

Voice based Email System for Blinds

Prateek Grover

Dr. A.P.J. Abdul Kalam Technical
University
Meerut, India

Naman Chikara

Dr. A.P.J. Abdul Kalam Technical
University,
Meerut, India

Priyam Khara

Dr. A.P.J. Abdul Kalam Technical
University
Meerut, India

Pallav Mulchandani

Dr. A.P.J. Abdul Kalam Technical University,
Meerut, India

Piyush Bhatti

Dr. A.P.J. Abdul Kalam Technical University
Meerut, India

ABSTRACT

'Helen keller' once said that "Someone having sight without any vision is worse than being blind". In the last two decades the world have seen unprecedented amounts of growth in the world of computers, but for the visually impaired people the world still remains dark and stunted. Ever since the early days of computers till now very few initiatives have been taken to escalate the no of blinds using these technologies. To amend the current scenario the well-versed technique of Audio feedback or IVR (previously applied in virtual environments like text readers) is being used excessively in the project helping the visually impaired people to be a part of the current revolution. The project aims at helping these people to feel connected by using email system through speech recognition in English language. The use of keyboard is eradicated. The system can be used by anyone including amateur users who are unable to read, write but only understand the English language.

Keywords

Speech-To-Text Converter, Text-To-Speech Converter, Interactive Voice Response (IVR), Speech recognition.

1. INTRODUCTION

Electronic mail is a popular method of sending messages ("mail") between two or more people using any electronic devices. It was invented by Ray Tomlinson, email was first used in the 1960s and by the mid-1970s it took the world by storm and came to be known as the term now known as email. Email particularly relies on the Internet as its interface to securely connect people and fulfil their daily goals.

But this conventional email system is not useful in all types of people particularly the older people who find the interface of the email too complicated to work with (perfectly explained in [6] and [7]) blind people and as they can't send the email via their visual senses. So, an Audio based email is a preferable option blind peoples as they can easily respond to the audio instructions in their native tone[10].

The Voice based system uses TTS (Text-to-Speech) (further explained in [11]) in order to provide the user with the ability

to control the system via his/her voice. It serves the purpose of an independent program, and can be easily installed on any smartphone without any spending. It accords the blind people to make their way up to the program. Extensive study dedicated to this subject study (Security aware email server. [12]) have been used in the technology to help blind people to connect to the global world.

The remaining paper structure accommodates 5 sections. Section II portrays the different technologies instilled within

the project and their initial uses in different areas. Section III compares the existing technologies of STT conversion then Section IV intends to delineate the working of the whole system whereas Section VI describes the conclusion and future work.

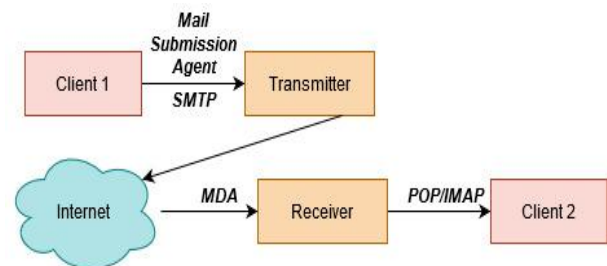


Figure 1. Complete process of Email System

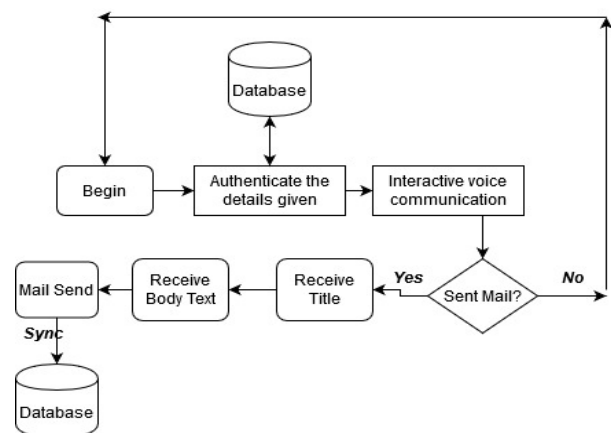


Figure 2. Prototype of Voice Based Email System

2. LITERATURE SURVEY

'PJM Mengibar' [1] discussed about the dialectic analysis of how a Voice recognition System intends to convert the voice input provided by the user into a text output. This output is generally used to process various dimensions of the decision tree. The input is broken down into segments where each segment is then explicitly studied to further determine the transcription for the various segments of the voice based input along with the stability criteria helping in drastically improving the accuracy of the voice recognition system. The Author further elaborated how each segment could have multiple matching criteria and how the stability factor plays a big role in helping to reach the right output.

'Dennis McLeod' [2] in their paper extensively discussed in their paper about determining the Emails for whether they are used properly or are they being misused. After using majorly 4 algorithms((SVM classifier, Neural Network, J48 classifier and Naive Bayesian Classifier) to classify the email on whether it's a spam or it is normal inbox mail the author found that simply using the J48 classifier along with the binary tree could help the dataset to be classified accurately.

'Pranjal Ingle' [3] comprehensively explained how the system utilized the speech to text technology to convert the words spoken by the user into text. Further the system utilized the Text to speech technology which is the opposite of Speech to text converting the text to synthesized speech. Author further illustrated another component of the system called IVR (Interactive voice response) describing it as an advanced technology acting as an interface between the user and the system .The users can easily make their own problems known by listening to the IVR dialogue .IVR makes use of the system keyboard to configure the tasks to be performed. Author presented the user with an interactive UI option for the blinds to easily use the Email facilities the system has imbibed.

'Naveen S' [4] elucidated how the existing GUI of the Email system combined with the new technological advancements made in the word of Speech Recognition can pave the way for the blinds to be the part of this unprecedented rise in the world of computers. They further described how Text-To-Speech (TTS) mechanism is the way forward as it can be easily instilled in any android device and can help the blinds type characters via their voice. Author explained some methods of decrypting the voice in an efficient way using Java API's and using a security module such as an App lock to protect any valuable information, for example , The password of the user.

'K Jayachandra' [5] explained in their paper about how the blinds can access e-mails with ease using it to send and receive messages using Speech Recognition in their native language by using a Computer or any other mobile device. With minimal requirements of a 1.44mb Floppy Drive and a nominal ram of 512mb showed the relative study of their GUI promised to achieve what is intended.

3. COMPARISON ON EXISTING TECHNOLOGIES

As the world of speech to text conversion go, different technologies are usually preferred in different circumstances. And the best way to compare them is through the universally accepted common metric system of Word Error Rate(WER).

For this purpose we conducted a simple experiment of testing three Speech-To-Text API's for different audio files varying in the category of 4000Hz to 16000 Hz.

The WER was recorded during the experiment for the sole reason of deciding which API would help us to increase the accuracy for the system to be used.

Word error rate was calculated by:

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

Figure 3. Formula to calculate Word Error Rate

where,

S represents the no of substitutions,

D represents the deletions,

I represents the insertions,

C represents the correct words,

N represents the number of words in the reference (N=S+D+C).

After comparing the error rate over the different audio recordings it was concluded that the Google API had the mean word error rate of 9% WER which was followed by Microsoft API standing at a mean word error rate of 12% WER and IBM Watson API Stood at 16% overall.

Based on the conclusion of the research the Google API was chosen as the basis of the project for the corresponding task of Speech-To-Text Conversion.

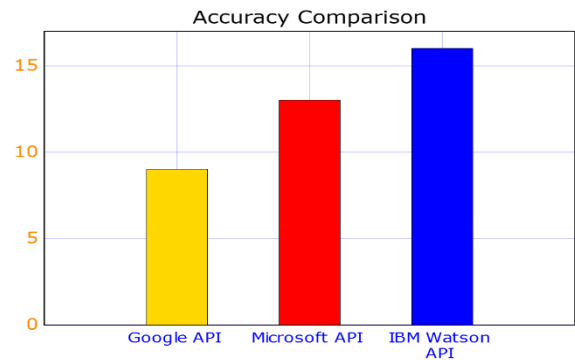


Figure 4. Comparison Between Three Systems

4. METHODOLOGY

The first and the foremost step is to take the input from the user by using a microphone at run time and thereby maintaining the stability factor of the input speech which is then is processed to recognize the spoken text.

The transcription is created for each segment of the input in order to increase the accuracy. So when the Speech to text converter is used it directly converts the input speech into text which can further be manipulated to produce the required results.

The system receptiveness for blind or physically handicapped can be enhanced by speech detection system.

The whole system is based upon the voice prompts within the terminal.

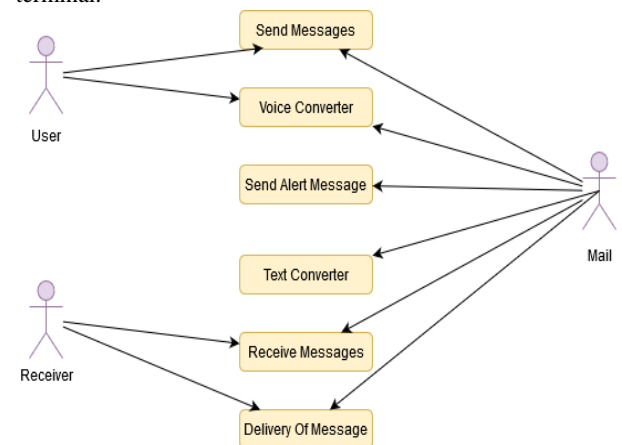


Figure 5. Use Case Diagram

While making use of this system the computer will ask the user to operate several operations in order to utilize the services being provided and if the user desires to perform a particular task he/she needs to perform a specific operation. So when the user speaks the input is taken and it is converted into text using the Speech-To-text Converter. Once the system is open it is totally governed by voice commands. After the input is recorded all operations are laid out from the start and there is a specific operation for every decision.

So the Text to speech converter is used to ask for the decision the user wants to take. Also because of the facilities provided, those who cannot read need not have to worry and the system will take care of all aspects by itself.

In between the decisions IVR (a sophisticated mechanism that elucidates the interaction between the blind and the system through the use of using voice and DTMF tones) is used. IVR helps customers to connect with a company's host system via speech recognition, after which the user can perform operation using the IVR dialogue.

IVR responds with pre-recorded or dynamically generated audio and which is further combined with Python Avbin64 to decode the voice input and to further helping the user to achieve their goal.

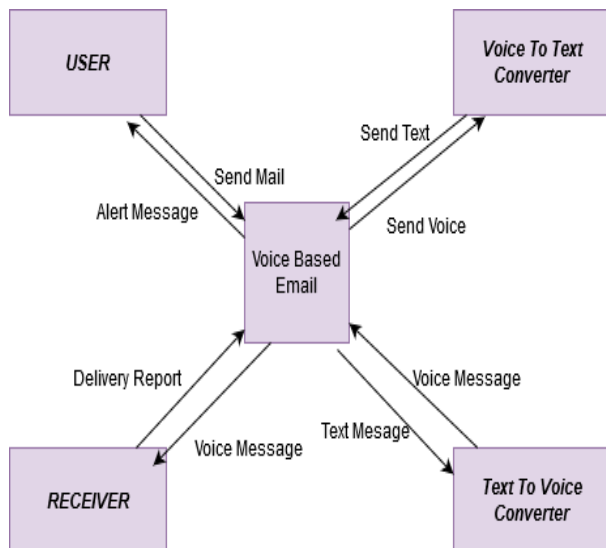


Figure 6. Data flow diagram

5. ACKNOWLEDGMENTS

We wish to acknowledge the help of provided by technical support staff in the Computer Science department of Dr.A.P.J. Abdul Kalam Technical University. We would also like to show our deep appreciation to our supervisor Mr. Vishal Choudhary who helped to finalize my project.

6. CONCLUSION AND FUTURE WORK

The main reason for creating this sort of system described in the paper is to make the visually impaired people feel a bit more connected in this small world. This mailing system will assist in overcoming the slight problems that are faced by visually impaired people by using their voice as the input. This system hence doesn't have to use screen readers and automatic speech recognizer because the system consumes Two-way conversations helping the user to by eradicating the use of the Keyboard. Since each operation has a specific result the decision tree follows a unique path through IVR thus making the system much more compatible. For now the application is only available for desktops. Also the interface of the system can include more features in the future

7. REFERENCES

- [1] 'A.P Aleksic, PJM Mengibar '- US Patent 10,049,666, 2018 - Voice Recognition System
- [2] 'Seongwook Youn' and 'Dennis McLeod'- A Comparative Study for Email Classification
- [3] 'Pranjal Ingle', 'Harshada Kanade,' , 'Arti Lanke' 2016 Voice based e-mail System for Blinds
- [4] 'Naveen S', 'Sreenidhi' 'K,Sreekanth P H', 'Ravi Kumar N' 2018 Mailing System Based Voice.
- [5] ' K Jayachandran', ' P Anbumani '- Int. J. Adv. Res. Ideas Innov. Technol, 2017 - Voice based email for blind people.
- [6] 'Robin N. Brewer', 'Raymundo Cornejo', 'Ted Schwaba, Darren Gergle,' and 'Anne Marie Piper'. 2016. Exploring Traditional Phones as an E-Mail Interface for Older Adults.
- [7] 'Anjali Kumari', 'Tanvi Nagrale, Tripti Pragyana Mohapatra, Anand Prem Rajan' 2019. Old age people problems and their importance in everyone's life.
- [8] ' Sourav Mukherjee' 2017 Popular SQL Server Database Encryption Choices
- [9] 'SM Simmons', 'JK Caird', ' P Steel '- Accident Analysis & Prevention, 2017 - A meta-analysis of in-vehicle and nomadic voice-recognition system interaction and driving performance
- [10] ' CS Jochumson' - US Patent 9,947,321, 2018 - Real-time interactive voice recognition and response over the internet.
- [11] 'Y Wang', ' RJ Skerry-Ryan'... 2017 Tacotron: A fully end-to-end text-to-speech synthesis Model.
- [12] 'GE Buddepalli', ' A Govindjee', 'JS Li', ' N Pattipati' 2017 Security aware email server.