ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

Grammatically Correct Speech by Using Speech Recognition

Vrushali Bhamare[1], Priyanka Kalokhe[2], Ketaki Kulkarni[3]

U.G. Student, Department of Computer Engineering, GESCOE Engineering College, Nashik, India[1]

U.G. Student, Department of Computer Engineering, GESCOE Engineering College, Nashik, India[2]

U.G. Student, Department of Computer Engineering, GESCOE Engineering College, Nashik, India[3]

ABSTRACT: It is often seen that people from vernacular medium of education does not have command on English language. So, they face tremendous difficulties in delivering fluent, grammatically correct and confident speeches. Typical grammar checking software use some form of natural language parsing to determine if errors exist in text. If a sentence is found ungrammatical, the grammar checker usually seeks a single grammatical error an explanation. Our program detects and explain grammatical mistakes made by non-native English speakers. The process will be done in two steps- speech recognition followed by grammar correction. The algorithm used for the speech recognition is Hidden Markov Model. In which Forward algorithm useful for isolated word recognition, Viterbi algorithm useful for continuous speech recognition and Forward-backward algorithm useful for training HMM. The algorithm used over here is rule - based checking for grammar correction. In old days, Grades (Grammar Diagnostic Expert System) is envisioned as a tool to help non-native English speakers learn to correct their English mistakes. Since the 1930s, speech recognition has been approached progressively, from a simple machine that responds to a small set of sounds to a sophisticated system that responds to fluently spoken natural language. Our aim is to provide the application which can be used on both the platform which will be very helpful for the people who are from the vernacular background.

KEYWORDS: Speech recognition, Hidden Markov Model

I. INTRODUCTION

Speech Recognition:

Speech recognition is a topic that is very useful in many applications and environments in our daily life. Generally, speech recognizer is a machine which understand humans and their spoken word in some way and can act thereafter.

Speech signal primarily conveys the words or message being spoken. Area of speech recognition is concerned with determining the underlying meaning in the utterance. Success in speech recognition depends on extracting and modeling the speech dependent characteristics which can effectively distinguish one word from another. The speech recognition system may be viewed as working in four stages as shown in Fig. 1

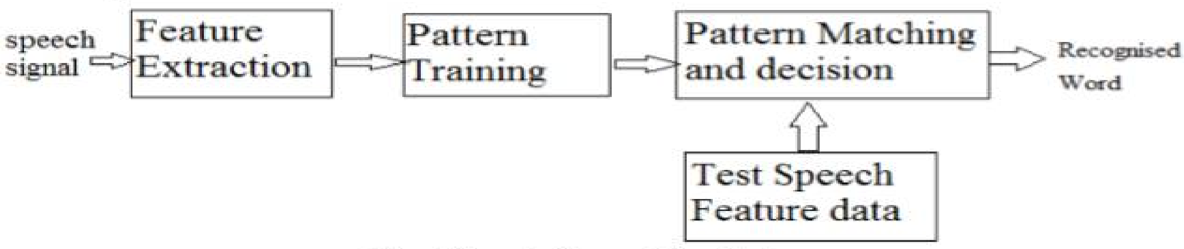

Fig. 1 Speech Recognition System

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

The feature extraction process is implemented using Mel Frequency Cepstral Coefficients (MFCC) in which speech features are extracted for all the speech samples. Then all these features are given to pattern trainer for training and are trained by HMM to create HMM model for each word. Then viterbi decoding will be used to select the one with maximum likelihood which is nothing but recognized word.

II. RELATED WORK

In 1952 three Bell Labs researchers built a system for single-speaker digit recognition. Their system worked by locating the formants in the power spectrum of each utterance. The 1950s era technology was limited to single-speaker systemswith vocabularies of around ten words.

Gunnar Fant developed the source-filter model of speech production and published it in 1960, which proved to be a useful model of speech production.

During the late 1960s Leonard Baum developed the mathematics of Markov chains at the Institute for Defence Analysis. A decade later, at CMU, Raj Reddy's students James Baker and Janet M. Baker began using the Hidden Markov Model (HMM) for speech recognition. Ames Baker had learned about HMMs from a summer job at the Institute of Defence Analysis during his undergraduate education. The use of HMMs allowed researchers to combine different sources of knowledge, such as acoustics, language, and syntax, in a unified probabilistic model.

In the early 2000s, speech recognition was still dominated by traditional approaches such as Hidden Markov Models combined with feedforward artificial neural networks.Today, however, many aspects of speech recognition have been taken over by a deep learning method called Long short-term memory (LSTM), a recurrent neural network published by Sepp Hochreiter&Jürgen Schmidhuber in 1997.LSTM RNNs avoid the vanishing gradient problem and can learn "Very Deep Learning" tasks that require memories of events that happened thousands of discrete time steps ago, which is important for speech. Around 2007, LSTM trained by Connectionist Temporal Classification (CTC)started to outperform traditional speech recognition in certain applications.

In 2015, Google's speech recognition reportedly experienced a dramatic performance jump of 49% through CTC-trained LSTM, which is now available through Google Voice to all smartphone users.

HMM Based Speech Recognition

A1.Pre-emphasis - In order to flatten speech spectrum, a pre-emphasis filter is used before spectral analysis. Its aim is to compensate the high-frequency part of the speech signal that was suppressed during the human sound production mechanism.

The most used filter is a high-pass FIR5 filter described in Eq. and whose transfer function corresponds.

$H_{preem}(Z)=1-a_{preem}z^{-1}$

A2. Windowing and frame formation - The speech signal is divided into a sequence of frames where each frame can be analyzed independently and represented by a single feature vector. Since each frame is supposed to have stationary behavior, a compromise, in order to make the frame blocking, is to use a 20-25 ms window applied at 10 ms intervals (frame rate of 100 frames/s and overlap between adjacent windows of about 50%), as Holmes & Holmes exposed in 2001. In order to reduce the discontinuities of the speech signal at the edges of each frame, a tapered window is applied to each one. The most common used window is Hamming window [2].

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi(n-1)}{\mathrm{N}-1}\right)$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

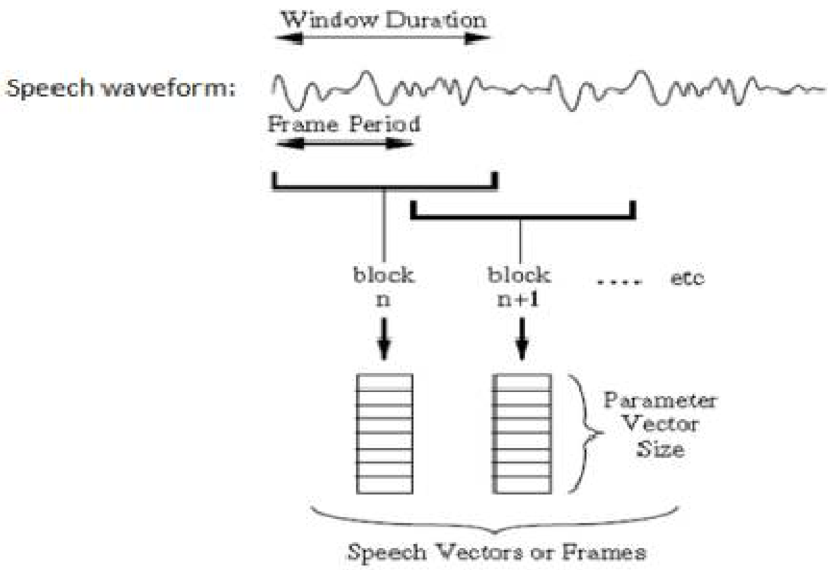

Fig. 2- Windowing and Frame Formation

Feature Extraction Techniques –

1. Speech Endpoint Detection - In the process of speech recognition, When the system receives a signal containing voice, system will detect and locate speech endpoint, removal of excess noise before and after the speech, complete voice will be submitted to the next level recognition. Voice endpoint detection algorithm is mainly based on the energy of the voice,zero crossing rate, LPC coefficients, information entropy, cepstral, band variance and so on. The endpoint detection effects and the actual environmental noise has a great relationship, Therefore, the endpoint detection of pre - denoising can improve the recognition rate. We introduce a traditional detection methods based on short -term energy, short-time zero-crossing rate.

2. Character Parameter Extraction- Sub-frame and endpoint detection is complete, the next feature extraction parameters. Feature extraction amount is theeffective characteristics of the signal is extracted from the speech signal, but also try to remove the noise informat ion of the speech signal, to improve the accuracy of identification. Since voice having a short-time characteristic, the speechcharacteristic parameters by frame information extract ion, frame feature vector. A voice after feature extraction, into a vector sequence. This vector sequence to train for and then some kind of model for speech recognition voice template. Voice characteristic parameter of extract ion is very important, directly affects the accuracy of the speech recognition.

A good speech features to meet the requirements of three:
(1)Can effectively extract the signal characteristics of the speech, including the channel characteristics of the human auditory model;
(2)Good independence between the order parameter;
(3)The characteristic parameters have an efficient method of calculating.

A3. MFCC Extraction- Davis &Mermelstein (1980) pointed the Mel Frequency Cepstrum6 Coefficients (MFCC) representation as abeneficial approach for speech recognition (Huang et al., 2001).[3] The MFCC is a representation of the speech signal defined as the real cepstrum of a windowed short -time signal derived from the FFT of that signal (Huang et al, 2001) which, is first subjected to a log-based transform of the frequency axis (mel-frequency scale), and then decorrelated using a modified Discrete Cosine Transform (DCT-II). Figure illustrates the complete process to extract the MFFC vectors from the speech signal. It is to be emphasized that the process of MFCC extraction is applied over each frame of speech signal independently. After the pre-emphasis and the frame blocking and windowing stage,

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

the MFCC vectors will be obtained from each speech frame. The process of MFFC extraction will be described below considering in any instant that all the stages are being applied over speech frames.

The first step of MFCC extraction process is to compute the Fast Fourier Transform (FFT) of each frame and obtain its magnitude. The FFT is a computationally efficient algorithm of the Discrete Fourier Transform (DFT). If the length of the FFT, is a power of two (K=2n), a faster algorithm can be used, so a zero-padding to the nearest power of two within speech frame length is performed. The next step will be to adapt the frequency resolution to a perceptual frequency scale which satisfies the properties of the human ears (Molau et al., 2001), such as a perceptually mel-frequency scale. This issue corresponds to Mel filterbank stage.
The last step involved in the extract ion process of MFCC is to apply the modified DCT to the log -spectral-energy vector, obtained as input of melfilterbank, resulting in the desired set of coefficients called Mel Frequency Cepstral Coefficients.

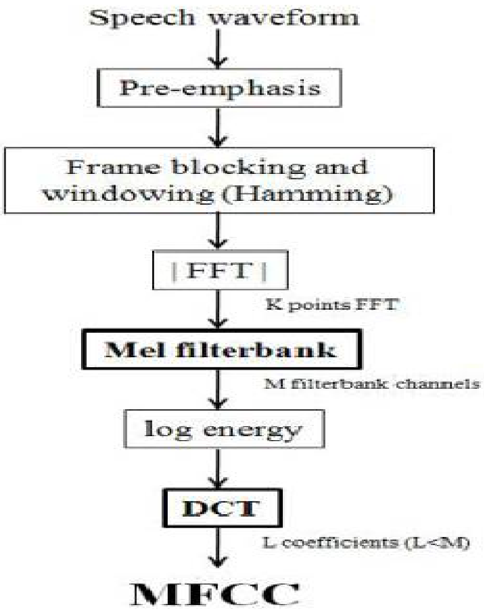

Fig. 4 – MFCC coefficient generation

HIDDEN MARKOV MODEL –
The problems associated with HMM are explained as follows:
(a) Evaluation: Evaluation is to find probability of generation of a given observation sequence by a given model. The recognition result will be the speech unit corresponding to the model that best matches among the different competing models. Now to find $P(O \mid \lambda)$, the probability of observation sequence $O = (o1, o2, \ldots, ot)$ given the model λ i.e .$P(O \mid \lambda)$.
(b) Decoding: Decoding is to find the single best state sequence, $Q = q1, q2, \ldots, qt$, for the given observation sequence $O = (o1, o2, \ldots, ot)$. Consider δ t *(i)* defined asthat is δ t*(i)* is the best score along single path at time *t*, which accounts for the *t* observations and ends in state *i*. byinduction,
(c) Training (Learning): Learning is to adjust the model parameters (A, B, π) to maximize the probability of the observation sequence given the model. It is the most difficult task of the Hidden Markov Modeling, as there is no known analytical method to solve for the parameters in a maximum likelihood model. Instead, an iterative procedure should be used. Baum-Welch algorithm is the extensively used iterative procedure for choosing the model parameters. In this method, start with some initial estimates of the model parameters and modify the model parameters to maximize the training observation sequence in an iterative manner t ill the model parameters reach a critical value.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

(d)Identification

Use the Viterbi algorithm to dynamically find the hidden Markov model state transition sequence (i.e identify the results), the time complexity is far less than the total probability formula.

The Viterbi algorithm is widely used in the dynamic programming algorithm in the field of communication, the algorithm in speech recognition applications. The total probability formula, you can calculate the output probability of the system, but were unable to find an optimum state transition path. Using the Viterbi algorithm can be found not only a good enough state transition path and the path corresponding to the output probability can also be obtained.

Meanwhile, with the Viterbi algorithm to calculate the output probability of the amount of computation required is much smaller than the amount of calculation of the total probability formula.

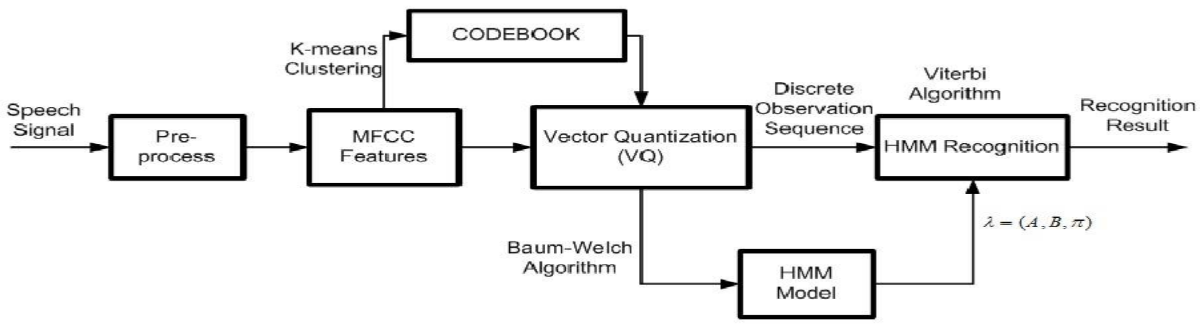

Fig. 5 – Speech recognition using HMM model

Grammar Correction Technique -

1. Rule-based Checking Algorithm - In this approach, a set of rules is matched against a text which has at least been POS tagged. It has many advantages:

1) A sentence does not have to be complete to be checked, instead the software can check the text while it is being typed and give immediate feedback.

2) It is easy to configure, as each rule has an expressive description and can be turned on and off individually.

3) It can offer detailed error messages with helpful comments, even explaining grammar rules.

4) It is easily extendable by its users, as the rule system is easy to understand, at least for many simple but common error cases.

5) It can be built incrementally, starting with just one rule and then extending it rule by rule.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

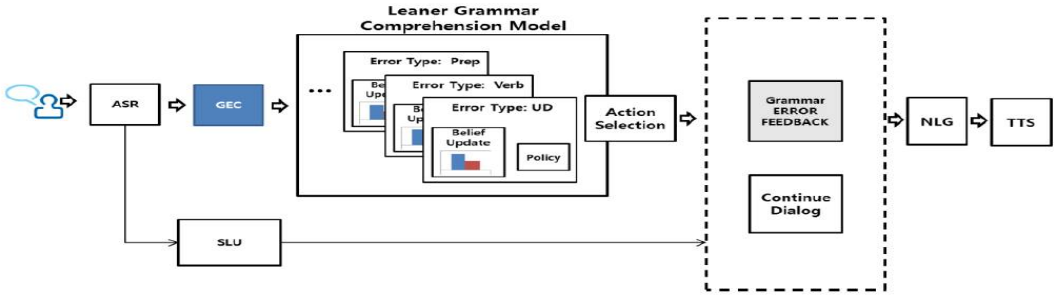

Fig. 6 – Grammar Checker Model

III .PROPOSED ALGORITHM

Proposed Approach:

- Our proposed approach is to give a grammar corrected speech from the input speech.
- The HMM algorithm used gives good and consistent result as compare to the other algorithms.
- The Grammar error correction model gives pretty accurate results.
- This system will surely help those who lacks in English grammar.

Mathematical Model :

The central outline of the proposed algorithm is as follows.

- Step 1: First the speech signal is given.
- Step 2: The pre-processing step is done which include silence removal, pre- emphasis, framing and windowing
- Step 3: Feature extraction is done by MFCC.
- Step 4: Then the HMM algorithm is applied to recognize the speech on basis of probabilities.
- Step 5: After having the recognized speech the grammar error correction model is applied.
- Step 6: The corrected text is then given to the user as result.

IV.PSEUDOCODE

If the indefinite article is followed by a word whose pronunciation starts with a vowel sound, an has to be used instead of a. If it is one of a, e, i, o, u, the word probably starts with a vowel - but there are exceptions . Here are some examples where the a,e,i,o,u rule applies, together with the correct indefinite article: a test, a car, a long talk an idea, an uninteresting speech, an earthquake

Here are some exceptions: a university, a European initiative an hour, an honor

This type of correction was not possible to be implemented by writing rules as mentioned above. We have used two text files containing these exceptional words for „a" and „an" for checking .Thus we created our own algorithm which works as follows:

Rule-based checking Algorithm-

Create a string array words[] to store space delimited words of a sentence.
for i=0 to array.length-2 do

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

if

word precedes with ‘a’ and starts with a vowel and if it is not exception then replace ‘a’ by ‘an’.

if

word precedes with ‘a’ and is an exception then replace ‘a’ by ‘an’.

if

word precedes with ‘an’ and it does not start a vowel and if it is not exception then replace ‘an’ by ‘a’.

if

word precedes with ‘an’ and is an exception then replace ‘an’ by ‘a’.

V. RESULTS

Speech recognized correctly with textual format and correct grammar by grammar correction algorithm

VI .CONCLUSION AND FUTURE WORK

This paper proposed a novel confidence measure forspeech recognition error correction. This measure can acquire syntactical and semantic information in a fixedwindows size.

The conclusion of this study of recognition and hidden markov model has been carried out to develop a voice based user machine interface system. These application can be related with disable persons those are unable to operate computer through keyboard and mouse, these type of persons can use computer with the use of Automatic Speech Recognition system, with this system user can operate computer with their own voice commands .

REFERENCES

[1] R.Cardin, Y. Normadin and E. Millen, “Inter-word coarticulation modeling and MMIE training digit recognition “,IEEE,ICASSP,Vol. II, pp.243-246-199,1993.

[2] Hong Ye,Youzheng Zhang and Jianwei Shen,” Study on Speech Recognition ion of Greeting Based on Biometric Pattern Recognition”,IEEE, 2nd InternationalWorkshop on Intelligent Systems and Applications (ISA), Wuhan, May 2010.

[3] Amita dev and Bansal poonam,” Robust Features for Noisy Speech Recognition using MFCC Computation from Magnitude Spectrum of Higher Order Autocorrelation Coefficients”,International Journal of Computer Applications, Number 8 - Article 7, pp.36-38, November2010.

[4]Xinguang li, Jiahua, chen, Zhenjiang li,” English sentence recognition based on HMM and clustering”, American Journal of Computational Mathematics, Vol. III, No.1, pp.38-42, 2013.

[5] MS. Vrinda,Mrs.Chander chekhar,” Speech Recognition System for English Language”,International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, pp. 919-922, January 2013.

[6] Kyusong Lee, Seong Han Ryu, Paul Hongsuck Seo, Seokhwan Kim, Gary Geunbae Lee,”Grammatical error correction based on linear compressionmodel inoral conversion”,IEEE,Spoken Language Technology(SLT), Vol-23,2014.

[7] Vidwath R Hebse, Anitha G,” The Learning Method of Speech Recognition Based on HMM “, International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, pp.971-977, 2015.

[8] Su Myat Mon, Hla Myo tun,” Speech to text conversion system using HMM “, International Journal of Scientific & Technology Research, Vol. 4,Issue 06, pp.350-352,June 2015

[9] Alaa Hassan Mahmoud, Salma Alzaki Ali et al "Speech to text conversion", SUST, 2014.

[10] Pahini A. Trivedi, ” Introduction to Various Algorithms of Speech Recognition: Hidden Markov Model, Dyanamic Time Warping and Artificial Neural Network,”, International Journal of Engineering Development and Research (IJEDR), ISSN:2321-9939, Volume.2, Issue 4, pp.3590-3596, December 2014

[11] Miss. Prachi Khilari et al ,”Pattern recognition approach”, International Journal of Advanced Research in Computer Engineering & Technology(IJARCET),Vol. 4, Issue 7, July 2015.

[12] John Lee, Stephanie Seneff, “An Analysis Of Grammatical Error In Non-Native Speech in English”,IEEE Workshop on Spoken Language Technology, SLT, pp.89-92, 2008

[13] Kelichi Tokuda, Tomoki Toda,” Speech Synthesis Based on Hidden Markov Models”,Proceedings of the IEEE, Vol. 101, No. 5, pp. 1234－1252, May 2013

[14] Mark Gales, Steve Young,” The Application of Hidden Markov Models in Speech Recognition”, Journal Foundations and Trends in Signal Processing, Vol. 1, Issue 03 ,January 2008.

[15] Youssef Bassil, Mohammad Alwani ,”Post-Editing Error Correction Algorithm For Speech Recognition Using Bing Spelling Suggestion”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3,
No.2, pp.95-101,2012.

[16] Dr Suresh D, Anu L B, Sanjeev Kubakaddi,” Person Identification Using MFCC and Vector Quantization “,IPASJ International Journal of Electronics & Communication (IIJEC), Vol.3, Issue 6, pp.19-24, June 2015.