# Artificial Neural Network Algorithm Based on Speech Recognition System To Improve The Utterance Rate Of Speech In Neural Network

**[1] V. Thiyagu**
Research Scholar, Department of Electronics and Communication Engineering, Vinayaka Mission's Research Foundation (Deemed to be University), Salem, Tamil Nadu, India.

**[2,]T.Sheela**
Associate Professor, Electronics and Communication Engineering Department,Vinayaka Mission's KirupanandaVariyar Engineering College, Vinayaka Mission's Research Foundation
(Deemed to be University), Salem, Tamil Nadu, India

**Abstract**:

The investigation to which recognition systems can train the model parameters to produce the best class of discrimination will primarily determine their ability to correctly identify speakers based on their speech waveform distribution. This report details the outcomes of an effort to identify each speaker's voice using their distributed continuous voice waveform employing the coupled artificial neural network frame works. For discriminative classification and training, a feed-forward multi-layer ANN structure with 30 hidden neurons was used. This model created scores that were moved to best match the speech features. The decision system uses coefficient of correlation analysis to assess how well speech features match known speakers. frames from the ANN structures that describe the detected speaker. Investigations were performed out using spoken utterances from 30 distinct speakers to verify the system's performance (7 males and 3 females).For cases of trained voice utterances, system performance demonstrated average recognition rates of 95 percent for 1-word utterances and 5 percent when the length of the utterances was raised to 1 words. For 1-word utterances with unknown speakers, a recognition rate of 98% was reached.

**Keywords:** artificial neural networks, voice recognition, coefficients, utterance.

## 1.Introduction

The challenge in voice recognition is transforming the information content of speakers' speech waveform into recognizable sets of features that contain all the relevant discriminating information required for speaker recognition. The capacity of a recognition system to accurately distinguish speakers' voices mostly depends on how well the time frequency and energy of the speech waveform are captured, and also how well the recognition parameters of the model are trained to create the best possible sets of discrimination. With the development of technology, the concept of using vocal signals for identification has found numerous practical applications in platforms like information access control, banking services, and secured databases management system, remote access through telephone services and automobile communication systems, etc.,.

Even though many successes have been shown, particularly for single words, recognition based on continuous speech signals is still a topic that has drawn a lot of interest since it captures the natural flow of speech. Thus, applications that call for speaker detection in

natural conversation may find use for speaker recognition based on continuous speech signals. Continuous speech signals do not have the gaps that isolated word recognition systems do, which makes it necessary for the recognition task to guess where each word in the utterances finishes and the others begin in order to produce the proper phrase.

In this regard, it is probable that errors that could occur due to the length of utterances could increase classification error and, as a result, influence the recognition accuracy of continuous speech signals by widening the variance of the speaker's class distribution. Over the years, a variety of algorithms and methods have been employed for the recognition of speech patterns, but the hidden Markov model (HMM), which has been demonstrated to have good recognition performance, is the most widely used technique.

## 2.Literature review

In the study of Katagiri et al, recognition rates of 67–83 percent were reported for isolated-word recognition. Ramesh et al [9] established recognition rates of 92–94.5 percent using HMM for solitary words (numbers). The standard HMM algorithm has been shown to exhibit poor discriminative learning due to the training algorithm, despite the fact that it is widely used in speech recognition technology. To address this shortcoming, various hybrid solutions have been proposed to boost the discriminative classification power [12]. Due to their discriminative and adaptive learning capabilities, machine learning techniques like artificial neural networks (ANN) have also been investigated as a potential technology to aid in statistical voice recognition [13–17].

Numerous applications, including isolated-word recognition, phoneme classifier, and probability estimator for voice recognizers, have shown the power of ANN [13, 14]. While ANN can train with good discriminative performance, particularly for short-term speech signals like isolated words, it also struggles to simulate the temporal fluctuations of long-term speech signals correctly. In the field of speech recognition, hybrid approaches that combine the best elements of the ANN and HMM frameworks have been shown to be effective. A hybrid ANN/HMM design was employed in the work of Trentin et al. [20] to achieve a voice recognition rate of 54.9–89.27 percent with corresponding SNR of 5–20 dB for isolated utterances.

Reynolds et al. used the statistical Gaussian mixture models (GMM) framework, a variation of the HMM, to achieve a speaker-independent voice recognition rate of 80.8-96.8% [21, 22] utilising isolated utterances. Even though continuous speech recognition is the subject of extensive study, the majority of that work is focused on either the accurate detection or recognition of words or their locations in utterances, such as identifying when a speaker uses a word that is not typical of continuous speech. However, the objective of this work is to identify speakers using the utterance waveform distribution of continuous speech. This could be especially helpful for application systems like forensics that need to identify speakers in real-world conversation settings.

The frameworks of ANN are used in this work to implement speaker voice recognition. The paradigm investigates how the ANN may learn in a discriminative and adaptive manner, as well as how to model underlying qualities, provide high classification accuracy, and be

resilient to errors in speech signals. Variable-length speech utterances that are both known to the system and unknown to it are used to assess the performance of the recognition system.

## 3.AN OVERVIEW OF SPEECH RECOGNITION

Speech recognition is a technology that allows a computer to capture human speech using a microphone. These words are later recognized by a speech recognizer, and the system outputs the recognized words at the end.

A speech recognition engine recognizes all words uttered by a human in an ideal situation, but in practices, the performance of a speech recognition engine is dependent on a number of factors. Vocabularies, multiple users and users, and a noisy environment are the major determining factors for a speech recognition engine.

### 3.1. Types of speech recognition

Speech recognition systems are classified into a number of classes based on their ability to recognize words and the number of words in their vocabulary. Speech recognition is divided into several classes, which are as follows:

**A. Isolated Speech**

Isolated words are distinguished by a pause between two utterances; this does not imply that it only accepts a single word, but rather that it requires one utterance at a time.

**B. Connected speech**

Connected words, also known as connected speech, are similar to isolated speech in that they allow for separate utterances with minimal pause between them.

**C. Continuous speech**

Continuous speech, also known as computer dictation, allows the user to speak almost naturally.

**D. Unintentional Speech**

On the most basic level, it can be defined as natural-sounding, unrehearsed speech.

An ASR system with spontaneous speech capability should be able to handle a wide range of natural speech features such as word runs, "ums" and "ahs," and even minor stutters.

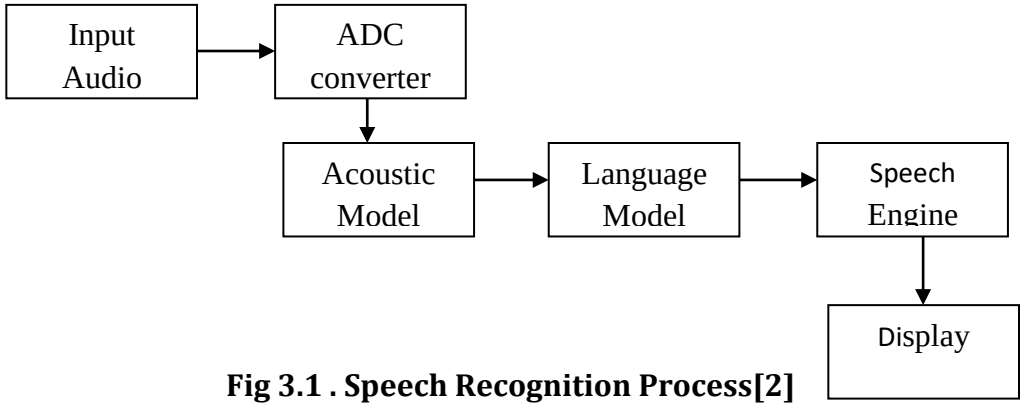


**Fig 3.1 . Speech Recognition Process[2]**

### 3.2 Speech Recognition weakness

Despite all of these advantages and benefits, it is impossible to develop a completely perfect speech recognition system. A variety of factors can impair the accuracy and performance of a speech recognition program.

Speech recognition is an easy process for a human, but it is a difficult process for a machine. When compared to a human mind, speech recognition programs appear less intelligent. This is due to the fact that a human mind is a God-given thing, and the ability to think, understand, and react is natural, whereas for a computer program it is a complicated task, as it must first understand the spoken words in terms of their meanings, and it must create a sufficient b A human has the ability to filter out noise from speech, whereas a machine requires training and assistance in separating the speech sound from other sounds.

## 4. ANN Based Speech Recognition System

In modern Engineering world speech recognition (SR) is one of the recent technologies which is not used commonly in the Outdoor environment due to environmental noise , not an Language barriers . Still there is many research is under gone in this technologies. Speech is common way to interacts human. in this paper, we discussed the Real time Artificial Neural Networks (ANN) based speech recognition. This project keeps that factor in mind, and an effort is made to ensure our project is able to recognize speech in indoor and outdoor noise environments for getting more accuracy rate. we Recorded five words, for each word we create five samples. Data set is a critical role in recognition accuracy of the neural network. If environment may changing the accuracy may depends. for more samples are there ,more could be the accuracy of neural network. Different samples will play a positive roles in total accuracy. By using this project we sampled a more speech data to check the accuracy of speech recognition in real time to getting more than 95% accuracy in any environmental area ,with the help of Artificial Neural Networks

In this project we are trying to increases the accuracy rate of speech recognition in the environmental noise while communication is done, between the devices or machine. Where the works is depends on the speech recognition technique. If we use ANN network model in machine learning all speech-trained systems that achieves better performance in any environmental area. In call centre applications and mobile wireless communications, speech emotion recognition is also used. As a result, we began to conceive of voice as a quick and powerful way to communicate with machines. Speech recognition is the process of translating an auditory signal collected by a microphone or other instrument into a collection of words . To achieve speech understanding, we apply linguistic analysis. Everyone needs to connect with individuals who work in the public sphere, and we need to see each other. Individuals have come to expect computers to have a spoken interface. Humans now require complicated languages for interactions with machines that are difficult to comprehend and use. by using this real time neural network model, we achieved better performance in the speech recognition system.

Speech synthesis is one of the artificial production of human speech. A computer used in this purpose called a speech computer, or speech synthesizer. This speech synthesizer is implemented in various software or hardware products to converts normal language text

into speech it also known as a text-to speech system(TTS) represented in fig.2. it also render an symbolic linguistic representations like phonetic transcriptions into speech

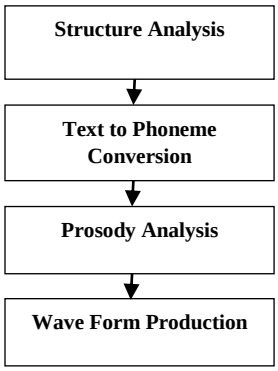


**Fig 2. Model of TTS (Text-To Speech synthesizer)**

In text-to speech system the structural analysis of the input text will be converted to the Phoneme conversion and then it will analysis the prosody of the text which will produce the Wave form of the signal which will also read the text as a speech(Audio Signal).in our project we use the Reverse system model of text to speech system to get the accurate output for communication with various devices in various environmental sphere.

## 4. Proposed method of Processing

The proposed model of the real time ANN based speech recognition system used to increases the accuracy rate of speech recognition in the environmental noise while communication is done, between the devices or machine by a human. Most of the existing models of various neural networks are having the accuracy rate up to 85%,because they are probably having the combined models like Power Normalized Cepstral Coefficient (PNCC) and Modified Group Delay Function (ModGDF) likewise for arrangement we use SVM algorithm, Gaussian models.,etc.

But in this paper we purely used the Artificial Neural network to increases the speech accuracy rate to understand the input speech signal in all the environmental area. In other network model the background noise is the main problem of identifying the input speech signal. For example if the person will say “Right” as the command for move the robot, some times the robot understand as “write” due to background noise.

To Avoid these kind of Malfunction in the robot we take five words as the sample. that is first we Recorded a five words, then for each word we create five samples. while recording input samples Data set is a critical role in recognition accuracy of the neural network. If environment may changing the accuracy may depends. for more samples are there ,more could be the accuracy of neural network. Different samples will play a positive roles in total accuracy.

### 4.1. Neural network architecture and training

The multi layer feed-forward ANN architecture with administered preparing of the removed MFCC include vectors of the expressions was carried out utilizing the Mat lab

platform. The back proliferation calculation, which has been demonstrated to be effective in the minimizing of acknowledgment mistake rates, is used in the preparation of the ANN. The ANN engineering plays a fundamental role in producing a good and advantageous result, and this heavily depends on the number of neurons present in the information. and covert layers, as well as the quantity needed for the task.. We started with the basic design of single information and secret layer neurons in a three layer framework and varied the number of neurons in each layer through organization preparation until the perfect number that delivered the best preparation was achieved.

Ten speakers' various facial expressions were used to evaluate the ANN design. Each expression's deleted element vectors were inserted at the information layer, and the most likely speaker and topic were identified at the output layer. We therefore set the goal yield to "1" for the appropriate speech signals (expression) and "0" for others because the back spread technique normally calls for arrangement of an objective yield that is used during preparation, which is typically not accessible for discourse acknowledgement. This in some way assisted the desirable output and hindered some undesirable produce. By adjusting the loads between the network's components, the network was made ready.

**The following steps should be practiced for great accuracy:**

• The phrases "ha" and "hmm," especially word ends, must be clearly communicated by the speaker. It is not to be used.

• It's critical to place the receiver correctly in order to avoid "rasp." This is especially important when understanding a live lecture or building a profile.

• The speaker should not speak too quickly or too slowly; instead, speak at a relaxed, average pace.

• If profile preparation is to be done, often used course-explicit watchwords that aren't found in word references should be prepared.

• During extended presentations, speakers will take a "ha" or "hmm" pause to assess the SR framework's trustworthiness.

• Speakers should do everything they can to avoid looking at the translated text while they're recording.

• Only the speaker's voice is reliably captured. When responding to questions from understudies, for example, the educator should either rehash the query and then respond or stop recording.

### 4.2.1. Background noise

Different noise evacuation techniques are routinely used when the disturbance source disturbing the indication of interest is noticed. Then, it is possible to dedicate an amplifier to recording the source of the noise and evaluating the space's acoustic environment's incentive to dampen the noise.

This response to motivation is frequently evaluated using the least mean square or recursive least square methods.. In a previous experiment, these strategies had encouraging results when the hubbub included speech or excellent music. In any event, Blind Source Separation (BSS) methods seem more appropriate if there should be an occurrence of hidden noise sources, such as washer or blender noise.

The speech and noise sources used to create the voice signals picked up by the receivers are mixed together.. Free Component Analysis might be a subcategory of BSS, which endeavors to distinguish the various sources based on their factual characteristics (i.e., absolutely information driven). This method is particularly effective for non-Gaussian signs (like discourse) and does not have the opportunity to take into account the producer's or receiver's situation. Instead, it assumes that the sign and the noise will be mixed together directly. This theory is by all accounts not compatible with reasonable histories. In this way, commotion separation in a tolerable dazzling home environment remains an open test despite the significant effort of the neighborhood.

**5. Experiment and Results in ANN**

In text-to speech system the structural analysis of the input text will be converted to the Phoneme conversion and then it will analysis the prosody of the text which will produce the Wave form of the signal which will also read the text as a speech (Audio Signal).in our project we use the Reverse system model of text to speech system to get the accurate output for communication with various devices in various environmental sphere.

The input speech signal is given to the input of the machine or a device via microphone which will be pre processing and converted to the text in the machine then the samples record signal will be matching to the data set of sample then the maximum rate of output will produces as the Output. This process will give the accurate output either noise in the environments.

The choice of a one-word expression was based on the fact that most conversations are fast, and most speakers can express two words at a time. From August to September 2021, each speaker repeated the expressed expression many times at various times. The framework was constructed using these expressions. Following the preparation, 10 speakers (7 female and 3 male) whose discourse waveform designs had been prepared were tested in a progressive atmosphere to perceive their voices. Every one of the 30 speakers was required to offer two expressions, one 1-word and one 20-word expression

known to the framework, followed by another two expressions, one 5-word and one 20-word expression unknown to the framework.

Each speaker four times repeated the utterance and the recognition rate average was estimated using following formula:

$$SR= (TCR/TSU)*100$$

where TCR represents the total number of correct recognition of a speaker from the system and TSU is the total number of spoken utterances presented to the system for recognition. Tables 1 and 2 below show the summary results for 10 speakers for the case of 1-word utterances. For each of the four speakers for a specific expressed expression with speaker 1, it can be shown that the framework effectively identified him as Arjun. an average correlation coefficient of 98.7%. The first attempt with speaker 2 ended in a false rejection, but the speaker was later identified as Yasika in the remaining cases with an average correlation coefficient of 92.3 %. With speaker 3, both first and second attempts

resulted in Recognition but the speaker was recognized in the third attempt whilst the fourth attempt resulted in false acceptance

The input of speaker speaks: "**right**".

| Speakers | 1st Attempt | 2nd Attempt | 3rd Attempt | 4th Attempt | 5th Attempt |
|---|---|---|---|---|---|
| **ARJUN** | Recognition | Recognition | Recognition | Recognition | Recognition |
| **YASIKA** | False Rejection | Recognition | Recognition | Recognition | Recognition |
| **ADJOA** | Recognition | Recognition | Recognition | False Acceptance | Recognition |
| **AARIF** | Recognition | False Acceptance | Recognition | Recognition | Recognition |
| **ASIQUE** | Recognition | Recognition | Recognition | Recognition | Recognition |
| **ANU** | Recognition | Recognition | Recognition | Recognition | Recognition |
| **AJAY** | Recognition | Recognition | Recognition | Recognition | Recognition |
| **BARBRA** | False Acceptance | False Acceptance | False Rejection | Recognition | False Acceptance |
| **BEETAL** | Recognition | Recognition | Recognition | Recognition | Recognition |
| **BEJAY** | Recognition | Recognition | Recognition | False Acceptance | Recognition |

**Table 5.1 1-word utterances Test results for voice recognition of speakers using**

| Speakers | 1st Attempt | 2nd Attempt | 3rd Attempt | 4th Attempt | 5th Attempt |
|---|---|---|---|---|---|
| **ARJUN** | 99.24% | 98.22% | 98.66% | 98.77% | 98.57% |
| **YASIKA** | 75.56% | 99.49% | 99.19% | 98.33% | 98.03% |
| **ADJOA** | 95.13% | 96.15% | 92.20% | 65.75% | 95.77% |

**Table 5.2 1-word utterances Correlation coefficient evaluation based on**

The false acceptance occurs while slow speaking, very high noise with unwanted background sounds louder than the speaker . When there is too much cross talk in the MIC during the test procedure and there is silence for longer than 30 seconds, the signal is rejected.

The normal speaker acknowledgment rates for the 10 speakers utilized in the testing of the framework for the 5-word and 20-word utterances are shown and Figure 3 underneath. Figure 3(a) shows that with the 10 testing discourse tests the framework can satisfactorily perceive the voice of the speakers at a triumph pace of 95% with bogus acknowledgment and bogus dismissal paces of 3% and 2% individually, when

1-word expressions like the prepared informational indexes were utilized.If more preparation data is used, it is possible to improve the acknowledgment precision even further. Since the mouthpiece's characteristics play a considerable role in the kind of acknowledgment precision, the deliberate exhibition value almost likely might be increased.The acknowledgment pace of 98% nonetheless, diminished to 95% when the length of the utterances was expanded to 2-word utterances as portrayed in Figure 3(b). The decline in could be attributed to the massive estimated utterances' increased complexity, which somehow affected how connections were learned.. Due to the increased degree of change abilities that affect the delivery of the message and, consequently, the acceptance rates, the issue of level of confusion related to the production and testing of the large measured discourse designs will also generally be crucial.. The outcomes in Figure 3(c) then again show acknowledgment execution for the instance of 2-word utterances that are obscure to the preparation framework. The speaker acknowledgment rates quickly declined to a normal of 3% with high bogus dismissal and bogus acknowledgment paces of 2% and 5%, separately. The shockingly low acknowledgment rates demonstrate how the testing discourse information's force unearthly thickness dispersions are not exactly consistent with those of the practise discourse expressions. Albeit the acknowledgment pace of 3% might be excessively low for application in speaker-free acknowledgment frameworks, it by one way or another likewise shows that the framework might be sufficient in a speaker-subordinate acknowledgment system to detect frauds.

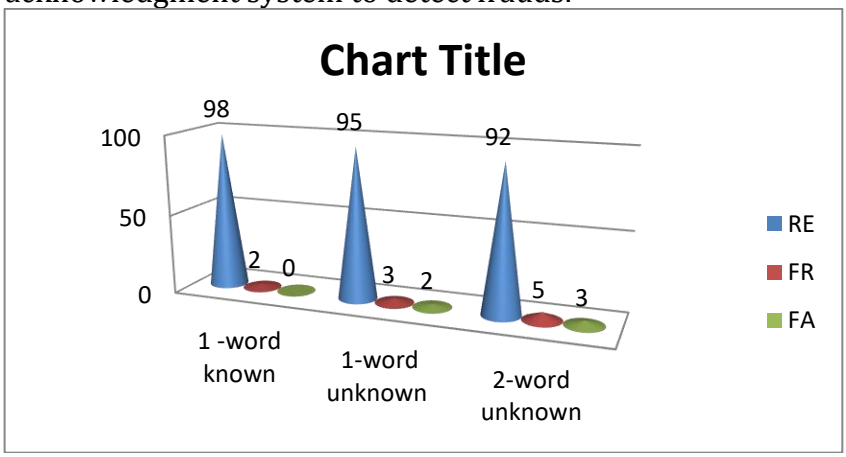


**Fig-3, Recognition rates of first three speakers using (a) 1-word utterances known to system (b) 1-word utterances known to the system (c) 2-word utterances unknown to the system: RE – recognition, FR – false rejection, FA – false acceptance**

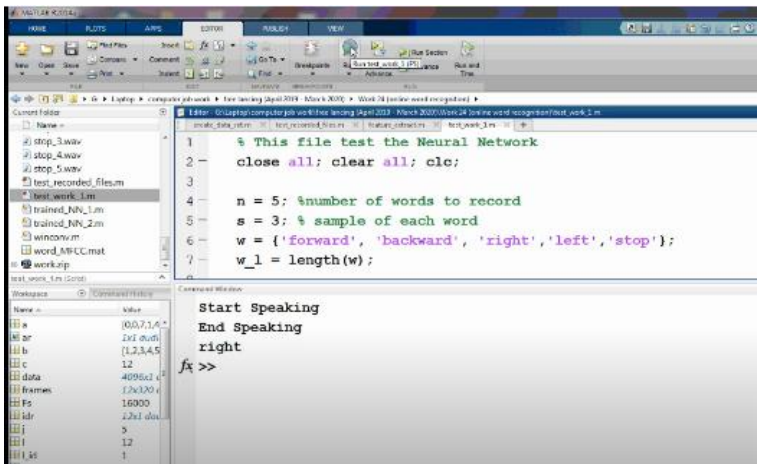


**Fig. 4.Real time speech recognition in Machine learning using ANN**

Figure 4 shows the machine learning output in Matlab stage for the discourse information streaming and preparing for acknowledgment. In experiment Every one of the 10 speakers was first made to absolute (read out) similar four distinct expressions (English words) involving of 1-word expressions, for example, "Right". The Figure 5 is the Interfacing Hardware of ARM 64 bit Processor. The ARM processor is 64bit processor which can be act as the main hatdware device with the name of ic is LPC2148.

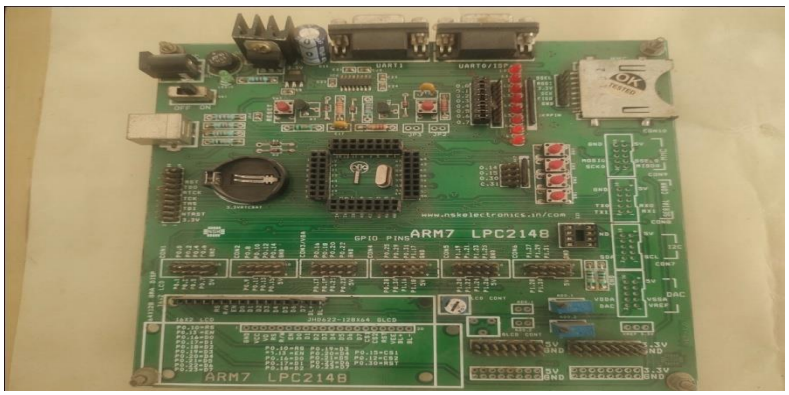


**Fig.5.Interfacing Hardware of ARM 64 bit Processor**

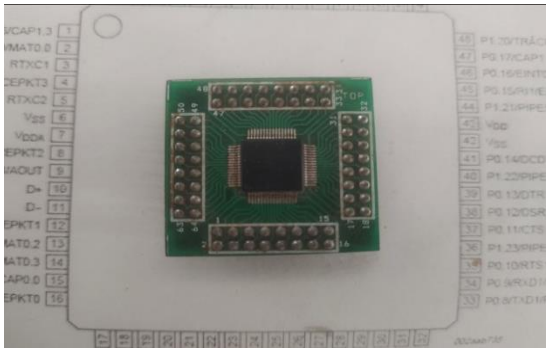

**Fig.6.ARM 64bit Processor**

The interfacing of speech recognition with the system is done by the communication slot PCI(Peripheral control interface ) by the system which is been communicate with the ARM(Advanced RICS Machine) processor .Figure.6 represent the ARM 64bit Processor is offent used in the Mobile phone and Tab which is easy to access with the communication between the System recognition the speech via MIC.ARM 32 bit Processor is user friendly Processor having 64 pins to interfaace many devices for getting high accuracy rate in output of speech recognition with best output in the results.

**Fig.7.PCI communication slot**

## Conclusion

The consequences of speaker acknowledgment dependent on the utilization of consistent discourse expressions and the consolidated structures of ANN in Reverse TTS. We have exhibited through testing of discourse expressions from 10 unique speakers with every speaker giving four distinct expressions and results show the capacity of the framework to perceive speakers with progress pace of 97% for the instance of 1-word utterances for circumstances where the utterances are known to the preparation ANN Model. On account of discourse expressions that are obscure to the framework, an acknowledgment pace of just 2% was unworkable for 1-word utterances due to vey high noise in background.When used as a speaker subordinate framework, this low rate makes it possible for the framework to spot hoaxes, although even lower rates may be required for productivity. The capacity to sufficiently perceive speakers utilizing their consistent discourse waveform might discover valuable applications in frameworks that require recognition of speakers in natural conversation.

## REFERENCE

[1] Rabiner, L. R., “Applications of speech recognition in the area of telecommunication”, IEEE Proc., 1997, pp. 501-510.

[2] Picone, J. W., “Signal modeling techniques in speech recognition,” IEEE Proc., Vol. 81, No. 9, 1993, pp. 1215-1247.
[3] Campbell, J. P. Jr., “Speech recognition: A tutorial”, IEEE, Vol. 85, No. 9, 1997, pp. 1437-1462.
[4] Morgan, N., and Bourland, H., “Continuous speech recognition using multilayer perceptrons with hidden Markov models”, International Conference on Acoustics, Speech and Signal Processing, Albuquerque, 1990, pp. 413-416.
[5] Bourlard, H., and Morgan, N., “Continuous speech recognition by connectionist statistical methods”, IEEE Trans on Neural Networks, Vol. 4, No. 6, 1993, pp. 893-909.
[6] Nichie, A., “Voice recognition system using artificial neural network”, Bachelor Thesis, Computer Engineering Department, University of Ghana, Legon, June 2012.
[7] Huang, X. D., Ariki, Y., and Jack, M., “Hidden Markov models for speech recognition”, Edinburgh University Press, Edinburgh, 1990.
[8] Rabiner, L. R., “A tutorial on hidden Markov models and selected applications in speech recognition”, IEEE Proc., Vol. 77, 1989, pp. 257-286.
[9] Ramesh, P., and Wilpon, J. G., “Modeling state durations in hidden Markov models for automatic speech recognition”, IEEE , Vol. 9, 1992, pp. 381-384.
[10] Katagiri, S., and Chin-Hui, L., “A new hybrid algorithm for speech recognition based on HMM segmentation and learning vector quantization”, IEEE Trans on Speech and Audio Processing, Vol. 1, No. 4, 1993, pp. 421-430.
[11] Frikha, M., Ben Hamida, A., and Lahyani, M., “Hidden Markov models (HMM) isolated-word recognizer with optimization of acoustical and modeling techniques”, Int. Journal of Physical Sciences, Vol. 6, No. 22, 2011, pp. 5064-5074.
[12] Johansen, F. T., “A comparison of hybrid HMM architectures using global discriminative training”, Proceedings of ICSLP, Philadelphia, 1996, pp. 498-501.
[13] Bengio, Y. “Neural network for speech and sequence recognition”, Computer Press, London, 1996. [14] Lippman, R. P., “Review of neural networks for speech recognition”, Neural Computing, Vol. 1, 1989, pp. 1-38.
[15] Renals, S., and Bourlard, H., “Enhanced phone posterior for improving speech recognition”, IEEE Trans on Speech, Audio, Language Processing, Vol.18, No. 6, 2010, pp. 1094-1106.
[16] Yegnanarayana, B., and Kishore, S., “ANN: an alternative to GMM for pattern recognition”, Neural Networks, 2002, pp. 459-469.
[17] Biing-Hwang, J., Wu, C., and Chin-Hui, L., “Minimum classification methods for speech recognition”, IEEE Transactions on Speech and Audio Processing, Vol.5, No. 3, 1997, pp. 257-265.
[18] Haykin, S. O., “Neural networks and learning machines”, 3rd Ed., Prentice Hall, 2008.
[19] Riis, S. K., and Krogh, A., “Hidden neural networks: A framework for HMM/NN hybrids”, International Conference on Acoustics, Speech, and Signal Processing, Munich, 1997, pp. 3233-3236.
[20] Trentin, E., and Gori, M., “Robust combination of neural network and hidden Markov models for speech recognition”, IEEE Transactions on Neural Network, Vol. 14, No. 6, 2003, pp. 1519-1531.
[21] Reynolds, D. A., Quatieri, T. F., and Dunn, R. B., “Speaker verification using adapted Gaussian mixture speaker models”, Digital Signal Processing, Vol 10, No. 103, 2000, pp. 19-41.

[22] Reynolds, D. A., and Rose, R. C., "Robust text-independent speaker identification using Gaussian mixture speaker models", IEEE Transactions on Speech and Audio Processing, Vol 3, No. 1, 1995, pp. 72-83.
[23] Brown, J. C., and Smaragdis, P., "Hidden Markov and Gaussian mixture models for automatic call classification", Journal of Acoustic Society of America, Vol 125, No 6., 2009, pp. 221-224.
[24] Furui, S., "Speaker dependent feature extraction, recognition and processing techniques", Speech Communication, Vol 10, No. 5-6, 1991, pp. 505-520.
[25] Furui, S., "Cepstral analysis technique for automatic speaker verification", IEEE Trans on Acoustics, Speech and Signal Processing, Vol 29, No. 2, 1981, pp. 254-272.
[26] Brookes, M., "Voicebox: speech processing toolbox for Matlab" Imperial College, http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html, (March 3, 2012)
[27] Linde, Y., Buzo, A., and Gray, R., "An algorithm for vector quantizer design", IEEE Transactions on Communications, Vol. 28, 1980, pp. 84-95.
[28] Xuan, G., Zhang, W., Chai, P., "EM Algorithms of Gaussian mixture model and Hidden Markov model", IEEE, 2001, pp. 145-148.
[29] Dempster, A., Laird, N., Rubin, D., "Maximum Likelihood from incomplete data via the EM algorithm", Journal of Royal Statistical Society, Vol 39, No. 1, 1977, pp. 1-38.