J.A.R.V.I.S

Balraj Saini

Computer Science and Engineering, SRM University Haryana, Sonipat, India

Balraj.saini100@gmail.com

Abstract — **This paper proposes the development of a personal digital assistant and home automation voice command system. Voice assistants are software agents that can interpret human speech and respond via synthesized voices. Apple's Siri, Amazon's Alexa, Microsoft's Cortana, and Google's Assistant are the most popular voice assistants and are embedded in smartphones or dedicated home speakers. Users can ask their assistants questions, control home automation devices and media playback via voice, and manage other basic tasks such as email, to-do lists, and calendars with verbal commands. This column will explore the basic workings and common features of today's voice assistants. It will also discuss some of the privacy and security issues inherent to voice assistants and some potential future uses for these devices. As voice assistants become more widely used, librarians will want to be familiar with their operation and perhaps consider them as a means to deliver library services and materials.**

I. INTRODUCTION

J.A.R.V.I.S. is a personal home automation assistant for controlling electrical home appliances integrated with an augmented reality app. Just A Rather Very Intelligent System (J.A.R.V.I.S.) was originally Tony Stark's natural-language user interface computer system, named after Edwin Jarvis, the butler who worked for Howard Stark A virtual assistant or intelligent personal assistant is a software agent that can perform tasks or services for an individual based on verbal commands. Users can ask their assistants questions, control home automation devices and media playback via voice, and manage other basic tasks such as email, to-do lists, and news report and even get the weather report.

II. EASE OF USE

A. Design and interface

First, digital assistants are easy to use! They are designed that way. Anyone can ask a digital assistant a question and receive a response. Aside from the ease of use, they are always affordable and use little technology. For instance, Jarvis is open source software, the speaker can carry throughout a classroom, and it only requires electricity and Bluetooth.

B. Maintaining the Integrity of the Specifications

Voice assistants could also easily be programmed to act as virtual tour guides in smaller gallery or exhibit spaces. Patrons could ask the assistant to tell them about an exhibit, and the assistant can read back prepared remarks. Libraries with a technology focus may want to consider lending these devices and providing basic training so that patrons can experiment with these devices in their homes as well.

III. PURPOSE OF THE RESEARCH

Advances and evolution of the intelligent personal assistants strictly stick to the recent developments in technology. Therefore, it is a research field which is subject to constant changes. In this regard, the purpose of this research is to reveal patterns and identify key concepts in research on intelligent personal assistants. Research is to reveal patterns and identify key concepts in research on intelligent personal assistants. The speech synthesizer will be very useful to any researcher who may wish to venture into the "Impact of using Computer speech program for brain enhancement and assimilation process in human beings". This text-to-speech synthesizing system will enable the semi-illiterates assess and read through electronic documents, thus bridging the digital divide. The technology will also find applicability in systems as banking, telecommunications (Automatic system voice output), transport, Internet portals, accessing PC, emailing, administrative and public services and many others. The system will be very useful to computer manufacturers and software developers as they will have a speech synthesis engine in their applications.

IV. TEXT-TO-SPEECH SYNTHESIS DEFINED

A speech synthesis system is by definition a system, which produces synthetic speech. It is implicitly clear, that this involves some sort of input. What is not clear is the type of this input. If the input is plain text, which does not contain additional phonetic and/or phonological information the system may be called a text-to-speech (TTS) system. A schematic of the text-to-speech process is shown in the figure 1 below. As shown, the synthesis starts from text input. Nowadays this may be plain text or marked-up text.
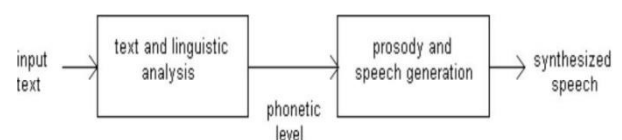

Fig 1: Schematic TTS

V. IDENTIFICATION OF SPEECH

Speech is an acoustic signal produced from a speech production system. In speech production, nature of speech is identified by excitation phenomenon. Based on the input excitation it classifies in to three levels. One if, the input excitation is nearly periodic known as voiced speech and other will be in the random noise like nature (unvoiced speech) otherwise in no excitation state (silence region).

As on the fig 2 If the input excitation is nearly periodic impulse sequence the corresponding speech looks like visually nearly periodic and is called as voiced speech.
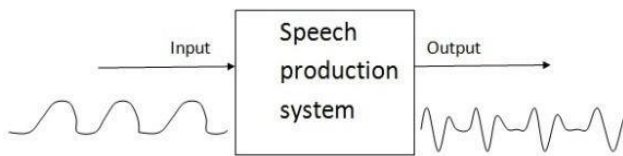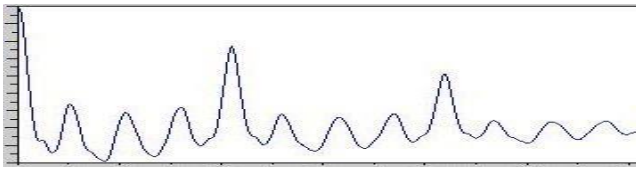
Fig 2: Speech system

Fig 3: Autocorrelation sequence

The periodicity can be measured by autocorrelation analysis and the period is termed as pitch period. **_Pitch period_** is defined as the largest peak n the autocorrelation sequence from the beginning. Fig 3 shows the sample autocorrelation sequence of a speech frame. Voice speech is periodic in nature and there should be some frequency and harmonics in the spectrum of speech. From the spectrum as shown in fig: it is clear that the frequency is repeating after a particular interval. This will be the harmonic structure. The signal having no periodicity will be called as unvoiced speech. In unvoiced speech no harmonic structure is present. The region identified by visual perception and automatic approach. In an intelligent speech the duration of the silence region is an important concern. But from the energy point of view it is less concerned. The fig shows a sample speech signal of word 'HelloHello'. The space in between the speech signal is the silent region and the set of varying signals be the voice region.
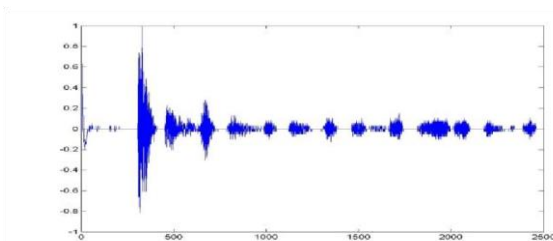

Fig 4: Spectrum

Fig 5: Speech signal waveform for 'HelloHello'

VI. THE HOUSE MODEL

By using voice commands, a user shall be able to switch on and off home devices such as lighting, fan, multimedia,

etc. A key feature of this system is its flexibility to adapt to different scenarios, and devices, as the user is able to add, delete, and edit rooms, user accounts, and devices. For security purposes, the systems require login to access it, and there are two kinds of users, the administrator user and the common user. Each one has a different scope of access in the system, for instance, the administrator has permission to add, delete and edit rooms, user account, and devices, whereas the common user is limited to only give voice commands. From a technical point of view, the system utilizes a local network in which the UI and the Arduino have different roles. The UI is JARVIS, an windows application client in the Bluetooth network, and its main role is to provide the user the means to interact with the system. It also processes the user's voice commands through a third party artificial neural network system from Google and the outcome of that is sent to the Arduino via socket in the local Bluetooth network. At last, it has a database to store data from the server, so the server does not need to be consulted every time, and it has a simple cryptography system to protect the data integrity in the system. The Arduino is the server in the Bluetooth network, and its main role is to process and manage all the client requests sent to it, execute the requests and send a reply back to the client.

In order to evaluate the system behavior in a real case scenario, it was developed a miniaturized residence with three rooms. This house model is filled with three devices symbolically represented by led's. The Table 1 details the rooms with the devices within.

TABLE 1

ROOM	DEVICE
KITCHEN	LIGHTS
BED ROOM	LIGHTS
DINNING ROOM	LIGHTS

The proposed system reached the expectations, by demonstrating a relevant flexibility in adapting to different scenarios and properly responding to the user voice requests. However, when it comes to processing time, the Microsoft Speech Recognition Engine seemed to take a while to deal with the voice requests. In addition, there were a few instances that the user speech could not be recognized.

The methodology adopted to simulate the system involved a user speaking a sequence of commands, and the result of that is photographed. Moreover, the appliances can be customized with the help of Jarvis and user interaction. For instance, if a user wants to turn on/off a fan, he can add the custom command of turning on/off the fan in Jarvis.

VII. METHODOLOGIES

Voice recognition works based on the premise that a person voice exhibits characteristics are unique to different speaker. The signal during training and testing session can be greatly different due to many factors such as people voice change with time, health condition (e.g. the speaker has a cold), speaking rate and also acoustical noise and variation recording environment via microphone.

TABLE 2

Process	Description
1) Speech	2 Female (age=20, age=53) 2 Male (age=22, age=45)
2) Tool	Mono Microphone Microsoft Speech software
3) Environment	College Campus
4) Utterance	Twice each of the following word 1) Volume Up 2) Volume Down 3) "wakeup jarvis" 4) Introduce yourself 5) Show date.
5) Sampling Frequency	16000 KHz

CONCLUSION AND FUTURE SCOPE

It can be presumed that HOME AUTOMATION SYSTEM USING ARDUINO was a victory. This framework comprises of an Arduino-Uno board, A Bluetooth Module, home apparatuses and Jarvis. It is easy to understand and it is financially savvy. Likewise, it can be inferred that the destinations of this venture have been effectively met and they are asper the following: Constructed a remote home computerization framework controlled by Jarvis. Designed an easy to use and a sheltered framework to control home apparatuses particularly pointed to help the older folks and debilitated. This app has large scope as it has following features:

☐ Accuracy will become better and better.

☐ Object and face recognition will be implemented.

☐ More search engines support will be there.

☐ Microphones and sound systems will be designed to adapt more quickly to changing background noise levels.

☐ More platforms will be supported in next version.

REFERENCES

[1] "Robomart.com" [Online]: Available all electronic items used in the project for smart home automation.
[2] Dr Arvind Kumar, Project Mentor and guide.
[3] Salim Sr. "Youtuber and Programmer, advanced voice recognition programming skills"
[4] [Online]. Available: http://arduino.cc/en/Guide/Introduction.
[5] Bluetooth," in Inigo Puy, 2008.
[6] [Online]. Available: http://www.bluetooth.com/Pages/Fast-Facts.aspx.