Virtual personal assistance

IOP Conf. Series: Materials Science and Engineering **263** (2017) 052022 doi:10.1088/1757-899X/263/5/052022

Virtual personal assistance

Aditya K, Biswadeep G, Kedar S and S Sundar
School of Electronics Engineering, VIT University, Vellore 632014, Tamil Nadu, India

E-mail: sundar.s@vit.ac.in

Abstract, Human computer communication has growing demand recent days. The new generation of autonomous technology aspires to give computer interfaces emotional states that relate and consider user as well as system environment considerations. In the existing computational model is based an artificial intelligent and externally by multi-modal expression augmented with semi human characteristics. But the main problem with is multi-model expression is that the hardware control given to the Artificial Intelligence (AI) is very limited. So, in our project we are trying to give the Artificial Intelligence (AI) more control on the hardware. There are two main parts such as Speech to Text (STT) and Text to Speech (TTS) engines are used accomplish the requirement. In this work, we are using a raspberry pi 3, a speaker and a mic as hardware and for the programing part, we are using python scripting.

1. Introduction

As the world is moving towards computing utopia, people are being more consensus towards the virtual world. Now a day's people are using more and more virtual tools for designing, i.e., from small circuits to large networks are made virtually for testing before implementation, simulations are down virtually. So, virtual reality has become a part of our life.

One of the most used tool of the virtual is an artificial intelligence. An artificial intelligence is the science of making an intelligent machine, or a computer controlled robots, or a software think intelligently, in the similar manner the intelligent human think. The core contributors in the development of an artificial intelligence are Computer Science, Biology, Psychology, Linguistics, Mathematics, and Engineering. A major thrust of AI is in the development of computer functions associated with human intelligence, such as reasoning, learning, and problem solving. Artificial intelligence has been dominant in various fields such as gaming, natural language processing and so on.

Speech processing has been an area of interest for the last four decades, but the last decade has witnessed significant progress in this area of research. This progress has been possible mainly due to the recent advantages made in the powerful and efficient speech processing techniques such as vector quantization, hidden Markov Model etc. Here we have used the Markov Model for predicting the command that we have given via the mic as input to our system, and for the output it segments the sentence into words and then check the words in the data set for its converted speech signal.

2. Methodology

Virtual Personal Assistance [1] works on real time, as its give the required output instantaneously. As we give the command to it via the mic, the speech or command that we have given is first processed and then it is converted to text, then form the text the keys words are extracted and then check with the modules which is stored in the local hard drive, if the keywords match with any of the modules then that particular module will be executed, if the key word doesn't matches with any of the modules than it will just tell the use to try again or it didn't understand what the user wants. As we are using witai as out speech to text converter, we get the option to store the conversation that we had with our Virtual Personal Assistance and we can use it later to make more modules easily.

2.1Hardware Components:

2.1.1Raspberry Pi 3 (Model B)

Raspberry Pi 3 is card size computer, which is shown in Fig 1, with a 64-bit quad core processor of ARM v7 having computing speed of 1.2 GHz, and 1GB RAM. It has 40 pins for GPIO and 4 USB ports which we need for connection of our USB mic and other peripherals. It also has 3.5mm audio jack which we are using to connect our speaker for output.

Figure 1. Raspberry Pi.

2.1.2 MIC

Mic, as shown in Fig: 2, is a input device which we have used here to give input to our Virtual Personal Assistance. Mic are a type of transducer- a device which converts energy from one form to another. Its convert acoustic energy i.e. sound waves into electrical energy into electrical energy i.e. audio signals.

Figure 2. USB Mic.

2.1.3 SPEAKER

Speaker is an output device that we have used for giving us the desired answer from the Virtual Personal Assistance. The speaker is also a kind of transducer which cover the electrical signal i.e. the audio signal, and converts it's to acoustic energy i.e. sound waves.

2.2. Software Component

Lizzy is the name of our AI, or to put in more simple word it is a dead simple API (application programming interface). In Lizzy, we have implemented the STT i.e. the witai and the TTS is the google now for the output voice. In between the STT and TTS, Lizzy performs her task i.e. Lizzy extract the keyword from the text that she receives from STT and check with the local modules that are given to her than give the output if the keywords matches with any of the modules.

3. Simulations and results

The working of system is done according to the steps shown in flowchart Fig 4.

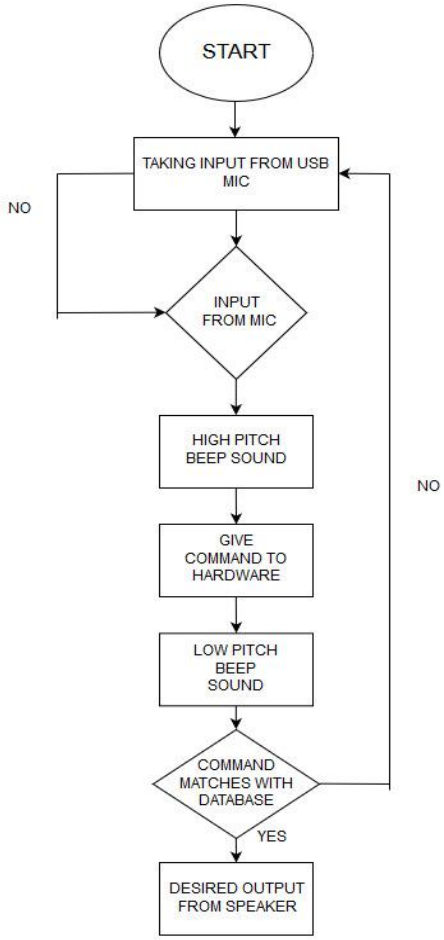

Figure 3. Working Flowchart

We give command to the hardware to get the desired output like today's weather, time, etc. For this purpose, we are using USB mic through which we are giving input in voice format to our system. To interact with system use 'J' as starting voice command. After matching our voice command with

database, the system gets start and gives high pitch sound. Then we give command to system and when input in voice format stops low pitch sound generated. The command which we provide in voice format to the system is get verify with defined database. If command matches then system gives output in the voice format or gives output "pardon" when command is not present in the defined database.

3.1 Speech processing (Speech to Text conversion)

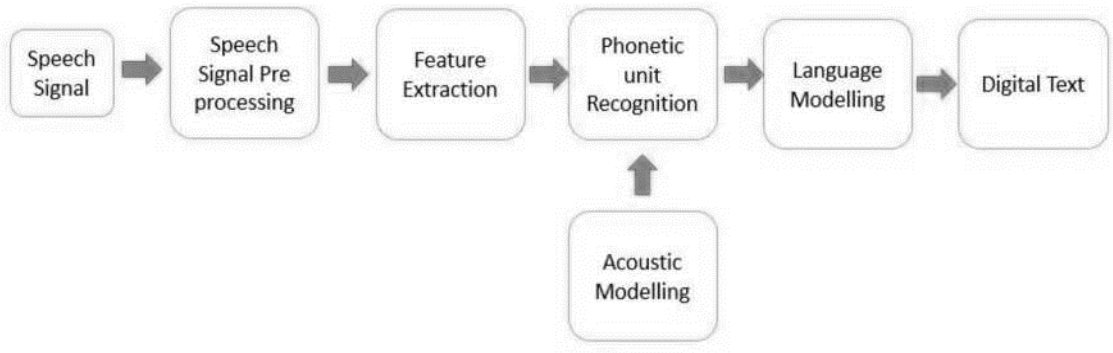

Figure 4. Speech to Text

As we know speech is a time-varying signal [2] so we need to perform windowing method for extraction of signal from speech signal as we can see from Fig: 5. Speech signal consist of voiced and unvoiced signals. We are using hamming window [3] for extraction of phoneme. After the extraction of phoneme, we are determining the which word belongs to that phoneme by extracting the features from that signal and giving it to phonetic unit recognition. There are many words and dialect available in the world, thus for determining the word we need to set the output language model. In our given system, we are setting it to English (UK). Language model also consist of hidden Markov chain model for predicting the next word. After determining the sentence, it will display it as the digital text on screen.

3.2 Text to Speech conversion

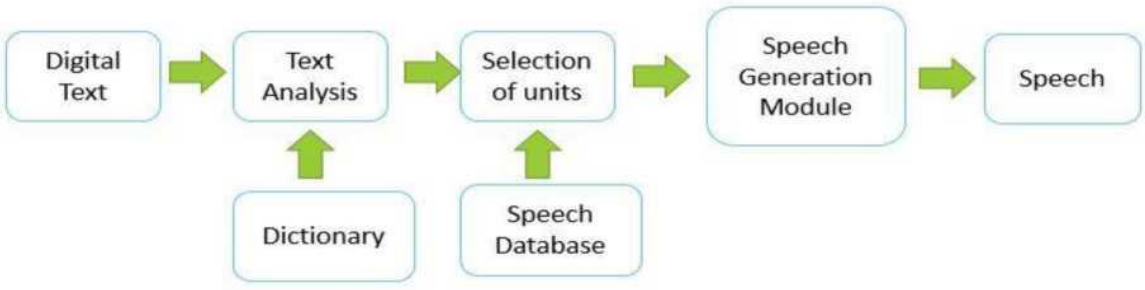

Figure 5. Text to Speech

In speech to text conversion, we are giving the digital text input to the system which will segment each word from given sentence [4] as we can see from the Fig: 6. After the extraction of word, the system

will search for phoneme which represent that extracted word. Then the system will combine all the phonemes and form the speech signal for that given digital text and will play it on speaker.

4. Application

Virtual Personal Assistance can have a lot of application, from simple application of telling time to playing songs or movies to managing our emails. She can also be used for security purpose, if camera module is interface he can do surveillance work. Also, Lizzy have the control over hardware as well i.e. we can just give command to Lizzy and control any hardware we have interface we her.

In future scope, we can implement deep learning [5] on Lizzy so that she can learn own her own and create her own modules. Also, mobile app can be developed for Lizzy, so that we can give her command via our smart phones.

5. Conclusions

Lizzy is fully working Virtual Personal Assistance, which can perform task in offline condition as we given her local modules to her. In online condition Lizzy gets more resources to work with. Also, any peripheral which is connected with the raspberry pi is can be control with the Virtual Personal Assistance, just by giving the command. The local modules can be added or removed by user as he sees fit. Also, there is simple option for conversation with Lizzy, where it learns further.

References

[1] Digital Personal Assistant for the Enterprise2013 *Intel White Paper.*

[2] Comerford, L, Frank, D, Gopalakrishnan P, Gopinath R and Sredivya J. 2001 The IBM Personal Speech Assistant Acoustics, Speech, and Signal Processing Proc. of International Conference on Speech and Signal processing, **1** 1-4.

[3] Lawrence Rabiner, Ronald W. Schafer 2013, Programs for supporting the teaching of digital speech processing *IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)* pp. 290 – 295.

[4] Paul Taylor 2007, Text-To-Speech Synthesis *University of Cambridge* pp 45-56.

[5] Tom M. Mitchell 1997, Machine Learning, *McGraw Hill* pp 88-92