SPEECH RECOGNITION BY IMPROVING THE PERFORMANCE OF ALGORITHMS USED IN DISCRIMINATION

Alahmar -Haeder Talib Mahde

ALIraqia University College of Engineering, Iraq

ABSTRACT

Speech recognition techniques are one of the most important modern technologies. Many different systems have been developed in terms of methods used in the extraction of features and methods of classification. Voice recognition includes two areas: speech recognition and speaker recognition, where the research is confined to the field of speech recognition.

The research presents a proposal to improve the performance of single word recognition systems by an algorithm that combines more than one of the techniques used in character extraction and modulation of the neural network to study the effects of recognition science and study the effect of noise on the proposed system.

In this research four systems of speech recognition were studied, the first system adopted the MFCC algorithm to extract the features. The second system adopted the PLP algorithm, while the third system was based on combining the two previous algorithms in addition to the zero-passing rate. In the fourth system, the neural network used in the differentiation process was modified and the error ratio was determined. The impact of noise on these previous systems.

The outcomes were looked at regarding the rate of recognizable proof and the season of preparing the neural network for every system independently, to get a rate of distinguishing proof and quiet up to 98% utilizing the proposed framework.

KEYWORDS

Speech Recognition, PLP, MFCC, Artificial Neural Networks (ANN).

1. INTRODUCTION

In the last four decades, computer experts and researchers have become more interested in speech recognition, in order to get to the stage of making the machine able to understand human speech and to receive orders and instructions in a voice-free manner without the need for traditional means of input in order to save a good time.

A number of research and studies have been conducted in the area of speech recognition during the past years, in the year 2015 particularly, an algorithm was designed to provide several systems for extracting features and then adopting /4 / words in a process called (Net Document), Shutdown, and restart and recorded with 33 different people in different ages[1].

In this research, MFCC, PLP, and LPCC algorithms were used to derive attributes and HMM as a class. The study reached the higher recognition rate (44.39/) when integrating the MFCC and PLP

DOI: 10.5121/ijcsit.2019.11102

systems, but at the expense of increasing identification time and training time. [2] Inge Gavant and Diana Mulitaru (2015) studied different workbooks and correlated with the vocabulary size used in the recognition process.[2]

PLP and MFCC algorithms were used to derive features. The results of this study showed that neuronal networks outperform the hidden Markov model by 10%. Neural networks have become the standard classification used in speech recognition since 2011.

In the year 2015, Veton Z. Këpuska and Hussien A. Elharati studied different algorithms in extracting attributes where 31 / attributes were extracted using each algorithm and merging the attributes to become the total number of attributes in each recognition / 93 / attribute.

The number of training samples / 2702 and number of test samples / 6842 /, where the system was tested in different noise conditions. The study found that the PLP algorithm With delta and delta samples are the best in high noise conditions up to /%25.39/ and speech recognition without the presence of speech without noise was LPCC algorithm is the best known by /%59.99/ [4].
The results of this experiment indicate that the integration of features did not produce good results compared to individual algorithms and were even poor in medium and high noise conditions.

In (2016), Poonam Sharma and Angali Garg conducted a study showing the rate of recognition of the Mandarin language, the influence of gender, the type of neural network, and the attribute extraction algorithm on the recognition ratio.

In the study, the female recognition rate was more than 2%. The combination of PLP with MFCC improved the recognition rate by 19% compared with PLP and 1% compared with MFCC.In 2017, Yousra Faisal Al-Rahim and Lajeen Abdel-Gadir conducted a speech discrimination study using the MFCC, PLP and RASRA-PLP algorithms, where 4 males and 5 females were recorded their ages between (30-16). The training samples included 864 training samples The mathematical model of the radiographic quantification was classified as modified with its parameters where the MFCC algorithm was the best known as /%5.89/ [5].

Previous studies have shown that the process of integrating features in one way or another may lead to better results in the identification process. It is preposterous to expect to decide the presence of a superior algorithm in light of the fact that the exactness of the recognizable proof is identified with the idea of the language, the sex of the speaker, age and numerous different components. And even the parameters of the workbook greatly affect the identification process. Therefore, the research results can't be adopted as comprehensive results because of the very wide variety of recognition conditions and their determinants.

Based on previous studies, the neural network was selected as a classifier, and the integration of the extracted features using PLP and MFCC was chosen for the overall performance.Figure (1) shows the general box diagram for suggested system and the merger process.

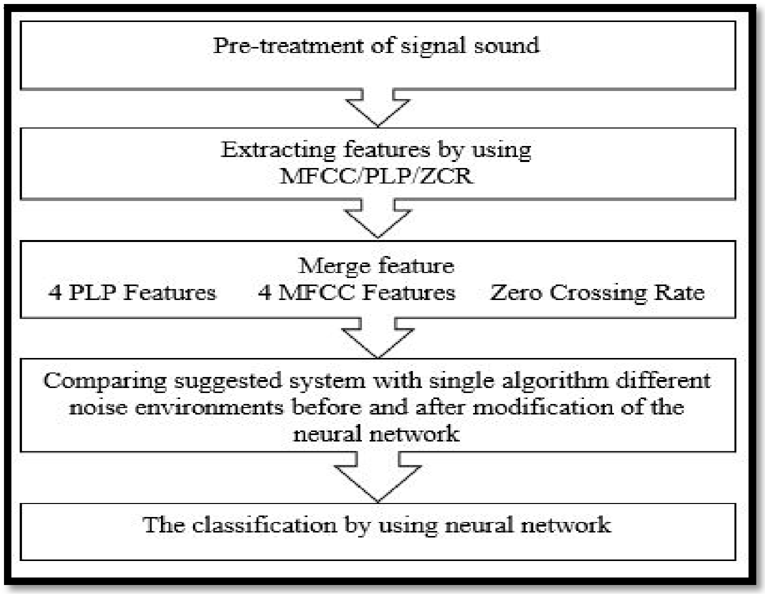

Pre-treatment of signal sound
Extracting features by using MFCC/PLP/ZCR
Merge feature
4 PLP Features
4 MFCC Features
Zero Crossing Rate
Comparing suggested system with single algorithm different noise environments before and after modification of the neural network
The classification by using neural network

Figure (1) shows the general box diagram for suggested system

2. THE IMPORTANCE OF THE RESEARCH AND ITS OBJECTIVES

The applications used for a voice interaction still suffer from a reduction in the accuracy of speech recognition, so the need for high- knowledge and small-time applications has emerged at the same time to suit the needs of different users.

The purpose of this research is to develop an application that integrates different techniques to extract features from speech before passing on the classifier, in order to study the effect of the integration on the accuracy of recognition and obtain the highest accuracy possible and in the shortest possible time.

3. RESEARCH METHODOLOGY

This study was conducted by using two methods:

Descriptive method where the effect of the type of neural network used in speech recognition on the recognition ratio and the time of training (neural network) was studied. In addition, the effect of integrating the features using different algorithms on the recognition ratio and the experimental method was studied by combining the PLP and MFCC algorithms, The rate of zero skipping at the level of features, measurement of the recognition ratio and the effect of noise on this ratio, then measure the recognition ratio when trying the proposed merge algorithm on different users.

Note: All experiments were conducted in a home room using a low-quality microphone.

4. RESULTS AND DISCUSSION

4.1. ARTIFICIAL NEURAL NETWORK

Neural networks with frontal feed are one of the most modern and highly efficient methods of giving satisfactory and good results in recognition. An important part of neuronal network construction is the use of a precise and powerful learning algorithm. The neural networks with frontal feed are one of the most modern methods of high efficiency in producing satisfactory and good results in recognition. An important part of the neural network construction is the use of a precise and powerful learning algorithm. The most common is the neural network with frontal feeding and back-error spreading.

The direction of signals entering the network is always forward, so that the signal coming out of any neuron depends only on the incoming signals. The neural networks with the frontal feed need to have two pairs of vectors: the input and output desire vectors, the training process begins with the input vector where it is applied to the network and produces the real output, compared with the corresponding expected vector and the difference between them represents the error that is used to adjust the weights according to the education algorithm, and the training continues until the error reaches as low as possible [6].

4.2. FEATURES EXTRACTION

The process in which the data is transferred is called the extraction process, and the methods used to extract the features have varied. In this research, we adopted the MFCC algorithm and the PLP algorithm at the feature extraction phase, Acoustic output and a beam of features represented by the frequency resonance frequencies of the audio stream extending from vocal cords to the lips.

4.2.1 MFCC ALGORITHM

The MFCC algorithm is one of the most common methods used in character extraction because of the sensitivity of its filters to human voice signal properties. MFCC is used extensively in speech recognition. It provided mile frequency coefficients in 1980 and has been a pioneer in this area since then.

Human-generated sounds are filtered according to vocal tract mode, so if the shape of the vocal track can be accurately determined, the phoneme produced can be determined. The audio track form in the short time power spectrum the purpose of the algorithm (MFCC) is to accurately represent this envelope.

The MFCC algorithm is based on changes in the human ear frequency bandwidth and is used to capture key speech characteristics. MFCC has linear linearity at frequencies below 1,000 Hz and logarithmic at frequencies higher than 1000 Hz. Figure (2) shows steps the MFCC algorithm works.

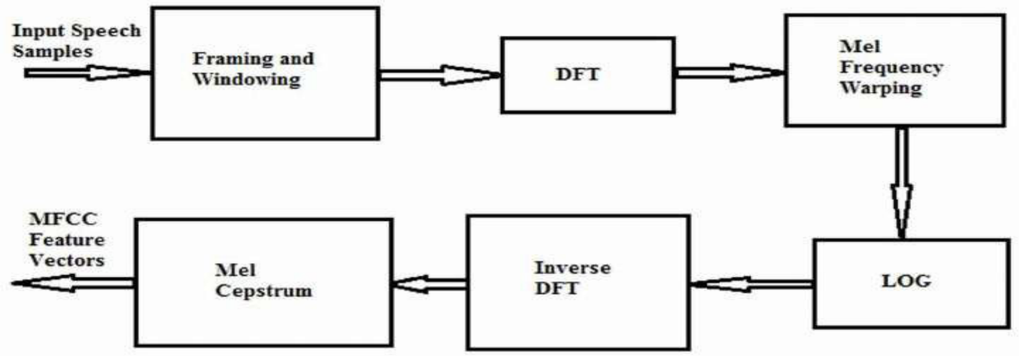


Figure (2) steps the MFCC algorithm works

4.2.2. PLP ALGORITHM

This technique relies on psychophysics of hearing, the relationship between a physical effect and influential perceptions where it excludes information that is not related to speech, thus improving recognition.

The PLP algorithm is similar to the LPCC algorithm but the spectral properties have been converted to match the characteristics of the human audio system. It is close to three main aspects:

1. The critical band resolution curve.
2. The equal loudness curve.
3. The intensity – loudness power – low relation.

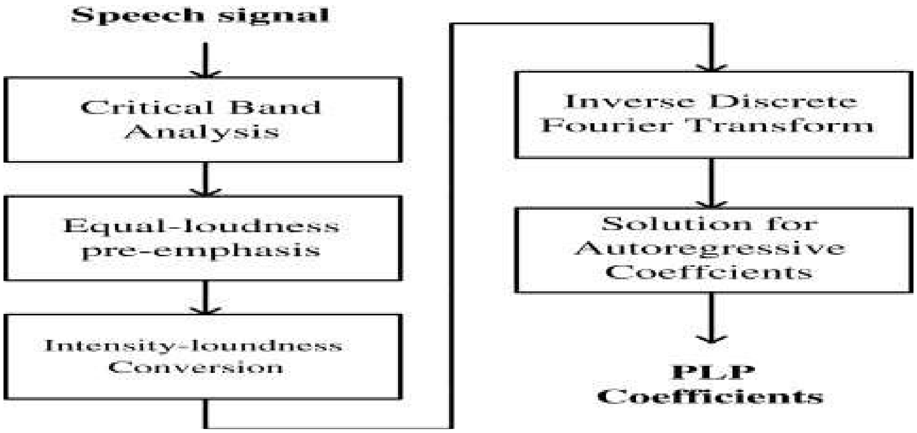

Speech signal
Critical Band Analysis
Equal-loudness pre-emphasis
Intensity-loundness Conversion
Inverse Discrete Fourier Transform
Solution for Autoregressive Coeffcients
PLP Coefficients

Figure (3) PLP algorithm for making the PLP

In Figure 4 we compare the Mel field used in the MFCC algorithm and the Bark field used in the .PLP algorithm{4}

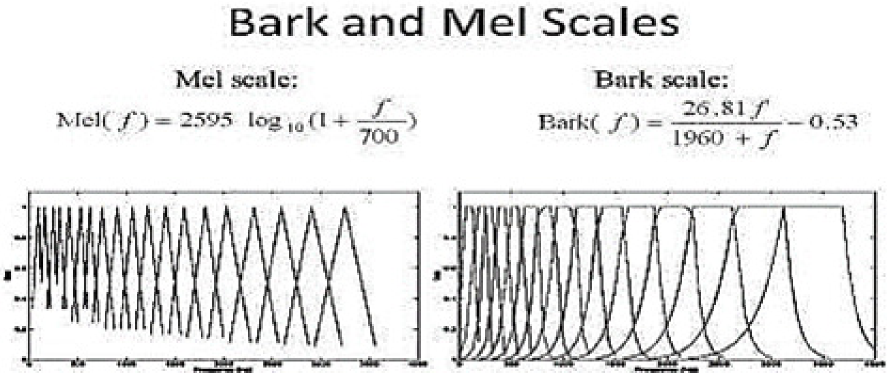


Figure (4): Comparison of Park Range and Mill Scope [01]

5. Practical Section

In this section we will build the neural network at first, and then use it in the process of identifying and comparing the results before and after the change, and studying the effect of noise on the recognition ratio.

5.1 Building The Proposed Neural Network

Most of the previous research in this area used the back tracking network and continued the training (train scg.) and the error rate is 0.00001 and its code in the matlab environment is the following code:

```
net=newff(minmax(p),[1000,200],{'tansig','logsig'},'trainscg');

net.trainParam.show=50;

net.trainParam.lr=0.035;

net.trainParam.epochs=1000;

net.trainParam.goal=0.00001;
```

The training of this network took 4 minutes 48 seconds, a relatively long time.

Therefore, the following network was used based on a detailed experimental study where the accuracy of the doubling was achieved in only 5 seconds by changing the training to train lm and the error rate to 0.00000001. Number of selected neurons / 5 / neurons to use the rounding process to identify output, the following code illustrates the network in a matlab environment:

```
net=newff(minmax(p),[20,15,5],{'tansig','tansig','purelin'},'trainlm');

%net.trainParam.mem_reduc=2;

net.trainParam.show=50;

net.trainParam.lr=0.07;

net.trainParam.epochs=50;

net.trainParam.goal=0.00000001;
```

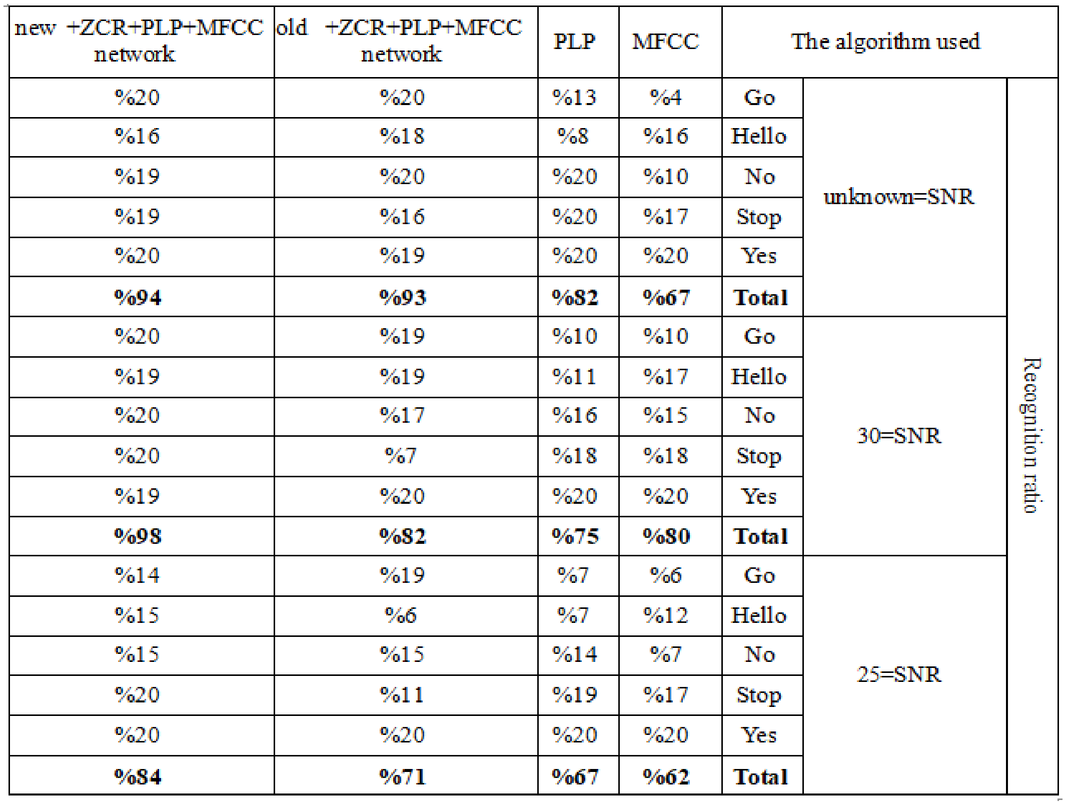
5.2. Work Scenarios

In this study, we used 5 words in the recognition process (Go, Hello, No, Stop, Yes). Experiments were conducted in a room with a medium quality microphone at a record frequency of 11025 hz and the recording time was 3 seconds.

The audio database consists of only 150 voice samples and /12 / features were adopted in the first system using MFCC and / 12 / attributes in the second system using the PLP algorithm. In the system based on the proposed merger, /4 / attributes were merged using The other two algorithms are the zero-pass rate. The number of attributes is/ 9 / which is less than three attributes of the conjugate algorithms./ 9 / attributes have been chosen instead of/ 11 / to obtain the account sources, thus the memory used in the processing and the identification and training time.

In the first scenario, the recognition process was done without adding noise or making any adjustments to the recorded sounds. In the second scenario, the recognition process was done after noise was added to the signal to become the noise ratio of the signal (snr = 30). In the third scenario, noise was added to the signal, 25. This table illustrates the effects of previous cases.

Table (1): Recognition ratios in all scenarios



new +ZCR+PLP+MFCC network	old +ZCR+PLP+MFCC network	PLP	MFCC	The algorithm used		
%20	%20	%13	%4	Go	unknown=SNR	Recognition ratio
%16	%18	%8	%16	Hello		
%19	%20	%20	%10	No		
%19	%16	%20	%17	Stop		
%20	%19	%20	%20	Yes		
%94	**%93**	**%82**	**%67**	**Total**		
%20	%19	%10	%10	Go	30=SNR	
%19	%19	%11	%17	Hello		
%20	%17	%16	%15	No		
%20	%7	%18	%18	Stop		
%19	%20	%20	%20	Yes		
%98	**%82**	**%75**	**%80**	**Total**		
%14	%19	%7	%6	Go	25=SNR	
%15	%6	%7	%12	Hello		
%15	%15	%14	%7	No		
%20	%11	%19	%17	Stop		
%20	%20	%20	%20	Yes		
%84	**%71**	**%67**	**%62**	**Total**		

Note that the highest recognition rate was in the signal to noise ratio (SNR = 30) in the system based on the proposed integration algorithm and this ratio was 98%.

From the previous table we see the marked improvement in the recognition ratio in the two systems, which relied on the integration of features compared to the single systems, although the number of features used is 9 / attributes in the third and fourth systems, while 12 / / Attribute using PLP in the second system.

We also notice that the error rate increases when we recognize the words GO and NO because of the similarity between them both are composed of a static character followed by a similar sound character and the same length.

The integration process has given a better rate of recognition in different noise situations, so this system is more stable. The diagram shown in Figure (5) shows the comparison of the previous algorithms with the difference in the noise on the signal to indicate clearly the superiority of the system designed on previous systems in all noise conditions.

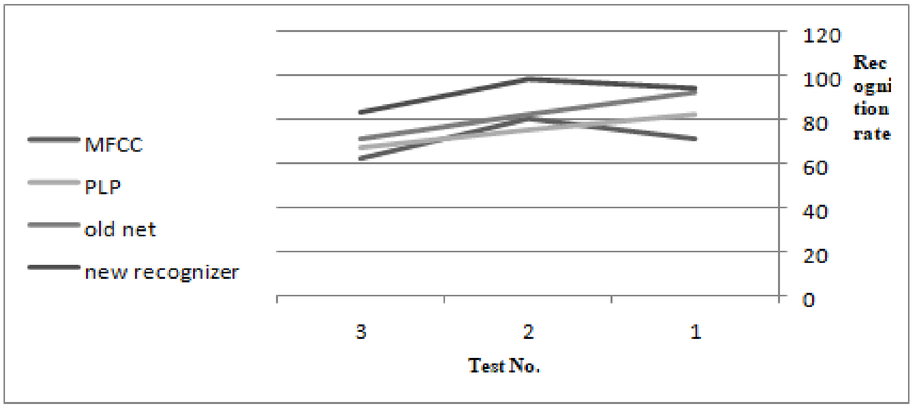


Figure (5): Comparison of the four algorithms studied

In the fourth scenario we modified the database to / 500 / audio file with different people's voices to get the following recognition ratio shown in Figure (6).

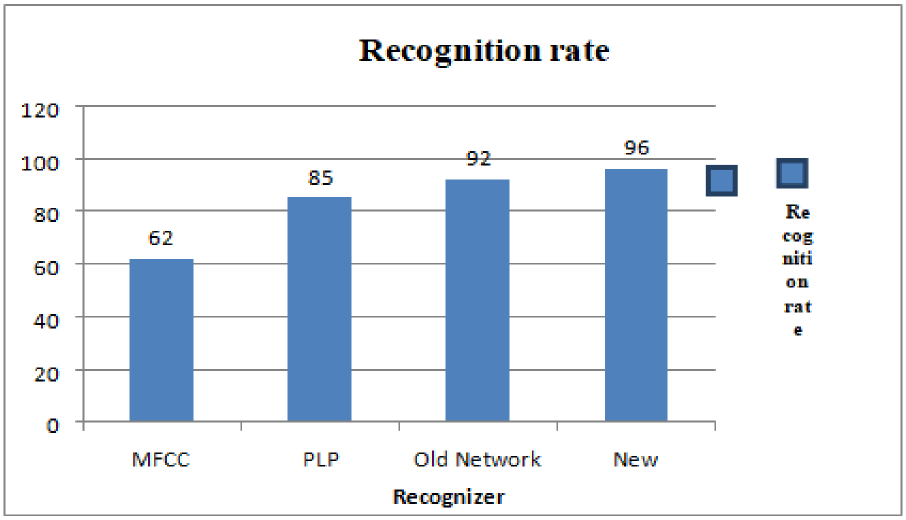
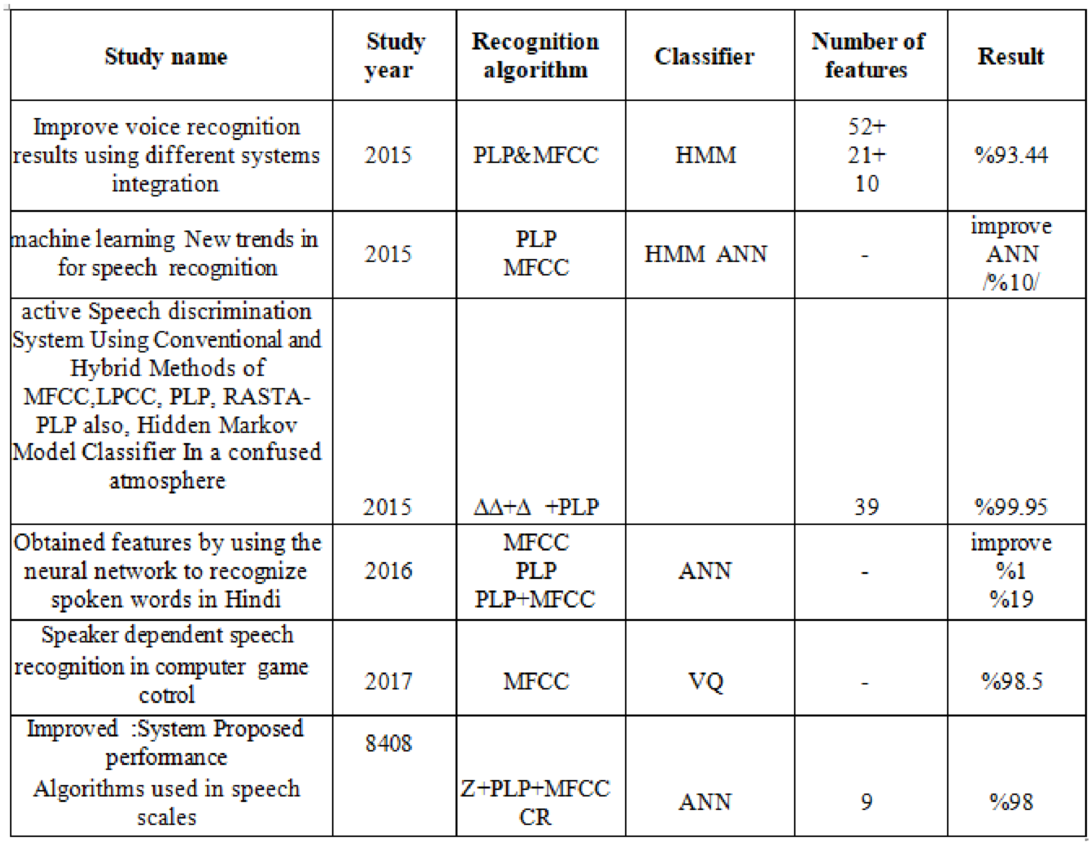


Figure (6): Identification ratio after increasing training samples to 500 samples

The work scenarios were selected in this paper by the researcher, but intersect with previous research in some points. The following table demonstrates The distinction the relationship between the proposed system and the past research.

Table (2): Comparison of the proposed system with the reference studies



Study name	Study year	Recognition algorithm	Classifier	Number of features	Result
Improve voice recognition results using different systems integration	2015	PLP&MFCC	HMM	52+ 21+ 10	%93.44
machine learning New trends in for speech recognition	2015	PLP MFCC	HMM ANN	-	improve ANN /%10/
active Speech discrimination System Using Conventional and Hybrid Methods of MFCC,LPCC, PLP, RASTA-PLP also, Hidden Markov Model Classifier In a confused atmosphere	2015	ΔΔ+Δ +PLP		39	%99.95
Obtained features by using the neural network to recognize spoken words in Hindi	2016	MFCC PLP PLP+MFCC	ANN	-	improve %1 %19
Speaker dependent speech recognition in computer game cotrol	2017	MFCC	VQ	-	%98.5
Improved :System Proposed performance Algorithms used in speech scales	8408	Z+PLP+MFCC CR	ANN	9	%98

6. Conclusions And Recommendations

A. The magnitude of the preparation database (sound examples) and the extent of closeness (disarray) between the words assumes an imperative job in the ID rate.The more the training samples increase the recognition rate and the greater the confusion, the lower the percentage.

B. There is a disparity between the rate of progress in the rate of acknowledgment and the quantity of highlights after some time as the expansion in the quantity of highlights improves the acknowledgment rate however to the detriment of expanding the ID time. It is best to design the appropriate neuronal network for each dataset as the network performance may be poor if the audio database is changed.

C. The learning algorithm of the neural network greatly affects training time.

D. The error rate of the neural network greatly affects the rate of recognition.

E. The process of integrating features in low noise conditions increases the recognition rate significantly from 12 to 23 percent and even in average noise conditions gives better results than PLP and MFCC algorithms from 16 to 21 percent. We conclude that the integration process gives greater stability in the recognition process under different conditions.

7. RECOMMENDATIONS

A. Do not increase the number of features from 12 attributes and may have a negatively affect the identification ratio.

B. Standardization of the parameters used in different systems such as frequency, the ratio of the reference to the sound and the recording environment to make the comparison process accurate and feasible.

C. Work on noise removal in more effective ways.

REFERENCES

[1] Rama Ghassan Hassan,(2015)"Improving the results of voice recognition based on the results of the integration of different systems", Tishreen University, Vol. 85 No85.

[2] INGE GAVAT, DIANA MILITARU,(2015)"New trends in machine learning in speech recognition" , SISOM Bucharest, pp 276.

[3] POONAM SHARMA, ANGALI GARG,(2016)"Feature Extraction and Recognition of Hindi Spoken Words using Neural Networks", International Journal of Computer Applications (0975 – 8887) Volume 142 – No.7, pp 17.

[4] VETON Z. KËPUSKA, HUSSIEN A. ELHARATI,(2015)"Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC,LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions", Journal of Computer and Communications, , pp 1-9,

[5] YUSRA FAISAL AL-IRAHYIM, LUJAIN YOUNIS ABDULKADER,(2017)"Speaker Dependent Speech Recognition in Computer Game Control", International Journal of Computer Applications (0975–8887)Volume 158–No 4 , , pp 37.

[6] DIAMANTARAS K. AND KUNG S,(2006)"Principle Component Neural Networks Theory and Applications", New York, John Wiley & Sons Inc, , , pp255.

[7] LAVNEET SINGH, GIRIJA CHETTY,(2012)"A Comparative Study of Recognition of Speech Using Improved MFCC Algorithms and Rasta Filters", Information Systems, Technology and Management Communications in Computer and Information Science Volume 285, , pp 304-.413

[8] BHAVNA SHARMA, K. VENUGOPALAN,(2014)"Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images". IOSR Journal of Computer Engineering, Volume 16, Issue 1, pp 35.

[9] PITZ, M, SCHLUTER R, NEY H, MOLAU S,(2001)"Computing Mel-frequency cepstral coefficients on the power spectrum", Print ISBN: 0-7803-7041-4 INSPEC Accession Number: 7120280 Acoustics, Speech, and Signal Processing, 2001, Proceedings. (ICASSP '01). IEEE International Conference on (Volume: 1) Page 73 - 76 vol.1, pp12.

[10] H. HERMANSKY,(1989)"Perceptual linear predictive (PLP) analysis of speech", Speech Technology Laboratory, Division of Panasonic Technologies, Inc. 3888 State Street, Santa Barbara, California 93105., pp 1752.

[11] H. DEMUTH, M. BEALE,(2002)"Neural Networks Toolbox User's Guide". The MathWorks, Inc. pp 826.

[12] P. RANI, S. KAKKAR, S. RANI,(2015)“Speech recognition using neural networks”. International conference on advancement in engineering and technology., pp 14.

[13] N. DAVE,(2013)“Features Extraction Methods LPC, PLP and MFCC in Speech Recognition”, International Journal for Advanced Research in Engineering and Technology. Vol.1, Issue VI, July, pp5.

[14] BHUSHAN C. KAMLE,(2016)“Speech recognition using artificial neural networks”, Int’l journal of Computing, Communication & Instrumentation Engg, (IJCCIE), Vol 3, Issue 1, pp 4.

[15] A. MANSOUR, G. SALH, H. Z. ALABDEN,(2015)“Speech recognition using back propagation algorithm in neural network”, International Journal of Computer Trends and Technology(IJCTT), Vol 23,Number 3, pp21.

AUTHOR

Haeder Talib Mahdi Al-Ahmar was born in AL-Musayyib, Iraq, on July 8, 1975. He received his bachelor's degree (Hons) in Computer Science In the year 1999. From the University of Sabha in Libya and received a Master's Degree in Information Technology (I.T) Very good grade in the year 2001. As well as from the University of Sabha from Libya received a PhD in Information Technology (I.T) (Image Processing) from the University of Neelain in Sudan In the year 2018. Grade (excellence). I have working in university academic teaching since 2002 at the Iraqi University College of Engineering, Department of Computer Engineering till now