International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-2, Issue-4, April 2014

Feature Extraction Using MFCC Algorithm

Chaitanya Joshi, Kedar Kulkarni, Sushant Gosavi, Prof. S. B. Dhonde

***Abstract*— There are various algorithms available, amongst that MFCC (Mel Frequency Cepstrum Coefficient) is quite efficient and accurate result oriented algorithm. Here in this algorithm Feature Extraction is used and Euclidian Distance for coefficients matching to identify speaker identification.**

***Index Terms*— Euclidian Distance, Feature Extraction, MFCC, Vector Quantization.**

I. INTRODUCTION

Speech is the primary, and the most convenient means of communication between people. The developments are done for the use of speech as for the security purpose in various fields. So speech recognition can be defined as Speech Recognition (is also known as Automatic Speech Recognition (ASR) or computer speech recognition) is the process of converting a speech signal to a sequence of words, by means of an implemented algorithm.

Speech recognition technology made it easy to follow the computer command and make it understand to human languages. The aspect of designing of speech recognition technology is to develop techniques and systems for speech input to machine and to represent it in some form of representation.

Now a days the ASR is used at various places such as updated travel information, stock price quotations, weather reports, Data entry, voice dictation, access to information: travel, banking, Commands, Avionics, Automobile portal, speech transcription, Handicapped people (blind people) supermarket, railway reservations etc.

One of the problems faced in speech recognition is that the spoken word can be vastly altered by accents, dialects and mannerisms. In South Africa, there is a large variety of languages and dialects. Even the most basic speech recognition systems perform poorly when trying to recognize words spoken by English second language speakers. The motivation behind this survey is to investigate speech recognition and more specifically what research has been around dealing with the problem of large variations in dialects. Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Many speech recognition applications, such as voice dialing, simple data entry and speech-to-text are in existence today.

The basic model of speech recognition is given below i.e., the basic speech recognition process shown below.

Manuscript received April 05, 2014.

Chaitanya Joshi, Department of Electronics Engineering, All India Shri Shivaji Memorial Society's Institute of Information Technology and University of Pune, Pune, Maharashtra, India, (+91) 9975766302

Sushant Gosavi, Department of Electronics Engineering, All India Shri Shivaji Memorial Society's Institute of Information Technology and University of Pune, Pune, Maharashtra, India, (+91) 9422152383

Kedar Kulkarni, Department of Electronics, All India Shri Shivaji Memorial Society's Institute of Information Technology and University of Pune, Pune, Maharashtra, India, (+91) 8149868068,

Prof. S.B. Dhonde, Assistant Professor, Department of Electronics Engineering, All India Shri Shivaji Memorial Society's Institute of Information Technology and University of Pune, Pune, Maharashtra, India.

.

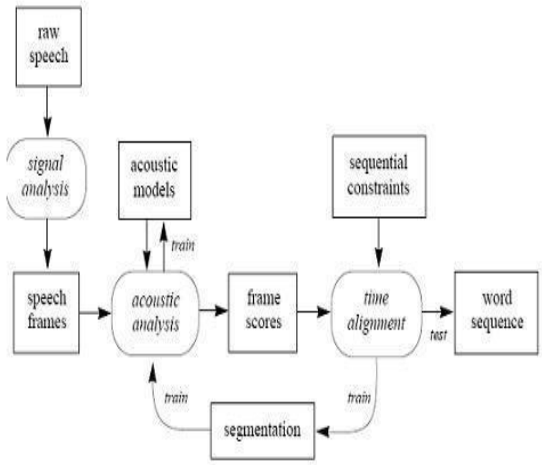

Fig 1. Block diagram of speech processing

II. TYPE OF SPEECH RECOGNITION

The different types of the speech recognition are available. Speech recognition systems can be divided in several different classes by describing what types of utterances they have the ability to recognize. These are classified as follows:

Isolated Words:

It accepts single words or single utterance at a time. These systems require the speaker to wait between utterances (during the pauses).It can be called as Isolated Utterance.

Connected Words:

Connected word systems (or more correctly connected utterances) are similar to isolated words, but allows separate utterances to be 'run-together' with a minimal pause between them.

Continuous Speech:

User speaks in a natural way and computer recognizes the speech and then it applies further procedure on it.

Spontaneous Speech:

Spontaneous speech are the words which have meanings depending on expression of humans. It must handle such spontaneous expression like "ums" and "ahs", and even slight stutters.

III. CLASSIFICATION

The following structure like tree shows the speech processing applications. The classification of ASR can be followed as:

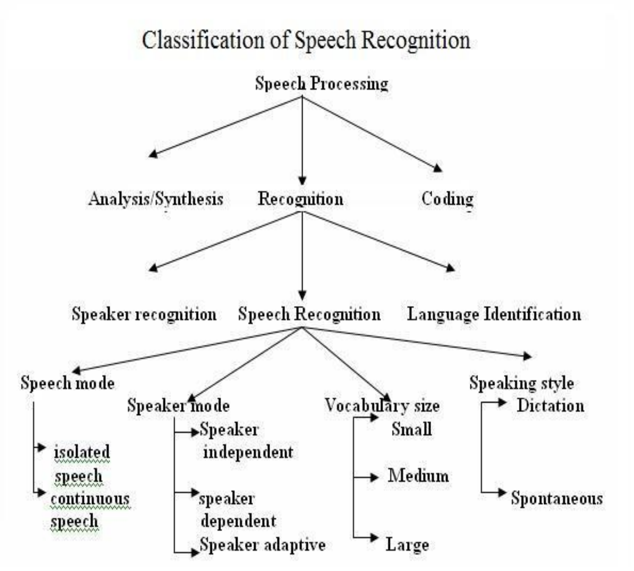

Fig 2. Classification of speech recognition

IV. APPROACHES TO SPEECH RECOGNITION

Basically there exist three approaches to speech recognition.

They are:

1. Acoustic Phonetic Approach
2. Pattern Recognition Approach
3. Artificial Intelligence Approach

A. Acoustic Phonetic Approach:

The first step in the acoustic phonetic approach is a spectral analysis of the speech combined with a feature detection that converts the spectral measurements to a set of features that describe the broad acoustic properties of the different phonetic units. After that segmentation and labeling phase is done in which the speech signal is segmented into stable acoustic regions, and each segmented region is labeled, result is a lattice characterization of the speech. Finally attempts to determine a valid word (or string of words) from the phonetic label sequences produced by the labeled it can be assured language constraints on the task are invoked in order to access the lexicon for word decoding based on the speech.

B. Pattern Recognition Approach:

The essential feature is that it uses a well formulated Mathematical framework and establishes consistent speech pattern represents set of labeled training samples via a formal training algorithm.

A speech pattern representation can be in the form of a speech template or a statistical model (e.g., a HIDDEN MARKOV MODEL or HMM) and can be applied to a sound, a word, or a phrase. In the pattern-comparison stage, a comparison is made between the unknown speeches (the speech to be recognized) with each possible pattern learned in the training stage in order to determine the identity of the unknown according to the goodness of match of the patterns.

C. Artificial Intelligence Approach:

The Artificial Intelligence approach is a hybrid of the acoustic phonetic approach and pattern recognition approach. In this, it tells about the concept of Acoustic phonetic and various pattern recognition methods. This uses the information regarding linguistic, phonetic and spectrogram.

This knowledge is came from careful study of spectrograms and is incorporated using rules or procedures.

It has limited success, largely due to the difficulty in quantifying expert knowledge. Another difficulty is the integration of many levels of human knowledge phonetics, phonotactics, lexical access, syntax, semantics and pragmatics. Knowledge enable the algorithms to work better. This system enhancement has contributed considerably to the design of all successful strategies reported.

V. FEATURE EXTRACTION

The main goal of the feature extraction step is to compute a sequence of feature vectors that provides a compact representation of the given input signal. The feature extraction is performed in three stages. In the first stage speech is analyzed. It performs on spectrums of frequencies and analyzed the signals to generate power spectrum envelopes of short speech intervals. The second stage compiles an extended feature vector composed of static and dynamic features. Finally, the last stage (which is not always present) transforms these extended feature vectors into more compact and robust vectors that are then supplied to the recognizer.

Various methods for Feature Extraction:

The various feature extraction methods are tabulated below:

Method	Property	Comment
Principal Component Analysis(PCA)	Nonlinear feature Extraction method, Linear map; fast; eigenvector-based	Traditional, eigenvector based method, also known as karhuneu-Loeve expansion; good For Gaussian data.
Linear Discriminant Analysis(LDA)	No linear feature Extraction method, Supervised linear map; fast eigenvector-based	Better than PCA for classification;
Independent Component Analysis (ICA)	Nonlinear feature Extraction method, Linear map, iterative non-Gaussian.	Blind course separation, used for de-mixing non-Gaussian distributed sources(features)
Linear Predictive coding	Static feature extraction method,10 to16 lower order co-efficient,	
Cepstral Analysis	Static feature Extraction method, Power spectrum	Used to represent spectral envelope
Mel-frequency scale analysis	Static feature extraction method, Spectral analysis	Spectral analysis is done with a fixed resolution along a subjective Mel scale.
Filter bank analysis	Filters tuned Required frequencies	
Mel-frequency cepstrum (MFCCs)	Power spectrum is computed by performing Fourier Analysis	
Kernel based Feature extraction method	Nonlinear transformation,	Reduction leads to better classification and it is used to remove noisy and redundant features, and improvement in classification Error
Wavelet	Better time resolution than Fourier Transform	It replaces the fixed bandwidth of Fourier transform with one proportional to frequency which allow better time resolution at high frequencies than Fourier Transform
Dynamic feature extractions 1)LPC 2)MFCCs	Acceleration and delta coefficients i.e. II and III order derivatives of normal LPC and MFCCs coefficients	
Spectral subtraction	Robust Feature extraction method	
Cepstral mean subtraction	Robust Feature extraction	
RASTA filtering	For Noisy speech	
Integrated Phoneme Subspace method	A transformation based on PCA+LDA+ICA	Higher Accuracy than the existing methods

Table 1. Types of feature extraction

Performance constraints:

Accuracy and speed are major constraints should be considered to calculate performance of speech. Accuracy may be stated with **Word Error Rate** (WER), whereas speed is measured with respect to time. Other terms are **Single Word Error Rate** (SWER) and **Command Success Rate** (CSR).

VI. Summary

In the last few years, research work is progressed in speech recognition area. It is spurred worldwide.

Sr. No.	Past	Present(new)
1.	Template matching	Corpus-based statistical modeling, e.g. HMM and n grams
2.	Filter bank/spectral resonance	Cepstral features, Kernel based function, group delay functions
3.	Heuristic time normalization	DTW/DP matching

4	Distance -based methods	Likelihood based methods
5	Maximum likelihood approach	Discriminative approach e.g. MCE/GPD and MMI
6	Isolated word recognition	Continuous speech recognition,
7	Small vocabulary	Large vocabulary
8	Context Independent units	Context dependent units
9	Clean speech recognition	Noisy/telephone speech recognition
10	Single speaker recognition	Speaker-independent/adaptive recognition
11	Read speech recognition	Spontaneous speech recognition
12	Single modality(audio signal only)	Multimodal(audio/visual)speech recognition
13	Hardware recognizer	Software recognizer
14	Speech signal is assumed as Quasi stationary. The feature vectors are extracted using FFT and wavelet	Data driven approach does not possess this assumption i.e. signal is treated as nonlinear And non-stationary. In this features are extracted using Hilbert Haung Transform using IMFs.

Table 2. Summary of speech recognition

VII. CONCLUSION

Speech recognition is widely used everywhere by now at the places where the security is an important issue so there are various techniques available that are used. Among all these algorithms the Mel Frequency Cepstrum Coefficient (MFCC) has efficient results that can be considered while performing speech recognition process.

VIII. FUTURE SCOPE

With this paper it can be concluded that future machines will be friendlier with humans to perform different tasks. This paper can be further developed and can be implemented in the public offices like banking sector, government services etc. to bridge the miscommunication gap between disabled and normal people. With the results as speaker identification various algorithms can be developed and they can be applied to operate devices to work on them in the future. Speech can be used as unique security feature for future innovations due to its complex nature. Human-machine interaction will be the key factor for future automation processes and industries.

ACKNOWLEDGMENT

We would like to express our special thanks of gratitude to Prof. S.B. Dhonde for their valuable guidance.

REFERENCES

[1] Yannick L. Gweth, Christian Plahl and Hermann Ney “Enhanced Continuous Sign Language Recognition using PCA” at 978-1-4673-1612-5/12 ©2012 IEEE.

[2] M.A.Anusuya and S. K. Katti “Speech Recognition by Machine” presented at (IJCSIS).

[3] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman “Speaker Identification Using Mel Frequency Cepstral Coefficients” ICECE 2004, December 2004.

Chaitanya Joshi, B.E. (Electronics Engineering), All India Shri Shivaji Memorial Society’s Institute of Information Technology, University of Pune, Pune.

Kedar Kulkarni, B.E. (Electronics Engineering), All India Shri Shivaji Memorial Society’s Institute of Information Technology, University of Pune, Pune.

Sushant Gosavi, B.E. (Electronics Engineering), All India Shri Shivaji Memorial Society’s Institute of Information Technology, University of Pune, Pune.

Prof. S.B. Dhonde, Assistant Professor, Department of Electronics Engineering, All India Shri Shivaji Memorial Society’s Institute of Information Technology, University of Pune., Perusing Ph.D. from Dr. Babasaheb Aurangabad University. Have 13 year experience in teaching and industry., Published several papers in National/International Journals.