Grenze International Journal of Engineering and Technology, Special issue

Development of Interface System for Personal Details Enrollement

V.Hemavathi[1] and P.Shanmugapriya[2]
[1-2] Saranathan college of engineering /Department of ECE, Trichy, Indian
Email: hemavathi150496@gmail.com, Shanmugapriya-ece@gmail.com

***Abstract*—Speech processing plays an important role in digital signal processing. System-human interaction plays an important role for many users such as visually challenged persons, illiterate or literate persons in order to access information. The interactions could be in many ways such as speech, text, gestures etc. Among these interaction, speech is the natural means of communication for human beings. speech recognition system has been developed by various researchers in different languages for accessing the information such as student personal details enrollement in the English language. The aim of the proposed work is to developing and deploying speech interface for form filling applications which replace the traditional keyboard for entering the response. The proposed system uses the keyboard for entering the response. In this paper the system prompts the users, with queries are generated, using speech to text synthesis system. The synthesis systems recognize the user response using ASR system and fill the fields. With the help of this system we can achieve the different fields in the forms like Name, Age, Date of birth etc..The system able to recognize the response using speech recognition and recognized word is converted into text using speech to text conversion. This system can be helpful for rural public as well as the visually challenged people. In future this system can be implemented in the education for students personal detail maintenance.**

***Index Terms*— Application programming interface, speech interface system, ASR, speech based conversation System, web speech API.**

I. INTRODUCTION

Communication and information are often interchangeably, but they signify quite different things. Information is giving out and communication is getting through. The process by which meanings are exchanged between people through the use of common set of symbols is called communication. The basic building block of good communication is the feeling that every human being is unique and of value. Speech communication is the most effective mode of communication used by the humans. It is the transfer of information from one person to another via speech, which consists of variations in pressure coming from the mouth of the speaker. Such pressure changes propagate as waves through air and enter the ears of the listeners, who decipher the waves into a received message. Speech signals are usually processed in a digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signal.

Grenze ID: 01.GIJET.5.1.6

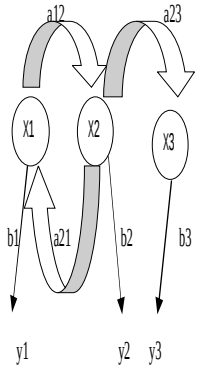
II. RELATED WORK

A. Feature extraction

The technique used for feature extraction is called Mel-Frequency Cepstral Coefficients. MFCC is used for to extract useful information from speech signal. MFC coefficients are based on known variations of human ear's critical bandwidth with frequencies. The important feature of Mel-filter bank is that the Mel-filters are spaced linearly at low frequencies and logarithmically at high frequencies to capture the phonetically important characteristics of speech.

B. Hidden markov model

Hidden Markov Model (HMM) is a doubly stochastic process so that it is not directly observable. This hidden stochastic process can be observed only through another set of stochastic processes which can produce the observation sequence. This technique is used to train a model for an utterance of a word. Later utterance is tested and calculating the probability of that the model has created the sequence of vectors. HMMs are widely used as acoustic models. It provides better performance than other methods. Initial value (π) finding is such a difficulty in HMM model and also states finding is also major problem.



X1, X2, X3 - No of possible input states.
Y1, Y2, Y3 - No of possible output states.
a12, a21, a23- Transition probability matrix
b1, b2, b3 - Emission probability matrix

There is a major difference between the Hidden Markov Model and observable markov model. In the HMM the state at each time (T) must be inferred from observations. But in the observable markov model the output state is completely determined at each time[4]. Observation is a probabilistic function of a state. Each hidden states are associated with the set of probabilities of emitting particular visible states. For training purpose the signal is used to ordinary utterances of the specific word whose MFCC matrix value is calculated. After find the MFCC value the threshold value is fixed for the utterance words. Testing is also done for the word utterance.

III. PROPOSED SYSTEM

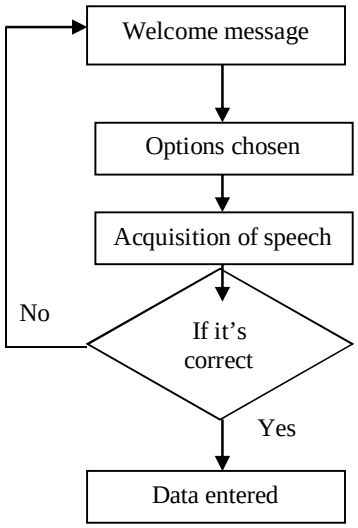
Automatic Speech Recognition (ASR) is initiated with speaker producing an utterance which consists of audio waves. The audio waves are recorded by a microphone and converted into electric signal which is again converted into digital signal in order to be understood by the speech system. The relevant information about the given utterance is extracted for accurate recognition. Finally, speech recognition system finds the best match by tuning the utterance. In other words, Automatic speech recognition is primarily used to convert spoken words into computer text. Additionally, automatic speech recognition is used for authenticating users via their voice (biometric authentication) and performing an action based on the instructions defined by the human. The proposed ASR system is modeled using some important blocks such as MFCC, HMM, WEB SPEECH API.

A. Flow Graph of Interface System

The fields names uttered are recorded through normal mic and saved as a wave (.wav) file. These recorded wave files are stored in a specified directory which is collectively called speech database. The speech signal from the

database is pre-processed and the MFCC features from the speech signals are extracted. After the feature extraction speech signal is given to the HMM model. In the HMM model for each word training &testing are done. After the HMM model the trained words are stored in the JSpeech Grammar Format. Using these dictionary required fields are filled.Field specific language models

- Name
- Age
- Date of birth
- Phone number
- Mail id



System: Welcome to online form filling system. To enter the Name say `1 ,to enter the Date of birth say 2, to enter Age say 3, and to end the task please say zero.

User: one
System: You have chosen one. To confirm it say yes or to deny say no.
User: Yes
System: Data for the field (Ex: Name)
User: Ragul
System: You have said Ragul. To confirm it say yes or to deny say no.
User: Yes
System: The systems finish the first field and verify do you want to continue?
User: No
System: Thank you

B. Web speech API

According to the Mozilla web document web speech API enables you to incorporate voice data into web apps. It has two parts.

- Speech synthesis
- Speech recognition

1. *Speech synthesis*
2. *Speech recognition*

Speechrecognition is accessed via the SpeechRecognition interface, which provides the ability to recognize voice context from an audio input (normally via the device's default speech recognition service) and respond appropriately. Generally you'll use the interface's constructor to create a new SpeechRecognition object, which

has a number of event handlers available for detecting when speech is input through the device's microphone. The SpeechGrammar interface represents a container for a particular set of grammar that your app should recognize. Grammar is defined using JSpeech Grammar Format (**JSGF**).
SpeechSynthesis interface, a text-t—speech component that allows program to read out their text content. Different voice types are represented by "speech synthesis voice" objects. you can get these spoken by paying them to the "speech synthesis.speak" method.

C. Requirements for web speech API

Web speech API has three requirements those are

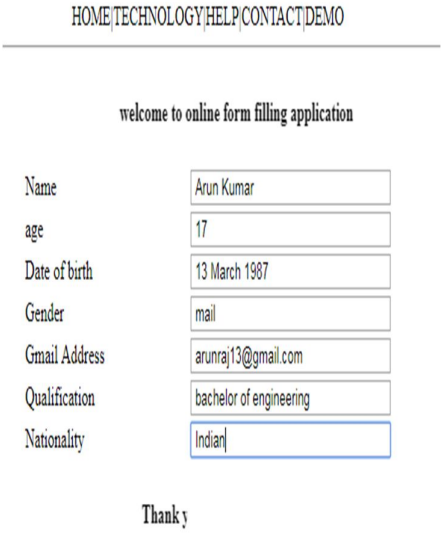
- Index.html (containing HTML)
- Style.css (containing css style)
- Index.js(containing JavaScript code)

IV. Result and Discussion

The evaluation of the performance of the speech recognition for the particular field is done. On testing the data, the following result is displayed. First the welcome message is initiated by the system. Then user selects the required field. The required field is the recognized.

The entire fields are displayed infront of the user

The first field is field using speech to text method



Similarly the entire fields are fields in the same manner

REFERENCES

[1] Ahmad A. M. Abushariah(1), Teddy S. Gunawan(2),"English Digits Speech Recognition System Based on Hidden Markov Models" International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia.

[2] Raja N. Ainon1, Roziati Zainuddin1, Othman O. Khalifa2" Human Computer Interaction Using Isolated-Words Speech Recognition Technology" International Conference on Intelligent and Advanced Systems 2007.

[3] Shweta Signas, Dr. Rajesh Kumar Dubey" Automatic Speech Recognition for Connected Words using DTW /HMM for English/ Hindi Languages" International Conference on Communication, Control and Intelligent Systems (CCIS)-2015.

[4] David Griol *, Lluı´s F. Hurtado, Encarna Segarra, Emilio Sanchis" A statistical approach to spoken dialog systems design and evaluation" Departament de Sistemes Informa`tics i Computacio´ , Universitat Polite`cnica de Vale`ncia, E-46022 Vale`ncia, Spain.

[5] Ms. Rupali S Chavan1, Dr. Ganesh. S Sable" An Overview of Speech Recognition Using HMM" International Journal of Computer Science and Mobile Computing-2013.

[6] Md. Sahidullah , Member, IEEE, Dennis Alexander Lehmann Thomsen, Rosa Gonzalez Hautam¨aki , Tomi Kinnunen," Robust Voice Liveness Detection and Speaker Verification Using Throat Microphones" ieee/acm transactions on audio, speech, and language processing, vol. 26, no. 1, january 2018.