I J C T A, 9(19) 2016, pp. 9305-9310

An Enhanced Proposed Hybrid DTW and HMM approach for Speech Recognition System

Neha*, R. C. Gangwar* and Rajeev Bedi*

ABSTRACT

In recognition of speech signals includes noise as well as they are not in the form of processing due to noisy, heavy and unidentified words. Systems create problem to recognize the input signals for process and not provide accurate results. Various models and approaches are defined by researches and claims that the recognition system is accurate and providing better results. Thus ubiquitous nature of input speech signals become potential source of crowd wisdom extraction especially in terms of voice words then classify and analyze the words for further research. This work is an effort to see the recognition system which recognize input signals, translate the unidentified speech words and provide better results than existing DTW. Proposed work is based on enhancement of DTW and experiments are carried out to observe the effect of proposed method on speech recognition which clearly indicates the improvement in various performance metrics.

Keywords: DTW,LPC,HMM,Slang.

I. INTRODUCTION

From previous century in the world, there has been interest in creation computers which can perceive and convert human signals into understanding natural language, process images and text into identified manner. Various researches are still working to find precious and simple system which can communicate efficiently with the computer to generate results. Speech is a efficient mode of communication and it is a advance topic of research from past few decades. It is very interesting phenomenon where the human beings are interacting with system and system's task is to use some machine learning as well some NLP for recognition. Previous techniques of speech recognition become popular and advance in 1970 where automatic speech recognition was used to analyze the speech signals. In Bell Laboratories, an isolated digit recognizer was developed which was for single speaker. This system was developed in 1952 based on dynamic programming. After this research many efforts was made by various researchers where the speaker independently can handle the voice signals. This voice recognition can be considered in many research fields such as in bioinformatics, biometric and multi model recognition. In military purpose this recognition plays a great role.

1.1. Speech Recognition Basics

In order to understand working principles of speech recognition systems, it is important to be familiar and understand some basic concepts used in this technology. In speech recognition utterance, pronunciation, grammar vocabulary, speaker dependency and independency matters a lot. These all factors affect the performance of the recognition system or we can say efficiency of the system depends on these all factors. Utterance is any stream of speech between two periods of silence. Stream is a flow of input signals which are defied as where S is defining the flow of input stream signals and are the various streams in time

* Department of Computer Science Engineering, BCET, Gurdaspur, Punjab, India, *E-mail: neha.aiet835@gmail.com*

interval T. Pronunciation defines the speech words which means how the word sound like and grammar in this recognition is important which extract domain knowledge of the system within the recognition engine and compare with the dictionary to extract the actual meaning. On the other hand Speaker Dependence is designed around a specific speaker and is much more accurate for that speaker but much less accurate for any other speaker. Thus these all are important part of speech recognition which helps to extract domain as well as complete knowledge of system to recognize the actual input signals.

1.2. Architecture of Speech Recognition

In speech recognition most of the voice signal captured noise in the environment which create problem to recognize the signals and make system complex. Simple architecture of speech recognition is described in the figure where input speech signals and provided to the system which is passed through the noisy channel. This noisy channel can add some high frequency waves in the input signal which may make input signals more complex or it may add some external variables which may harm the sound quality and the system will not recognize the actual input signals in efficient manner. This input signals which carries noise will pass through the decoder which will suppress the noise as well and will recognize the actual input signal to output signal.

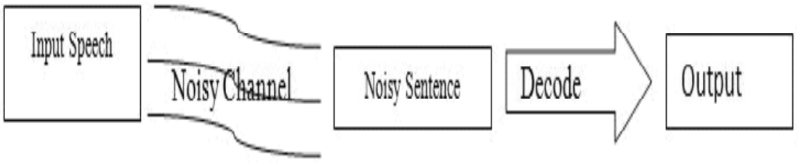


Figure 1: Architecture of Speech Recognition

II. LITERATURE SURVEY

Various researchers worked on speech recognition. In this broad area of research literature indicates that efforts can be made to develop such kind of system which can recognize the system efficiently and in dynamic time wrap thus less complexity is to be faced. For this synthesis,

Anupam Choudhary et. al. (2012) [1] The AI approach is used in describing the speech recognition process. The language mode, trigram model and acoustic model is used in this recognition method. Acoustic model interfaced with telephony system to test the and manage the dialogues of speakers' .No any graphical user interface is used in this process.

Alexandre Trilla (2012) [2] described the ASR approach used by Natural language processing techniques. It describes the text to speech synthesis and vice versa i.e. an Automatic speech recognition approach.

D D Doye et. al. (2015) [3] they used the non linear approach time alignment model instead of old DTW approach. Time alignment algorithm is implemented on Marathi language. In the testing they used 46 monosyllabic confusing alphabets and 46 confusing name . Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC) and Linear Prediction Coefficient (LPC) are used in this research.

Dr. Kavita R. et. al. (2014) [4] They used the sampling technique by using the method of digitizing of audio samples. The Mel Frequency Cepstral Coefficients is used for extraction process. These coefficients woks with DTW approach on matching with Tamil database. The main interest of this research is security during extracting and matching during using the approaches of DTW and mathematical calculations.

Elyes et. al. (2014) [5] The SVM/HMM hybrid approach is used on Arabic Automatic Speech Recognition on triphone modeling. They emphasize the Arabic SR system that is based on triphones are HMM with 64.68%, MLP/HMM with 72.39% and SVM/HMM with 75.01% hybrid approach

Fook C.Y et.al. (2012) [6] The various approaches of speech recognition methods are compared and summarized in this research.

Jayashree Padmanabhan et.al. (2015) [7] The ASR with various models and techniques like Gaussian mixture model, machine learning and HMM is reviewed. Scanning of speech, preprocessing of speech, extraction and classification of speech input is done by using the acoustic, bottleneck and MLP feature.

Kenji Sagae et.al. (2009) [8] worked on incremental model or system which complete the speech by using its processing capabilities. It completes the speech before its utterance.

Mohammad et.al. (2014) [9] The main area of this research is emotion recognition of any speech by its analysis. Same utterance is done by user 6 times and recorded. The various parameters are used for evaluating like pitch, frequency and intensity. "PRAAT" a freeware software is used in this research.

Poonam.S.Shetake et. al. (2014) [10] The main focus of the researchers in this research is on character recognition and Text to speech conversion techniques or approaches.

Siva Prasad Nandyala et. al. (2014) [11] In this, they used hybrid approach of DTW and HMM using kernel adaptive filters for analyzing and recognition of speech. They used the noise filtration techniques also. This hybrid approach resulted better than traditional one.

Xiang-Lilan et. al.(2014) [12] The new merged weight DTW algorithm is introduced in this paper. By using DTW, template confidence index is used for measuring the similarities between testing data and training. Merge approach of HMM, DTW and SD speech recognition datasets, gave six times better result than DTW overall.

Zue, V. et. al. (2011) [13] N word string matching and filtering of components is implemented in this work. They defined an approach for audio text, dialogues, icons and graphics. An understandable urban penetration and speech recognition system is used. The main interest of this research is pairing of words which can help in searching and navigations.

III. PROPOSED SCHEME

This proposed scheme algorithm deals with voice recognition system where words with different entities are present. This algorithm is based on identification of voice signals and probability of co-occurrence if binding words in voice system and unidentified words using various performance metrics. Steps of algorithm are given below:

Input

Voice signal having un-identified words, slang words or noisy words.

Output

Identification of voice signals using various performance metrics. If unidentified words are present then it should be weeded out from the system.

Procedure

- Find the close binding of speech words in the input audio signal based on DTW and proposed approach. Let S is is combination of speech input signals where s_1, s_2 are sub set of S. $S = (s_1 , s_2)$ *and* $s_1, s_2 \in S$,

s_1 , s_2 Contains combination of words in input signal ***I*** which is defined as:

$$S_1 = (w_1, w_2, ..., w_n)$$
$$S_2 = (w_1, w_2, ..., w_n)$$

Combine HMM and DTW for recognition process and to enhance the speech recognition system. Following steps are used:

- Thus similarity between two voice input signal $s_1 = (w_1 , w_2 , ...w_n)$ and $s_2 = (w_1 , w_2 , ...w_n)$ is a n-dimensional space which can be computed as: $dist(s_1 , s_2) = | s_1 - s_2$

Combine HMM and DTW for recognition process and to enhance the speech recognition system

- This optimum function computes the optimum value of DTW between s_1 and s_2. Analysis of these speech signals is analyzed at the part of recognition using rules.

a. If W_1 occurs before and after a proper noun then it is significant.

b. If is coexisting with collective noun and has reference to a proper noun then W_1 is less significant.
c. Else W_1 is insignificant and W_1can be weeded out of input signal.

Compute the performance metrics with the comparison of proposed approach. Update the recognition channel with T_{new} than T_{old}.

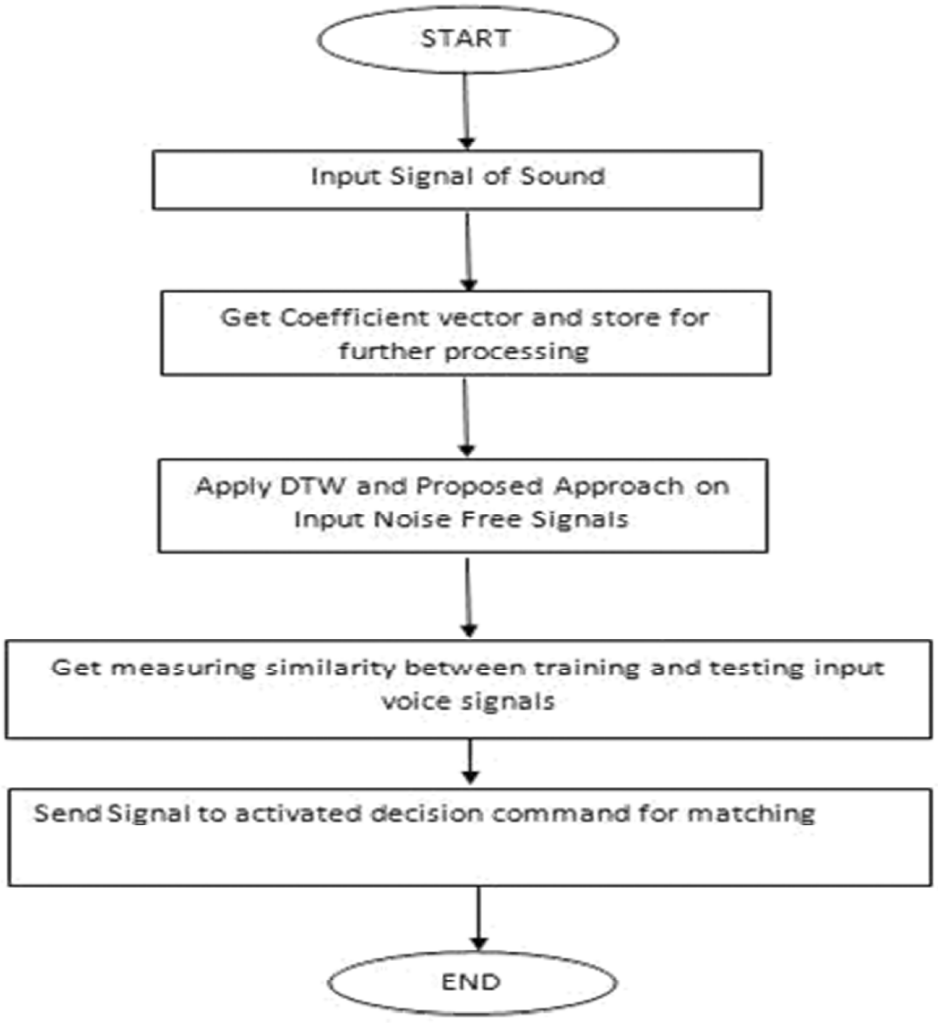


Figure 2: Flowchart of Proposed Approach

4.1. Pseudo code

1. *Begin*
2. *Initialize input (w1,w2) to NULL,*
3. *Initializes streaming S at time t,*
4. *Enter input Speech w1 for processing at time t1*
5. *Receive sound output (W1,W2)*
6. *Calculate Simf(Sf) between two speech signals*
7. *Repeat (until end of input speech)*
8. *End*

IV. RESULTS AND DISCUSSIONS

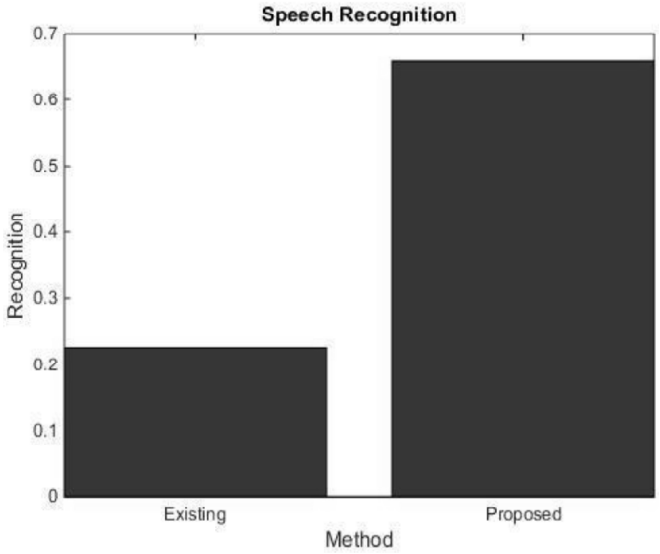


Figure 3: Speech Recognition Proposed vs DTW

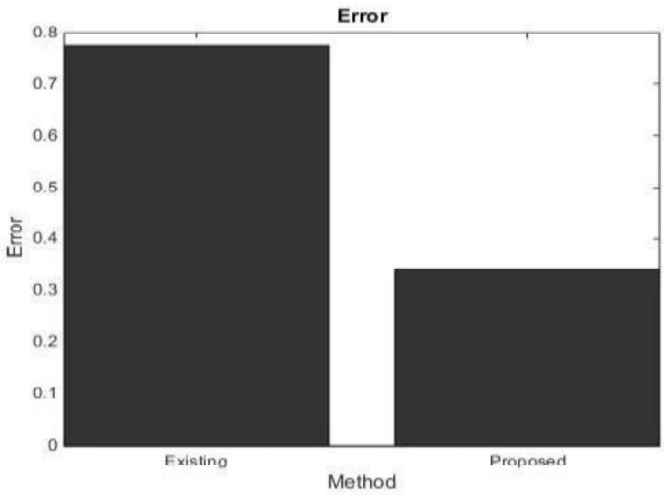
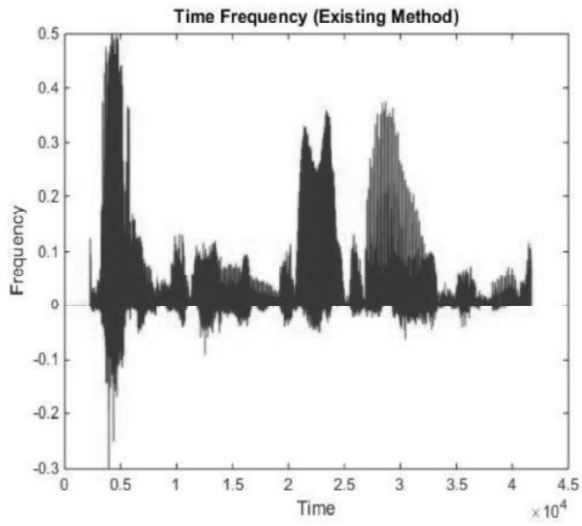


Figure 4: Error Rate Recognition Proposed vs DTW



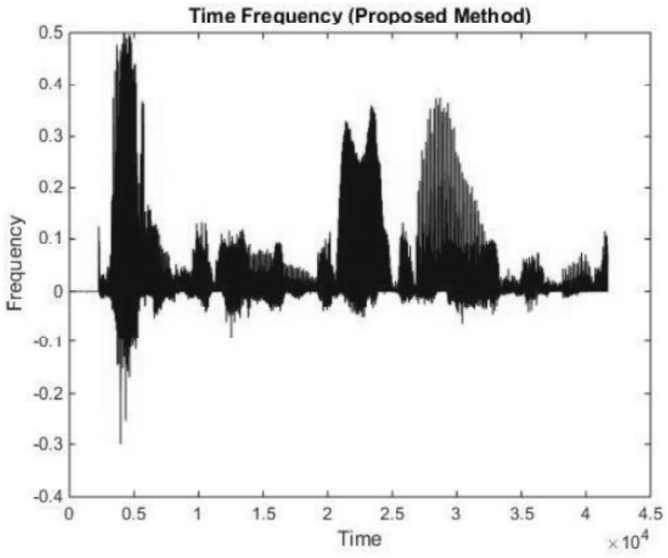


Figure 5: Time Frequency DTW Vs Proposed

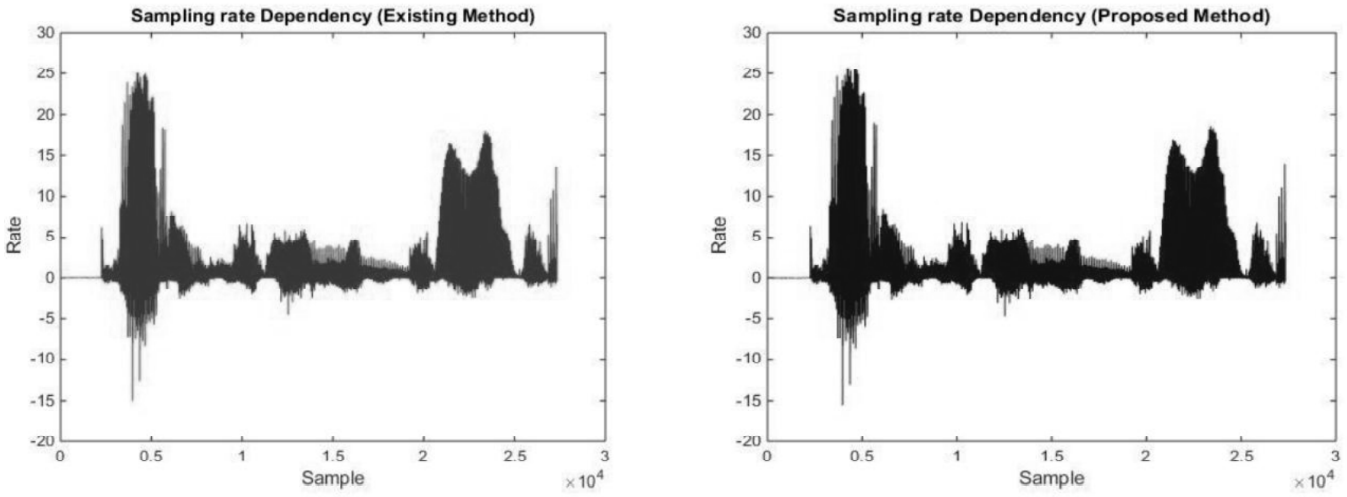


Figure 6: Sampling Rate Dependency DTW Vs Proposed

REFERENCES

[1] Anupam Choudhary, Ravi Kshirsagar, "Process Speech Recognition System using Artificial Intelligence Technique", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-5, November 2012.

[2] Alexandre Trilla, "Natural Language Processing techniques in Text-To-Speech synthesis and Automatic Speech Recognition", IEEE, Vol. 4, 2012.

[3] Dr. Kavitha, Nachammai, Ranjani, Shifali., "Speech Based Voice Recognition System for Natural Language Processing", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5301-5305.

[4] Zue V, Glass, J., Goodine, D., Leung, H., Phillips, M, Polifroni, J., Seneff, S,"Integration of speech recognition and natural language processing in the MIT VOYAGER system", IEEE, 2011.

[5] Kenji Sagae and Gwen Christian and David DeVault and David R. Traum, "Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue System", Proceedings of NAACL HLT 2009: Short Papers, pages 53–56, Boulder.