International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 4, Issue 2 (March-April, 2016), PP. 331-333

SPEECH TO TEXT CONVERSION FOR CHEMICAL ENTITIES: REVIEW

Farhaan Kaleem, Shruti Kanchan, Pradnya Kalbhor, Aditya Kakde
Department of Computer Engineering, Savitribai Phule Pune University
Pune, Maharashtra, INDIA
farhaan.kaleem@gmail.com

Abstract— **In last many years, work has been done in audio processing. However it was not much used in the fields of Electronics and Computers due to its variety of speech signals and complexity. But now with the help of modern algorithms, it is possible to easily recognize the text from the given speech. In this project, we will develop an application, which will tkae the speech as input and give the text which will be specifically related to chemical names. It can also be further extended to provide a baase for various applications like when the text is recognized it can give all the related information of that chemical. We know that there are applications already developed for audio processing and extracting text but they do not work that accurately with chemicals. Hence we make this appliction specially for chemical entities. Here, we use the technique based on Hidden Marcov Model (HMM).**

Index terms- **HMM, Phoneme, Chemicals, Speech Recognition.**

I. INTRODUCTION

Speech recognition is the process of capturing spoken words using microphone or telephone and converting them into a digitally stored set of words. Speech to text conversion is one of the application of speech recognition. The main goal or objective of this paper is how to recognize the various sound in speech and provide the output as the text. Accuracy or correctness of speech recognition depends on various factors such as the size of the vocabulary, mode of speech like isolated, continuous, etc. To recognize different voice patterns, a technique based on Hidden Marcov Model (HMM) is used, as it is one of the mot efficient algorithm for speech recognition[1].

Speech recognition system can be divided into different modules such as: Speech Acquistion, Speech Preprocesing, Hidden Marcov Model and Text Storage.

Finally we can extend this application to extract the chemical related text from the speech and provide the information about that chemical entity.

II. RELATED WORK

There are some speech recognition techniques such as Natural Language Processing (NLP), etc. The idea from some of the base papers is used to extract text from the speech. Due to various reasons like variety of sounds in the speech, it becomes difficult to recognize the text from the input audio.

In past, many scientists were doing research on text extraction from input audio signals. As a result many algorithms were developed. Natural Language Processing (NLP) techniques can be used for this proces. However these techniques do not work that effficiently when it comes to the chemical names. Therefore in this paper, we specifically stress on the chemical entities.

The microphone input port with the audio codec receives the audio signal and produces the output as the text.

III. METHOD TO EXTRACT TEXT FROM INPUT AUDIO SIGNAL

A. Overview

In this paper, the speech is taken as the input through microphone which is provided to the speech acquistion module. Then that speech is provided to the speech preprocessing module. Then the operations of Hidden Marcov Model (HMM) is applied on the input audio to finally obtain the desired text as the output.

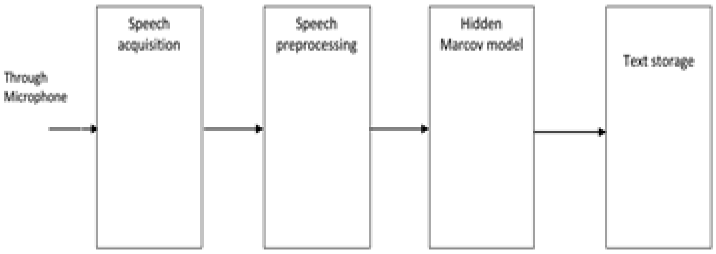


Fig. 1.1 Flowchart of Proposed Method

B. Speech Acquistion

In this module, the microphone input port with the audio codec receives the signal, amplifies it, and converts it into 16-bit PCM digital samples at a sampling rate of 8 KHz. Here the analog audio signal is sampled on time and amplitude axes for digitization of audio, so that further processing can be done on this digitized data. Speech signal is analyzed in even interval which is usually 20ms[2]. Because the speech sinal within this time interval is considered as stable. The final output of this module will be the digitized audio signal. This data is then stored in memory for further processing.

C. Speech Preprocessing

This digitized data is then given to the further module known as speech preprocessing, which extracts the features of the digitized audio signal like phonemes, unique patterns, etc.

Preprocessing involves taking the speech samples as input, and breaking the samples into frames, and returning a unique pattern for each sample. The unique pattern can be achieved based on certain basic factors like phonemes.

The unique pattern can be achived by following steps:-

1. The digital samples are divided into overlapped frames.
2. The system checks the frames for voice activity using endpoint detection and energy threshold calculations.
3. The speech samples are passed through a pre-emphasis filter.
4. The frames with voice activity are passed through a Hamming window. Hamming window returns only positive sample. It discards negative and zero values.
5. The system finds linear predictive coding (LPC) coefficients for frames .
6. From the LPC coefficients, the system determines the cepstral coefficients.

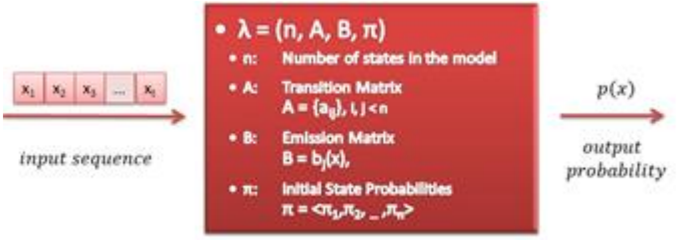
The cepstral coefficients serve as feature vectors.

We know that the frequency of speech of humans lies within specific range. An adult male will have a fundamental frequency from 85 to 180 Hz, and that of a typical adult female from 165 to **255 Hz**. Thus pre-emphasis filter is used to extract only that audio, which lies in this frequency range.

Linear predictive coding (**LPC**) is a tool used mostly inaudio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a **linear predictive** model.

D. Hidden Marcov Model (HMM)

Hidden Marcov Model (HMM) is an important alorithm, which is used to convert audio into the text. The input to this module is filtered diitized audio sinals, which is then converted into text using this module[2].



Hidden Markov Model[6]

HMM consists basically of three steps:-

1. Training
2. HMM-based recognition
3. Digit Models

1. *Training*

It is an imortant part of speech to text conversion. It involves putting all the speech which sounds similar into one class. So that in future any audio is easy to search from the existing database. Here the trainig is given to the database so that the searching is done easily from the database.

2. *HMM-based recognition*

It is the process of comparing each input speech to existing database. It involves comparison of unknown patterns to the existing sound class reference pattern and thus computing the similarity between them.

3. *Digit Models*

Digit Models are nothing but the probability of occurrence of certain events. HMM-based recognition is used to compare these probabilities with the digit models and the model with maximum probability is choosen as the recognized text.

E. Text Storage

This is the final module, which gets the input as the extracted text from the Hidden Marcov Module (HMM). This text is then stored into the suitable memory space.

IV. Mathematical Model

Let us consider a set S = { I, O, Fn, Su, F }
Where,
I => Input
I = {Audio or Speech}
O => Output
O = {Text which is extracted from the audio signal}

Su => Success
Su = {Proper word or sentence which is spoken is obtained}

F => Failure
F = {Unable to extract the patterns from the input audio signal and thus failure in giving the correct word as output in the form of text}

Fn => Functions
Fn = {speechAcquistion(), formFrames() }

speechAcquistion()
{
This function takes the audio signal as the input and using Pulse Codde Modulation can encode the input audio into 16-bit encoded data.
}
formFrames()
{

This function is used to form the input encoded audio signals into frames of size 960.

short pcm_[960];

}

Preprocessing()

{

It is used to extract the feature vector in the form of ceptral coeffecients

}

HiddenMarcov()

{

It is used to compare the words by using probability distribution and give the desired text as the output.

}

V. Conclusions

Therefore we conclude that Hidden Marcov Model (HMM) can be used to efficiently extract the text from the input audio. It is used to recognize the speech and thus give an output in textual format. The older techniques, were not able to efficiently extract chemical names. But HMM can efficiently extract the chemical names. And give the desired text as an output.

References

[1] B. Raghavendhar Reddy,, E. Mahender, Speech to Text Conversion using Android Platform. January -February 2013, Parvathapur, Uppal, Hyderabad, India.

[2] MJF Gales. Semi-tied full-covariance matrices for hidden Markov models. 1997.

[3] H. Hermansky. Perceptual linear predictive (PLP) analysis of speech. Journal of the Acoustical Society of America, 87(4):1738{1752,1990.

[4] B. Harb, C. Chelba, J. Dean, and G. Ghemawhat. Back-o_ Language Model Compression. 2009

[5] Maryam Kamvar and Shumeet Baluja. A large scale study of wireless search behavior: Google mobile search. In CHI, pages 701{709, 2006.

[6] Wikipedia