2023). https://doi.org/10.32985/ijeces.14.7.6
28. Daouad, M., Dadi, E.W.: Amazigh speech recognition using 1D CNN. In: Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security, pp. 1–6 (2023). https://doi.org/10.1145/3607720.3607760