ISSN (Online) : 2455 - 3662
SJIF Impact Factor :5.148

EPRA International Journal of

Multidisciplinary Research

Monthly Peer Reviewed & Indexed
International Online Journal

Volume: 5 Issue: 3 March 2019

Published By :EPRA Publishing

EPRA International Journal of
Multidisciplinary Research (IJMR) *Peer Reviewed Journal*

MULTI-MODAL INTERFACED DATA ANALYSIS SYSTEM

JayashreeChelladurai C
Assistant Professor,
Dr.Mahalingam College of Engineering and Technology,
Pollachi, Coimbatore,
Tamil Nadu, India

Preethi M
Student, Dr.Mahalingam College of Engineering and Technology,
Pollachi, Coimbatore,
Tamil Nadu,
India

Nandhini A
Student, Dr.Mahalingam College of Engineering and Technology,
Pollachi, Coimbatore,
Tamil Nadu,
India

Rohith T
Student, Dr.Mahalingam College of Engineering and Technology,
Pollachi, Coimbatore,
Tamil Nadu,
India

ABSTRACT

***Multi** modal interaction is a form of human-machine interaction using multiple modes of input and output channels. Multimodal interaction includes input channels such as keyboard text, speech recognition, haptic perception, etc., and output channels such as visual output, auditory output etc. The proposed system involves speech and text for input recognition and graphical representation of report for output generation.*

Speech is a human vocal communication using the phonetic combinations of words in definite meaning. Human voice varies from people to people accordingly. Speaker– independent recognition system is designed to recognize the speech of the user irrespective of the gender, age, style, social & cultural background, voice, texture and volume of speech. The corpus includes heterogeneity of data with variety of speakers to enhance the recognition rate of a Speaker Independent Recognition System.

The proposed system is for small vocabulary of speech recognition concerning text output for the given speech input. Speech recognition is process of converting acoustic signal captured by a microphone to textual representation. The contiguous speech has been recognized from the dataset trained for the given system.

Data capturing is a field that requires frequent text entry operations, which can be alternatively replaced by a speech recognition system. The recognized data are processed and the information is stored in a disk and the reports can be generated from the data in required format. Text entry is a time-consuming process, alternative ways to make the process more efficient by speech recognition. The data is recognized and stored in a database for easy retrieval of data. The stored data is analyzed and the result can be viewed by graphical representation such as pie chart, bar graph for easier understanding. The report is generated in user required format such as pdf, excel.

KEYWORDS: ***Gaussian Mixture Model, Hidden Markov Model, Speech Recognition***

1. INTRODUCTION

Speech is the most natural and common way of communication between people. The automatic speech recognition is the technology that allows human beings to use their voices to speak with a computer interface in a way that in its most sophisticated variations, resembles normal human conversation. Accent will differ between speakers. Changes in the physiology of the organs of speech production will produce variability in the speech waveform. The speech signal will also vary considerably according to emphasis or stress on words. Environmental or recording differences also change the signal. The data may come from hand-written forms or be audio files.

In this system, data entry is done through keyboard and speech where data entry is a field that requires frequent text entry operations which can be alternatively replaced by a speech recognition system. it can allow documents to be created faster because the software generally produces words as fast as they are spoken which is generally much faster than a person can type.

The recognized data is stored in the database and it is processed to retrieve the information from stored disk and the reports can be generated from the data in needed format such pdf, excel.

2. METHODOLOGY

The proposed system is explained with different phases. The login validation is used to accept username and password from the user and validates the input with the data stored in the database. If valid data are given user can access the further pages else the page shows an error.

The selection of batch/department/semester is used to navigate through the required needs such as subject faculty mapping, biodata, sections, course details and mark entry.

The mark entry is done by speech recognition with several models such as the Hidden Markov Model for speech recognition and Gaussian Mixture Model for feature extraction.

Finally, the stored data is analyzed and the result with several analysis is sent to email of the faculty members. They can also fetch the data with their convenience and the data can be prioritised based on the marks of the student. That can be viewed by graphical representation such as pie chart, bar graph for easier understanding. The report is generated in user required format such as pdf, excel.

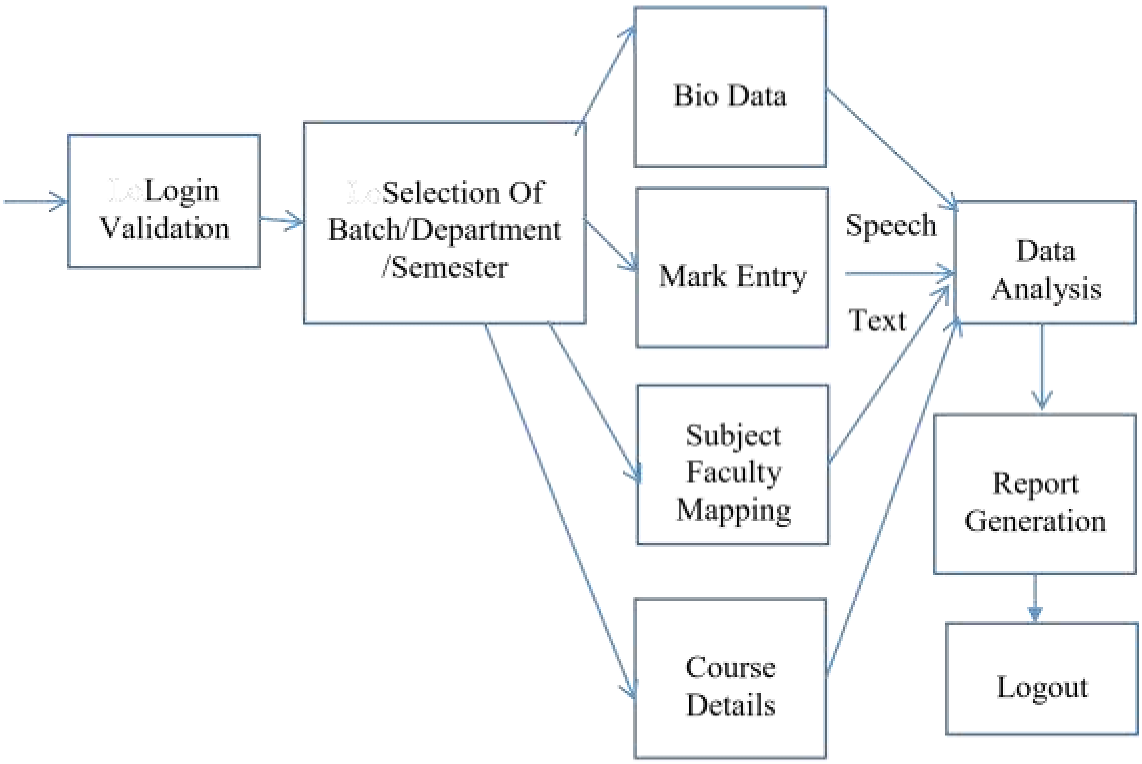

Figure 1 Block Diagram Of Proposed System

2.1 Hidden Markov Model

The Hidden Markov Model, along with the Gaussian Mixture Model, is used to recognize speech from the user. In markov model the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters, while in the hidden markov model, the state is not directly visible, but the output dependent on the state, is visible. Size of the set is depends on the nature of the observed variable. If the observed variable is discrete with M possible values, governed by a categorical distribution, there will be a M-1 separate parameters, for a total of N(M-1) emission parameters over all hidden states, there will be M parameters controlling the covariance matrix for a total of,

$$N\left(M+\frac{M(M+1)}{2}\right)=\frac{NM(M+3)}{2}=O(NM^2)$$

2.1.1 Speech input for HMM systems

Implementing a HMM for speech recognition makes the assumption that the features can be broken up into a series of quasi-stationary discrete segments. The segments are treated independently and in isolation. The frame rate must be sufficiently large such that the speech is roughly stationary over any given frame. Speech features are usually based upon the short-term Fourier transform of the input speech.

2.1.2 Recognition units

For very small vocabulary recognition tasks, it would be possible to build a HMM model for each word. However, this presents problems of identifying adequate HMM topologies and establishing the optimal number of states for each word. In addition, with a medium or large vocabulary there will be insufficient data to robustly estimate parameters for each whole word model. The most commonly used approach is to split words up into smaller subword units, such as syllables or phones.

2.2 Gaussian Mixture Model

Gaussian Mixture Model is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population. Gaussian Mixture Model have been used for feature extraction from speech data. Feature Extraction, a component of the front end speech processing block, is the process of extracting unique information from voice data. A voice feature is a matrix of numbers in which each number represents the energy or average power that is in a particular frequency band during a specific interval of a speech signal. Gaussian mixture model is parameterized by the mean factors, covariance matrices and mixture weights from all component densities.

2.2.1 Speaker adaptation

An approach is to apply feature mean normalisation to the formant features. This approach removes any linear shift from each formants and has been likened to a vocal tract normalisation transform. The assumption made is that each formant should be distributed about its mean for a given utterance. Any shift on the mean position is assumed to be an inter speaker variation based on the change in vocal tract length, and can thus be removed. This approach is a linear subtraction and will have different effects from the application of a scaling factor to the extracted parameters. One difference is that observed range of the formants for a given speaker will be unchanged when a linear shift is applied, whereas using a linear scale would compress or expand the effective range of the formants. Applying feature variance normalisation could possibly compensate for this effect if it is an issue. As the means are shifted up for a speaker, the variances will also be expected to increase as the separation of the peaks increases.

3. RESULTS

3.1 Evaluation Metric

The networks predict floating point numbers, which were rounded off to the nearest integer. In the OEIS dataset, to get a better picture of the accuracy, the count of sequences which differed upto +/- 10 from the actual numbers were also recorded. For each network architecture, the number of sequences solved is used as the evaluation metric.

3.2 Speech Accuracy Rate

An HMM recognizer with ten-state, five-mixture/state models trained with the traditional ML method achieved an accuracy of 89% for the training data set and 76% for the test set.

$$WER=\frac{S+D+I}{N}=\frac{S+D+I}{S+D+C}$$

- D is the number of deletions,
- I is the number of insertions,
- C is the number of correct words,
- N is the number of words in the reference (N=S+D+C)
- S is the number of substitution

4. CONCLUSIONS

Thus, simple data entry system with speech recognition is obtained by neural network models. The presence of hidden markov model process with unobserved (*hidden*) states improves the performance of the speech recognition and Gaussian Mixture model

(GMM) improves the performance in feature extraction. An HMM recognizer with ten-state, five-mixture/state models trained with the traditional ML method achieved an accuracy of 89% for the training data set and 76% for the test set. The Hidden Markov Model, along with the Gaussian Mixture Model, is used to recognize speech from the user.

REFERENCES

1. Shivakumar K M , Varsha V Jain and Krishna Priya P, "A study on impact of language model in improving the accuracy of speech to text conversion system", International Conference on Communication and Signal Processing, April 2017
2. N.Uma Maheswari, A.P.Kabilan, R.Venkatesh "Speaker Independent Speech Recognition System Based On Phoneme Identification" ,International Conference on Computing Communication and Networking,2008.
3. Colleen.G ,Le Prell , "Effects Of Noise On Speech Recognition: Challenges For Communication By Service Members", Hearing research,2017.
4. AlexGraves,NavdeepJaitly,"Towards End-To-End Speech Recognition With Recurrent Neural Networks",International Conference on Machine Learning,2014.
5. Has im Sak, Andrew, Kanishka Rao, Franc¸oise Beaufays,"Fast And Accurate Recurrent Neural Network Acoustic Models For Speech Recognition", Google, 2015.
6. Jan chorow,Dzmitry Bahdanau,Kyunghyun Cho,"End-to-End Continuous Speech Recognition Using Attention-Based Recurrent NN:First Results", arXiv:1412.1602v1 [cs.NE],2014.
7. Liang Lu, Xingxing Zhang, Kyunghyun Cho, and Steve Renals,"A Study of the Recurrent Neural Network Encoder-Decoder for Large Vocabulary Speech Recognition",ISCA,2015.
8. Xiangang Li, Xihong Wu,"Constructing Long Short-Term Memory Based Deep Recurrent Neural Networks For Large Vocabulary Speech Recognition",Speech and Hearing Research Center,2015.
9. Michiel Bacchiani, Andrew Senior, Georg Heigold,"Asynchronous, Online, GMM-Free Training Of A Context Dependent Acoustic Model For Speech Recognition",ISCA,2014 .
10. Yanmin Qian,Mengxiao Bi,Tian Tan,and Kai Yu, "Very Deep Convolutional Neural Networks for Noise Robust Speech Recognition",IEEE/ACM transactions on audio, speech, and language processing, vol. 24, no. 12, December 2016.