VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT

Article *in* Journal of Applied and Fundamental Sciences · November 2015

CITATIONS
22

READS
67,298

4 authors, including:

Pranab Das
Assam Don Bosco University
14 PUBLICATIONS **29** CITATIONS
SEE PROFILE

Vijay Prasad
Assam Don Bosco University
7 PUBLICATIONS **70** CITATIONS
SEE PROFILE

VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT

Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad*
Department of Computer Science & Engineering and Information Technology, School of Technology, Assam Don Bosco University, Assam, India
*For correspondence. (vpd.vijay82@gmail.co)

Abstract: VOICE RECOGNITION SYSTEM:SPEECH-TO-TEXT is a software that lets the user control computer functions and dictates text by voice. The system consists of two components , first component is for processing acoustic signal which is captured by a microphone and second component is to interpret the processed signal, then mapping of the signal to words. Model for each letter will be built using Hidden Markov Model(HMM). Feature extraction will be done using Mel Frequency Cepstral Coefficients(MFCC). Feature training of the dataset will be done using vector quantization and Feature testing of the dataset will be done using viterbi algorithm. Home automation will be completely based on voice recognition system.

Keywords: Voice recognition, MFCC, HMM, Vector quantization, Viterbi algorithm, Feature extraction

1. Introduction:

Voice is the basic, common and efficient form of communication method for people to interact with each other. Today speech technologies are commonly available for a limited but interesting range of task. This technologies enable machines to respond correctly and reliably to human voices and provide useful and valuable services.
As communicating with computer is faster using voice rather than using keyboard, so people will prefer such system. Communication among the human being is dominated by spoken language, therefore it is natural for people to expect voice interfaces with computer.

This can be accomplished by developing voice recognition system:speech-to-text which allows computer to translate voice request and dictation into text. Voice recognition system:speech-to-text is the process of converting an acoustic signal which is captured using a microphone to a set of words. The recorded data can be used for document preparation.

2. Classification of speech recognition system:

Speech recognition system can be classified in several different types by describing the type of speech utterance, type of speaker model and type of vocability that they have the ability to recognize. The challenges are briefly explained below:

A. Types of speech utterance

Speech recognition are classified according to what type of utterance they have ability to recognize. They are classified as:
1) Isolated word: Isolated word recognizer usually requires each spoken word to have quiet (lack of an audio signal) on bot
h side of the sample window. It accepts single word at a time.
2) Connected word: It is similar to isolated word, but it allows separate utterances to 'run-together' which contains a minimum pause in between them.
3) Continuous Speech: it allows the users to speak naturally and in parallel the computer will determine the content.
4) Spontaneous Speech: It is the type of speech which is natural sounding and is not rehearsed.

B. Types of speaker model

Speech recognition system is broadly into two main categories based on speaker models namely speaker dependent and speaker independent.

1) Speaker dependent models: These systems are designed for a specific speaker. They are easier to develop and more accurate but they are not so flexible.
2) Speaker independent models: These systems are designed for variety of speaker. These systems are difficult to develop and less accurate but they are very much flexible.

C. Types of vocabulary

The vocabulary size of speech recognition system affects the processing requirements, accuracy and complexity of the system. In voice recognition system:speech-to-text the types of vocabularies can be classified as follows:
1) Small vocabulary: single letter.
2) Medium vocabulary: two or three letter words.
3) Large vocabulary: more letter words.

3. Survey of research papers:

Kuldip K. Paliwal and et al in the year 2004 had discussed that without being affected by their popularity for front end parameter in speech recognition, the cepstral coefficients which had been obtained from linear prediction analysis is sensitive to noise. Here, the use of spectral subband centroids had been discussed by them for robust speech recognition. They discussed that performance of recognition can be achieved if the centroids are selected properly as in comparison with MFCC. to construct a dynamic centroid feature vector a procedure had been proposed which essentially includes the information of transitional spectral information [1].

Esfandier Zavarehei and et al in the year 2005, studied that a time-frequency estimator for enhancement of noisy speech signal in DFT domain is introduced. It is based on low order auto regressive process which is used for modelling. The time-varying trajectory of DFT component in speech which has been formed in Kalman filter state equation. For restarting Kalman filter, a method has been formed to make alteration on the onsets of speech. The performance of this method was compared with parametric spectral substraction and MMSE estimator for the increment of noisy speech. The resultant of the proposed method is that residual noise is reduced and quality of speech in improved using Kalman filters [2].

Ibrahim Patel and et al in the year 2010, had discussed that frequency spectral information with mel frequency is used to present as an approach in the recognition of speech for improvement of speech, based on recognition approach which is represented in HMM. A combination of frequency spectral information in the conventional Mel spectrum which is based on the approach of speech recognition. The approach of Mel frequency utilize the frequency observation in speech within a given resolution resulting in the overlapping of resolution feature which results in the limit of recognition. In speech recognition system which is based on HMM, resolution decomposition is used with a mapping approach in a separating frequency. The result of the study is that there is an improvement in quality metrics of speech recognition with respect to the computational time and learning accuracy in speech recognition system[6].

Kavita Sharma and Prateek Hakar in the year 2012 has represented recognition of speech in a broader solutions. It refers to the technology that will recognize the speech without being targeted at single speaker. Variability in speech pattern, in speech recognition is the main problem. Speaker characteristics which include accent, noise and co-articulation are the most challenging sources in the variation of speech. In speech recognition system, the function of basilar membrane is copied in the front-end of the filter bank. To obtain better recognition results it is believed that the band subdivision is closer to the human perception. In speech recognition system the filter which is constructed for speech recognition is estimated of noise and clean speech[10].

Puneet Kaur, Bhupender Singh and Neha Kapur in the year 2012 had discussed how to use Hidden Markov Model in the process of recognition of speech. To develop an ASR(Automatic Speech Recognition) system the essential three steps necessary are pre-processing, feature Extraction and recognition and finally hidden markov model is used to get the desired result. Research persons are continuously trying to develop a perfect ASR system as there are already huge advancements in the field of digital signal processing but at the same time performance of the computer are not so high in this field in terms of speed of response and matching accuracy. The three different technique used by research fellows are acoustic phonetic approach, pattern recognition approach and knowledge based approach[4].

Chadawan Ittichaichareon and Patiyuth Pramkeaw in the year 2012 had discussed that signal processing toolbox has been used in order to implement the low pass filter with finite impulse response. Computational implementation and analytical design of finite impulse response filter has been successfully accomplished by performing the performance evaluation at signal to noise ratio level. The results are improved in terms of recognition when low pass filters is used as compared to those process which involves speech signal without filtering[3].

Geeta Nijhawan, Poonam Pandit and Shivanker Dev Dhingra in the year 2013 had discussed the techniques of dynamic time warping and mel scale frequency cepstral coefficient in the isolated speech recognition. Different features of the spoken word had been extracted from the input speech. A sample of 5 speakers has been collected and each had spoken 10 digits. A database is made on this basis. Then feature has been extracted using MFCC.DTW is used for effectively dealing with various speaking speed. It is used for similarity measurement between two sequence which varies in speed and time[5].

4. Table of comparison:

Table 1: Table of comparison.

Author(s)	Year	Paper name	Technique	Results
Kuldip K. Paliwal	2004	Recognition of Noisy Speech Using Dynamic Spectral Subband Centroids	Use of spectral subband Centroids	It showed that the new dynamic SSC coefficients are more resilient to noise than the MFCC features.
Esfandier Zavarehei	2005	Speech Enhancement using Kalman filters for Restoration of short-time DFT trajectories	Concept sequence modelling, two-level semantic-lexical modelling, and joint semantic-lexical modelling	Increase the semantic information utilized and tightness of integration between lexical and semantic items
Ibrahim Patel	2010	Speech Recognition Using HMM with MFCC-an analysis using Frequency Spectral Decomposition Technique	Resolution Decomposition with Separating Frequency is the mapping approach	It show an improvement in the quality metrics of speech recognition with respect to computational time, learning accuracy for a speech recognition system
Kavita Sharma	2012	Speech Denoising using Different Types of Filters	FIR, IIR, WAVELETS, FILTER	Use of filter shows that estimation of clean speech and noise for speech enhancement in speech recognition
Bhupinder Singh	2012	Speech Recognition with Hidden Markov Model	Hidden Markov Model	Develop a voice based user machine interface system
Patiyuth Pramkeaw	2012	Improving MFCC-based speech classification with FIR filter	FIR Filter	Shows the improvement in recognition rates of spoken words
Shivanker Dev Dhingra	2013	Isolated Speech Recognition using MFCC and DTW	Dynamic Time Warping(DTW)	It shows that the DTW is the best non linear feature

				matching technique in speech identification, with minimal error rates and fast computing speed

5. Overview of voice recognition system:speech-to-text:

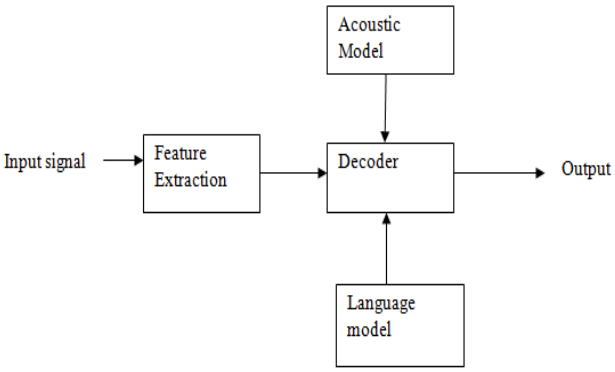

Figure 1: Overview of Voice Recognition System:Speech-to-text.

Input signal- Voice input by the user.
Feature Extraction- it should retain useful information of the signal, deduct redundant and unwanted information, show less variation from one speaking environment to another, occur normally and naturally in speech.
Acoustic model- it contains statistical representations of each distinct sounds that makes up a word.
Decoder- it will decode the input signal after feature extraction and will show the desired output.
Language model- it assigns a probability to a sequence of words by means of a probability distribution.
Output- interpreted text is given by the computer.

The main of the project is to recognize speech using MFCC and VQ techniques. The feature extraction will be done using Mel Frequency Cepstral Coefficients(MFCC). The steps of MFCC are as follows:-
1) Framing and Blocking
2) Windowing
3) FFT(Fast Fourier Transform)
4) Mel-Scale
5) Discrete Cosine Transform(DCT)
Feature matching will be done using Vector Quantization technique. The steps are as follows:-
1) By choosing any two dimensions, inspection on vectors is done and data points are plotted.
2) To check whether data region for two different speaker are overlapping each other and in same cluster, observation is needed.
3) Using LGB algorithm Function Vqlbg will train the VQ codebook.
The extracted features will be stored in .mat file using MFCC algorithm. Models will be created using Hidden Markov Model(HMM). The desired output will be shown in matlab interface.

6. Conclusion:

In this paper the fundamentals are discussed and its recent progress is investigated. The various approaches available for developing a Voice Recognition System based on adapted feature extraction technique and the speech recognition approach for the particular language are compared in this paper. The main aim of our project is to develop a system that will allow the computer to translate voice request and dictation into text using MFCC and VQ techniques. Feature extraction and feature matching will be done using Mel Frequency Cepstral Coefficients and Vector Quantization technique. The extracted feature will be stored in .mat file. A distortion measure which is based on minimizing the Euclidean distance will be used while matching the unknown speech

signal with the database of the speech signal.In near future, home automation will be completely based on Voice Recognition System.

Reference:

[1] Jingdong Chen, Member, Yiteng (Arden) Huang, Qi Li, Kuldip K. Paliwal, "Recognition of Noisy Speech using Dynamic Spectral Subband Centroids" IEEE SSIGNAL PROCESSING LETTERS, Vol. 11, Number 2, February 2004.
[2] Hakan Erdogan, Ruhi Sarikaya, Yuqing Gao, "Using semantic analysis to improve speech recognition performance" Computer Speech and Language, ELSEVIER 2005.
[3] Chadawan Ittichaichareon, Patiyuth Pramkeaw, "Improving MFCC-based Speech Classification with FIR Filter" International Conference on Computer Graphics, Simulation and Modelling (ICGSM'2012) July 28-29, 2012 Pattaya(Thailand).
[4] Bhupinder Singh, Neha Kapur, Puneet Kaur "Sppech Recognition with Hidden Markov Model:A Review" International Journal of Advanced Research in Computer and Software Engineering, Vol. 2, Issue 3, March 2012.
[5] Shivanker Dev Dhingra, Geeta Nijhawan, Poonam Pandit, "Isolated Speech Recognition using MFCC and DTW" International Journal of Advance Research in Electrical, Electronics and Instrumentation Engineering, Vol.2, Issue 8, August 2013.
[6] Ibrahim Patel, Dr. Y. Srinivas Rao, "Speech Recognition using HMM with MFCC-an analysis using Frequency Spectral Decomposition Technique" Signal and Image Processing:An International Journal(SIPIJ), Vol.1, Number.2, December 2010.
[7] Om Prakash Prabhakar, Navneet Kumar Sahu,"A Survey on Voice Command Recognition Technique" International Journal of Advanced Research in Computer and Software Engineering, Vol 3,Issue 5,May 2013.
[8] M A Anusuya, "Speech recognition by Machine", International Journal of Computer Science and Information security, Vol. 6, number 3,2009.
[9] Sikha Gupta, Jafreezal Jaafar, Wan Fatimah wan Ahmad, Arpit Bansal, "Feature Extraction Using MFCC" Signal & Image Processing:An International Journal, Vol 4, No. 4, August 2013.
[10] Kavita Sharma, Prateek Hakar "Speech Denoising Using Different Types of Filters" International journal of Engineering Research and Applications Vol. 2, Issue 1, Jan-Feb 2012