10th International Conference on Control Engineering &Information Technology (CEIT-2025)
Proceedings Book Series –PBS- Vol 23 Vol 23, pp.64-70

Intelligent Control system for Adaptive Speech Recognition

Jalila Alferjany#1, Rabeia N Abd Alati Elfrjane *2, Khamisa A Yousef #3

 # 1Electrical and Electronic Department, University of Benghazi, Al-Marj Branch, Libya

2 Faculty of Education University of Benghazi/Libya

3 computer department, TOBRUK university, Libya

1 jaleela1993@gmail.com

2 rabeia.elfargane@uob.edu.ly

3 Khamisa.yousef@tu.edu.ly

Abstract— **Speech recognition has emerged as a sophisticated and effective modality for facilitating human-machine interaction. In this study, we introduce an innovative methodology for the recognition of isolated Arabic words by amalgamating feature extraction utilizing discrete wavelet transform (DWT), subtractive clustering for the initialization of rules, and an Adaptive Neuro-Fuzzy Inference System (ANFIS) for the classification process. The proposed framework was executed in MATLAB and assessed using a dataset obtained from various speakers. The experimental findings revealed that the proposed technique attains a recognition rate of approximately 94.5%, suggesting that the incorporation of intelligent control mechanisms can significantly enhance the efficacy of speech recognition systems..**

Keywords— ***Speech Recognition, Intelligent Control, ANFIS, Wavelet Transform, Subtractive Clustering*.**

I. INTRODUCTION

The swift advancement of voice-driven interfaces has amplified the demand for precise and effective speech recognition systems. Traditional methodologies—predominantly reliant on Fourier or linear predictive coding techniques—frequently encounter difficulties when addressing non-stationary attributes and overlapping speech phenomena. In order to mitigate these challenges, this research examines the implementation of intelligent control via an Adaptive Neuro-Fuzzy Inference System (ANFIS) in conjunction with wavelet-based feature extraction. The proposed methodology is specifically designed for the recognition of isolated Arabic lexemes, thereby tackling the issues related to acoustic variability and non-linearity inherent in speech signals.

II. RELATED WORK

J. Linke, S. Wepner, G. Kubin, and B. Schuppler[1] presented advance Automatic Speech Recognition (ASR) for conversational Austrian German by creating a Kaldi-based system that addresses the complexities of spontaneous speech, achieving a Word Error Rate (WER) of 0.4% for read speech and 48.5% for conversational speech, while underscoring the critical role of pronunciation modeling through a knowledge-based lexicon in low-resource environments, and providing cross-validation insights that highlight the impact of speaker variability on recognition accuracy, ultimately advocating for a synergistic approach between knowledge-based and data-driven strategies in ASR development.

S. Tamura, T. Hattori, Y. Kato and N. Noguchi [2] presented a novel method for developing a speech recognition system tailored for indigenous languages with limited resources is presented. To overcome the problem of little training data for minor languages, it makes use of text autoencoder techniques, neural machine translation (NMT), and a speech recognizer from a large language. In order to improve feature extraction and enable efficient representation even with limited resources, the authors concentrate on self-supervised learning (SSL). All things considered, the study shows how NLP methods and SSL can be combined to improve speech recognition for native languages.

Selina S. Sung et al.[3] presented system of an automated framework for the detection of speech sound disorders (SSDs) in pediatric populations through the application of deep learning methodologies. It underscores the multifaceted challenges encountered by children afflicted with SSDs, which encompass social

ISSN: 2961-6611

and educational impediments, as well as the intricate nature of diagnosing these disorders attributed to the variability in severity and the ambiguity of diagnostic criteria. The investigation employs a comprehensive dataset consisting of audio recordings from 573 children, meticulously assessed by speech-language pathologists, to accurately identify 92 children diagnosed with SSDs. A range of methodologies, encompassing automatic speech recognition (ASR) and audio classification (AC), are rigorously evaluated to enhance the precision of detection outcomes. The most effective approach yielded an unweighted average recall of 73.9%, thereby illustrating the promising capabilities of deep learning within this domain, whilst simultaneously acknowledging the necessity for continued inquiry to bolster the reliability of the models employed.

III. Methodology

This paper introduces the ANFIS intelligent control system for voice recognition. Using a corpus of Arabic speech, the proper learning algorithm is applied to individual words. The wavelet transform was utilized to extract features from the speech signal, and subtractive clustering was used to determine the best structure for the fuzzy inference system. The parameters of the network were then learned using hybrid learning, which combines gradient decent and least square estimation (LSE).

A. DataAcquisition and Preprocessing

A specialized speech database was created by recording separated Arabic digits and sentences from four speakers.To make sure the feature extraction procedure was reliable, the recorded audio signals were preprocessed to lower noise and normalize amplitude.

B. Feature Extraction

Feature extraction was conducted utilizing the discrete wavelet transform (DWT). The wavelet transform systematically decomposes the speech signal into multi-resolution sub-bands, thereby effectively encapsulating both temporal and frequency characteristics. In this investigation, the Daubechies 8-tap (db8) filter bank was chosen owing to its capacity to closely approximate the original signal and its remarkable efficacy in capturing the nuances of speech. Feature extraction was conducted utilizing the discrete wavelet transform (DWT). The wavelet transform systematically decomposes the speech signal into multi-resolution sub-bands, thereby effectively encapsulating both temporal and frequency characteristics. In this investigation, the Daubechies 8-tap (db8) filter bank was chosen owing to its capacity to closely approximate the original signal and its remarkable efficacy in capturing the nuances of speech.

C. Clustering and Rule Generation

In order to mitigate the intricacies associated with the fuzzy inference system (FIS), the methodology of subtractive clustering was employed on the dataset comprising wavelet coefficients. This singular-pass clustering algorithm effectively identifies the most advantageous cluster centers, which are subsequently utilized to formulate the fuzzy rules pertinent to the Adaptive Neuro-Fuzzy Inference System (ANFIS) model. Various parameters, including the range of influence, squash factor, acceptance ratio, and rejection ratio, were meticulously adjusted to achieve an optimal equilibrium between the number of clusters and the system's capacity for generalization.

D. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS integrates the human-like cognitive reasoning inherent in fuzzy logic with the adaptive learning proficiencies characteristic of neural networks. The preliminary Fuzzy Inference System (FIS), developed through the application of the subtractive clustering methodology, underwent further enhancement through a hybrid learning algorithm that synergistically merges gradient descent techniques with least squares estimation. The training regimen necessitated the iterative modification of the parameters associated with the membership functions to effectively minimize the discrepancy between the target outcomes and the outputs produced by the model.

ISSN: 2961-6611

IV. DATABASE USED IN THIS EXPERIMENTAL STUDY

The database used was created from 5 Arabic words taken from the Arabic speech corpus for isolated words. These 5 Arabic words are given in table 1.

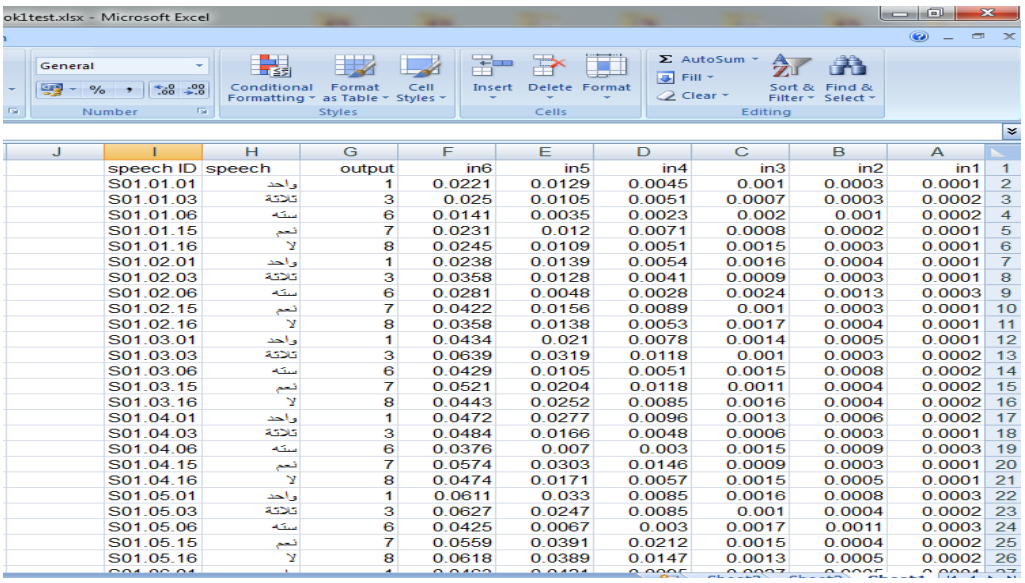
TABLE 1
ALL THE WORDS THAT HAVE BEEN INCLUDED IN THE CORPUS WITH THE
ENGLISH APPROXIMATION AND TRANSLATION

Arabic	**Translation**	**English Approximation**	**IPA**
واحد	One	Wahed	/ wa:hid /
ثلاثة	Three	Thlatha	/ θala: θh /
ستة	Six	Setah	/ sitat /
نعم	Yes	Naam	/ nʕm /
لا	No	Laa	/ la: /

following percentages:
Training subset = 50% (100)
Testing subset = 30% (60)
Checking subset = 20% (40)

The dataset utilized for training, testing, and validation was meticulously constructed through the application of wavelet transformation, with the specifics of wavelet decomposition delineated; each detail corresponds to a distinct frequency band, after which the standard deviation for each detail was computed. Utilizing a MATLAB program derived from the jal function detailed in the appendix, the standard deviation for every detail was systematically determined across all speech signals. Table 2 presents a selection of samples extracted from the curated database.

TABLE2
SOME OF SAMPLE OF THE DATABASE PREPARED



	I	H	G	F	E	D	C	B	A	
	speech ID	speech	output	in6	in5	in4	in3	in2	in1	1
	S01.01.01	واحد	1	0.0221	0.0129	0.0045	0.001	0.0003	0.0001	2
	S01.01.03	ثلاثة	3	0.025	0.0105	0.0051	0.0007	0.0003	0.0002	3
	S01.01.06	ستة	6	0.0141	0.0035	0.0023	0.002	0.001	0.0002	4
	S01.01.15	نعم	7	0.0231	0.012	0.0071	0.0008	0.0002	0.0001	5
	S01.01.16	لا	8	0.0245	0.0109	0.0051	0.0015	0.0003	0.0001	6
	S01.02.01	واحد	1	0.0238	0.0139	0.0054	0.0016	0.0004	0.0001	7
	S01.02.03	ثلاثة	3	0.0358	0.0128	0.0041	0.0009	0.0003	0.0001	8
	S01.02.06	ستة	6	0.0281	0.0048	0.0028	0.0024	0.0013	0.0003	9
	S01.02.15	نعم	7	0.0422	0.0156	0.0089	0.001	0.0003	0.0001	10
	S01.02.16	لا	8	0.0358	0.0138	0.0053	0.0017	0.0004	0.0001	11
	S01.03.01	واحد	1	0.0434	0.021	0.0078	0.0014	0.0005	0.0001	12
	S01.03.03	ثلاثة	3	0.0639	0.0319	0.0118	0.001	0.0003	0.0002	13
	S01.03.06	ستة	6	0.0429	0.0105	0.0051	0.0015	0.0008	0.0002	14
	S01.03.15	نعم	7	0.0521	0.0204	0.0118	0.0011	0.0004	0.0002	15
	S01.03.16	لا	8	0.0443	0.0252	0.0085	0.0016	0.0004	0.0002	16
	S01.04.01	واحد	1	0.0472	0.0277	0.0096	0.0013	0.0006	0.0002	17
	S01.04.03	ثلاثة	3	0.0484	0.0166	0.0048	0.0006	0.0003	0.0001	18
	S01.04.06	ستة	6	0.0376	0.007	0.003	0.0015	0.0009	0.0003	19
	S01.04.15	نعم	7	0.0574	0.0303	0.0146	0.0009	0.0003	0.0001	20
	S01.04.16	لا	8	0.0474	0.0171	0.0057	0.0015	0.0005	0.0001	21
	S01.05.01	واحد	1	0.0611	0.033	0.0085	0.0016	0.0008	0.0003	22
	S01.05.03	ثلاثة	3	0.0627	0.0247	0.0085	0.001	0.0004	0.0002	23
	S01.05.06	ستة	6	0.0425	0.0067	0.003	0.0017	0.0011	0.0003	24
	S01.05.15	نعم	7	0.0559	0.0391	0.0212	0.0015	0.0004	0.0002	25
	S01.05.16	لا	8	0.0618	0.0389	0.0147	0.0013	0.0005	0.0002	26

The ANFIS architecture which has six input standard deviations of all sample training and one output. The structure of the ANFIS model is shown in Fig.1.

ISSN: 2961-6611

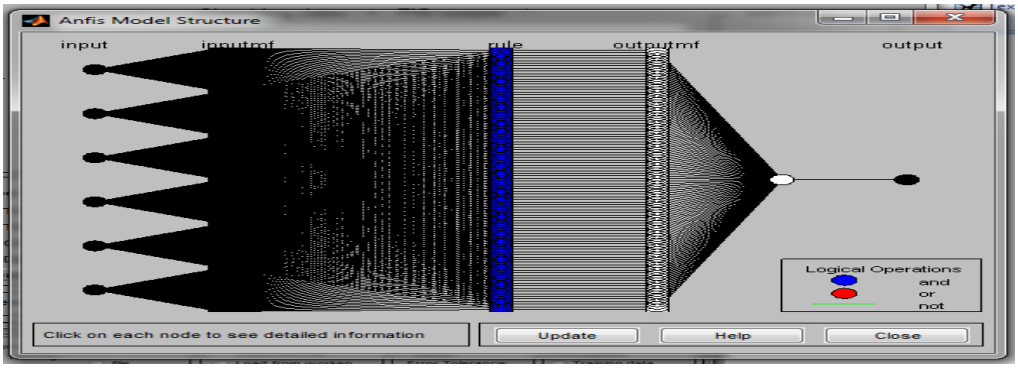


Fig. 1 The ANFIS architecture

A typical input–output (input standard 1 (in 1), input standard 2 (in 2) and output of ANFIS) surface of the training phase of samples is plotted in Fig.2.

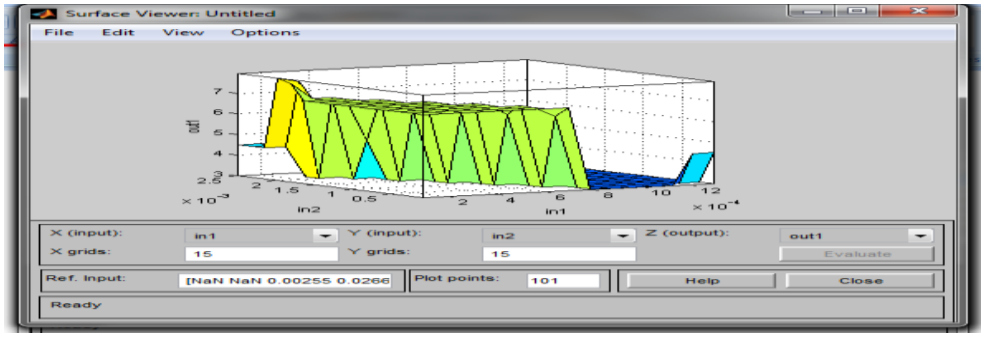


Fig. 2 Overall input–output surface of standard deviation input1,

standard deviation input2 and the output of ANFIS

The system had 100 fuzzy rules in the rule layer; the structure of the rule is shown in Fig.3.

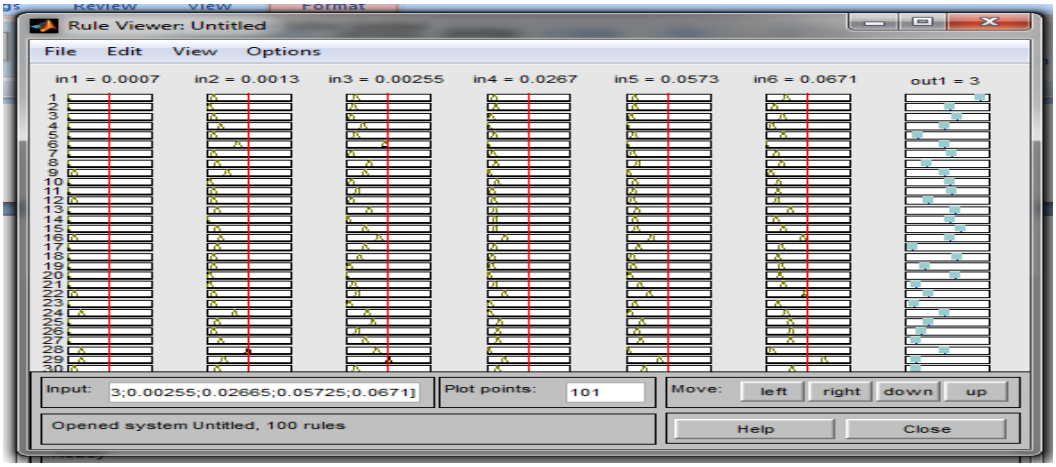


Fig. 3 Rule structure

ISSN: 2961-6611

V. EXPERIMENTAL RESULTS

The proposed system underwent a rigorous evaluation employing a comprehensive array of training, testing, and validation datasets. The conducted experiments demonstrated that the amalgamation of wavelet-based feature extraction techniques alongside Adaptive Neuro-Fuzzy Inference System (ANFIS) facilitates a substantial recognition rate, with the system attaining an approximate accuracy of 94.5% in the identification of isolated Arabic lexemes. A meticulous examination of the error trajectories and rule configurations substantiated the resilience of the methodology, particularly in addressing the non-linear characteristics that are intrinsic to speech signals.

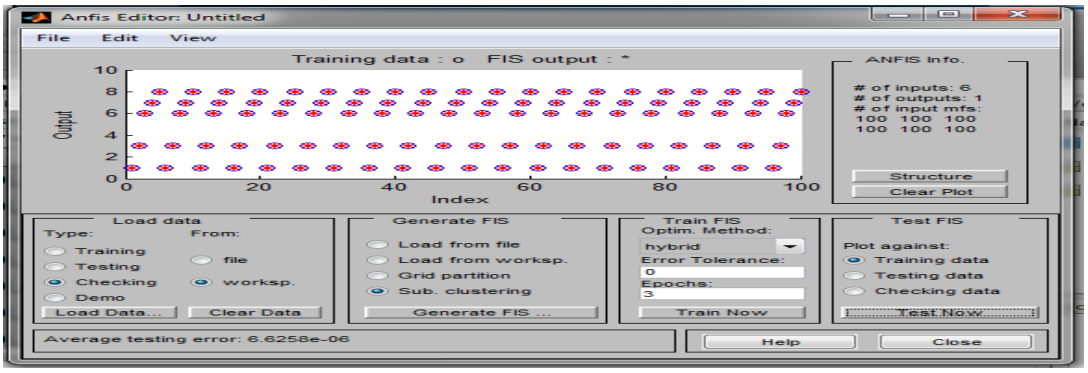


Fig.4 Testing the FIS with Training data set

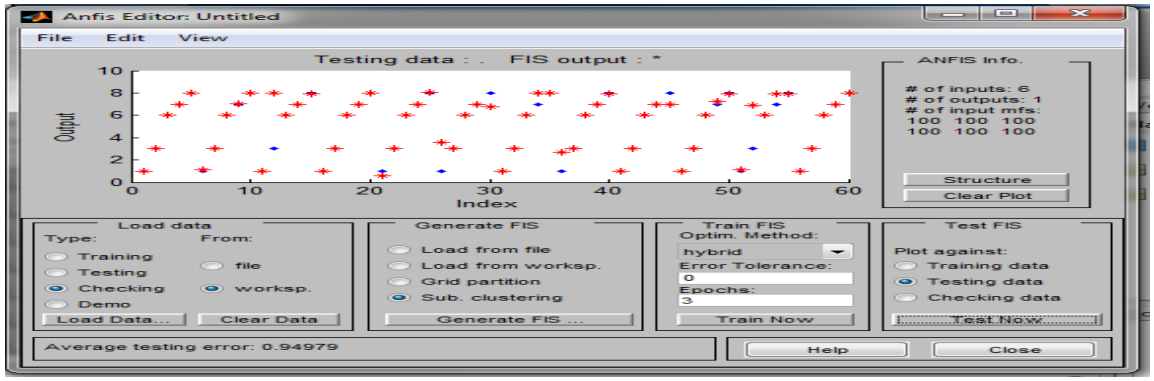


Fig.5 Testing process of the ANFIS model

The average error for testing the training, testing and checking data against the trained FIS was shown in table 3

TABLE2
SOME OF SAMPLE OF THE DATABASE PREPARED

The average error for test data against train FIS	**Value**
The average error for test training data against the trained FIS	6.6258e-6
The average error for test testing data against the trained FIS	0.94979
The average error for test checking data against the trained FIS	0.73542

VI. CONCLUSION

ISSN: 2961-6611

This investigation elucidates a proficient methodology for enhancing speech recognition efficacy through the amalgamation of advanced control strategies. The implementation of discrete wavelet transform for feature extraction, in conjunction with subtractive clustering and Adaptive Neuro-Fuzzy Inference System (ANFIS) for classification, culminates in a resilient recognition system tailored for isolated Arabic lexemes. The noteworthy recognition rate of 94.5% accentuates the promise inherent in this hybrid methodology. It is recommended that subsequent inquiries aim to broaden the applicability of this technique to encompass a wider array of speech applications and to optimize the system's performance in real-world environments characterized by auditory noise.

VII. DISCUSSION

The empirical evidence indicates that the sophisticated control mechanism employed by ANFIS is instrumental in augmenting the efficacy of speech recognition systems. Through the adept amalgamation of feature extraction utilizing wavelet transformation and adaptive learning derived from fuzzy inference mechanisms, the proposed methodology is capable of encapsulating the inherent variability of speech while preserving computational efficiency. Prospective research may investigate the potential for expanding this paradigm to encompass continuous speech recognition and the integration of supplementary robust feature extraction methodologies.

REFERENCES

[1] J. Linke, S. Wepner, G. Kubin, and B. Schuppler, “Using Kaldifor automatic speech recognition of conversational Austrian Ger-man,” arXiv, p. doi: 10.48550/ARXIV.2301.06475, 2023.

[2] S.Tamura, T. Hattori, Y. Kato and N. Noguchi"Speech Recognition for Indigenous Language Using Self-Supervised Learning and Natural Language Processing"In Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024), pages 779-784

[3] Selina S. Sung,JungminSo,Tae-Jin Yoon, and Seunghee Ha "Automatic detection of speech sound disorder in children using automatic speech recognition and audio classification", phonetics and Speech Sciences Vol.16 No.3 (2024) 87-94

[4] Naresh, P.V., & Visalakshi R.(2019). FEATURE EXTRACTION TECHNIQUES INSPEECH RECOGNITION: A REVIEW. International Journal ofManagement, Technology And Engineering, Volume IX, Issue II. Retrievedfrom http://ijamtes.org/gallery/123-feb2019.pdf..

[5] Rady, E. R., Yahia, A. H., El-Dahshan, E. A., & El-Borey, H. (2013). Speechrecognition system based on wavelet transform and artificial neuralnetwork. Egyptian Computer Science Journal (ECS), 37(3).

[6]] JAGTAP, P. (2014). Neuro-Fuzzy Systems for Modelling and ControlApplications (Doctoral dissertation, INDIAN INSTITUTE OF TECHNOLOGYROORKEE)

ISSN: 2961-6611