Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT-2020)
IEEE Xplore Part Number:CFP20F70-ART; ISBN:978-1-7281-4685-0

A Review on Voice based E-Mail System for Blind

[1]Paulus A. Tiwari, [1]Pratiksha Zodawan, [1]Harsha P. Nimkar, [1]Trishna Rotke, [1]Priya G. Wanjari, [2]Umesh Samarth
[1]BE Students, Department of Information Technology, J. D. College of Engineering and Management, Nagpur, Maharashtra, India
[2]Assistant Professor, Department of Information Technology, J. D. College of Engineering and Management, Nagpur, Maharashtra, India.

Abstract— **Due to its simplicity and accessibility, Internet is widely used in almost all the communication applications. In the recent times, number of application based on internet have been developed to make the communication as a more reliable and efficient in nature. Out of this numerous applications, E-mail is the most widely used and reliable way to communicate with each other. The usage of e-mail is quiet easy and lucid for regular users but when it comes to the user with visual defect, the system is yet very difficult to use. The current system is not useful for people with visual defect as the available system are based on the visual perceptions. There are huge up gradation in the technologies now a days, especially for the visually challenged people. Still the current emailing system is yet not upgraded for the use of visually impaired. This arises a significant need to upgrade the existing system to make it more useful for the visually impaired. Thus, in this study we present an email system working on the voice controlling principle for the people with visual impairment to deliver a simple and easy access to the email system. This framework will also helpful for the individuals with other weaknesses alongside the visually impaired individuals.**

***Index Terms*— Speech Recognizer, Text to Speech Converter, Visually Impaired People, Speech to Text Converter, Screen Reader.**

I. INTRODUCTION

With the boom in internet technologies, the communication has become a lot easier. Internet is considered as the vault of innovation, technologies and information. Numerous networking and social media sites. The most conventional way of online communication is e-mail. It is estimated that there are more than 4.5 Billion email accounts. By the end of 2020, this figure in estimated to rise up to 5.9 Billion, which is a improvement of over 29.5 %. There are 2.586 B email clients overall along with both business and purchaser clients as per [5].

Along these lines, email remains as the accepted standard for delivering noteworthy communication. For utilizing these offices of Internet each individual require visual ability. Since on visual discernment to comprehend what substance are available onscreen. Henceforth this type of frameworks are of no utilization for visually impaired people.

For making this frameworks helpful for these visually challenged individuals. There are different advances given to them such as Automatic speech recognizer, screen reader, text to speech to speech to text, braille console and so on. Nonetheless, these advancements are not so much valuable for those individuals as it couldn't give the best possible response like an ordinary framework. To use above frameworks visually challenged individuals' faces numerous issues.

II. INTERACTIVE VOICE RESPONSE(IVR)

Interactive voice response (IVR) is an innovation that enables a PC to associate with people using voice and DTMF tones input through a keypad. In broadcast communications, IVR enables clients to connect with an organization's host framework by means of a phone keypad or by speech recognition, after which administrations can be asked about through the IVR exchange. IVR frameworks can react with pre-recorded or dynamically produced sound to additionally guide clients on the best way to continue. IVR frameworks sent in the network are measured to deal with large call volumes and furthermore utilized for outbound calling, as IVR frameworks are cannier than numerous prescient dialer frameworks.

IVR frameworks can be utilized for portable buys, banking installments and administrations, retail orders, utilities, travel data and climate conditions. A typical misguided judgment alludes to a robotized attendant as an IVR. The terms are particular and mean various things to conventional broadcast communications experts. The reason for an IVR is to take input, process it, and return an outcome, while that of a mechanized specialist is to course calls. The term voice response unit (VRU) is sometimes utilized too. DTMF decoding and discourse acknowledgment are utilized to decipher the guest's reaction to voice prompts. DTMF tones are entered by means of the phone keypad. Different advances incorporate utilizing text-to-speech (TTS) to speak complex boggling and dynamic data, for example, messages, news reports or climate data. IVR innovation is additionally being brought into vehicle frameworks for sans hands activity. TTS is PC created synthesized discourse that is never again the robotic voice generally connected with PCs. Genuine voices make the speech in pieces that are joined together (linked) and smoothed before being played to the guest.

Another technology which can be used is using text to speech to talk advanced and dynamic data, such as e-mails, reports and news and data about weather. IVR used in automobile systems for easy operations too. Text to Speech is system originated synthesized speech that’s not the robotic

voice historically related to computer. Original voices produce the speech in portions that are joined together and rounded before played to the caller.

III. Speech Recognition

Speech recognition (SR) is the ordered sub-field of computational linguistics (CL) that generate techniques and advancements to empower the acknowledgment and interpretation of communicated in language into text by PCs. It is also known as "automatic speech recognition "(ASR), "computer speech recognition ", or only "speech to text" (STT). It consolidates information and study in the linguistics, software engineering, electrical engineering fields. Some SR frameworks require "training" where an individual speaker understands message or isolated vocabulary into the framework. The framework breaks down the individual's particular voice and utilize it to recognition the acknowledgment of that individual's speech, for bringing about expanded exactness. Frameworks that do not use preparing are called as "speaker independent "frameworks. Frameworks that utilization preparing are named as "speaker dependent ". SR applications incorporate voice UIs, for example, voice dialing, call routing, household apparatus control, search (for example discover a web recording where specific words were verbally expressed), basic information entry (e.g. Visa number), readiness of organized reports for instance a radiology report, discourse to-content handling for instance word processors or messages, and airplane (ordinarily named Direct Voice Input). The term voice recognition or speaker recognizable proof alludes to identifying the speaker, instead of what they are saying. Perceiving the speaker can enhance the undertaking of deciphering discourse in frameworks that have been generated on a particular individual's voice or it tends to be used to confirm or check the personality of a speaker as a major aspect of a security procedure. From the innovation point of view, SR has a long history with a several waves of significant advancements. Most as of late, this field has profited by progresses in big data and deep learning (DL). The advances are proving not just by the flood of academic papers distributed in the field, however more critically by the overall worldwide industry appropriation of an assortment of DL strategies in planning and conveying SR frameworks.

SR works utilizing calculations or algorithms through and language modeling and acoustic. Acoustic modulation speaks to the connection among linguistic part of speech and audio signals; language modeling matches sounds and word successions to help recognize words that sound comparable. Frequently, the algorithm called hidden Markov models are utilized also to perceive transient examples in speech to enhance exactness inside the framework. The most continuous uses of SR inside the endeavor incorporate call routing, speech to text preparing, voice dialing and voice search. While advantageous, speech recognition innovation still has a couple of issues to work through, as it is consistently created. The good things of speech recognition softwares are it is easy to utilized and promptly accessible. Speech recognition software is currently much of the time introduced in PCs and cell phones, considering simple access. The drawback of speech recognition incorporates its failure to catch words because of varieties of elocution, its absence of help for most languages outside of English and its powerlessness to figure out foundation commotion. These variables can prompt errors.

SR execution is estimated by precision and speed. Accuracy estimated with word error rate (WER). WER works at the word level and distinguishes mistakes in interpretation, despite the fact that it can't recognize how the error occurred. Speed estimated with the real time factor. An assortment of elements can influence PC SR execution, accent, consist of pronunciation, pitch, background noise and volume. It is essential to take note of the terms speech recognition and voice acknowledgment are some of the time utilized conversely. Be that as it may, the two terms mean various things. SR is utilized to distinguish words in communicated in language. Voice recognition is a biometric innovation utilized for recognizing a specific person's voice or for speaker identification.

IV. Literature Review

Paper [1] explains the "Voice Based System in Desktop and Mobile Devices for Blind People". Voice message engineering encourages blind individuals to get to email and other interactive media elements of working framework. Additionally, in mobile application SMS can be perused by framework itself. Presently the headway made in PC innovation opened stages for visually defective individuals over the world. It has been seen that almost about 60% of absolute visually impaired populace over the world is available in INDIA. Here authors depict the voice message design utilized by blind individuals to get to E-mail and media elements of working framework effectively and productively. This design will likewise decrease intellectual burden taken by oblivious in regards to recall and type characters utilizing console.

There is main part of data accessible on innovative advances for visually challenged individuals. This incorporates advancement of content to Braille frameworks, readers and magnifiers of screen. As of late, endeavors have been made so as to create devices and advances to assist Blind with people to get to web advances. Between the early endeavors, voice information and contribution for surfing was received for blind peoples. In the home page of IBM the website page is a simple to-utilize interface also changes over the content to-discourse having diverse sexual orientation voices for understanding texts, links. In any case, the burden of this is the engineer needs to plan a composite new interface for the complex graphical website pages browsed and for the screen reader to perceive.

Basic perusing arrangement, which isolates a web page in two measurements. This enormously rearranges a web page structure and makes it simple to browse. Another internet

browser created a tree structure from the HTML record through links analyzation. As it endeavored to structure the pages that are connected together for improving Travers ability, it didn't demonstrate exceptionally proficient for surfing. After, it didn't deal with needs in regards to traversability and ease of use current page itself. Another browser produced for the visually challenged individuals was e-Guide which had an incorporated TTS engine.

Some exceptional data extraction mechanism or algorithm has been applied for representing the web page into an easy to understand way. Be that as it may, at present it didn't fulfill the necessary guidelines of business use. Thinking about Indian situation, Shruti Drishti and the "web browser for the blind individual" are two web surfing application that are utilized by blind individuals to get to the web including the messages or email. Both frameworks are coordinated with Indian language ASR and TTS frameworks. Be that as it may, the accessible frameworks are not portable for little gadgets like cell phones.

In paper [2] authors build up a web index which supports Man-Machine collaboration absolutely as voice. Web-page Reader and a novel Voice based Search Engine presented that enables the clients to order and control the internet browser through their voice. The current Search Engines get demand from the client as content and react by recovering the pertinent records from the server and shows as content. Despite the fact that the current internet browsers are fit for playing sounds and recordings, the client needs to demand by composing some content in the hunt content box afterward the client can play the interested videos with the assistance of Graphical User Interfaces (GUI). The Voice based Search Engine tries to serve the clients particularly the visually impaired in perusing the Internet. The client can talk with the PC and the PC will react to the client as voice. The PC will help the client in perusing the reports too.

Voice enabled interface [3] with expansion support for signal based information and yield approaches are for the "Social Robot Maggie" changing over it into aloud reader. This voice recognition and synthesis can be influenced by n number of reasons, for example, the voice speed, its pitch, its volume and so forth. It depends on the Loquendo (Emotional Text-To-Speech) ETTS software. Robot additionally communicates its state of mind through signal that depends on geostationary. Speech recognition precision can be improved by expulsion of noise. In A Bayesian plan is applied in a wavelet space to isolate the speech and noise parts in a proposed iterative speech upgrade algorithm. This proposed technique is created in the wavelet area to misuse the chose highlights in the time frequency space portrayal. It includes two phases: a noise gauge arrange and a sign partition organize.

In the Principle Component Analysis (PCA) based HMM for the visual philosophy of broad media annals is used. PCA (Principle Component Analysis) [4] and PDF (Probabilistic Density Analysis).

Presents an approach to manage discourse acknowledgment using fluffy displaying and fundamental initiative that dismisses commotion instead of its identification and expulsion. In the discourse spectrogram is changed over into a fluffy semantic delineation and this depiction is used instead of accurate acoustic features. In Voice affirmation technique got together with facial component collaboration to assist virtual skilled worker with upper limb debilitations to make visual cut in a mechanized medium, spare the peculiarity and authenticity of the artistic work. Strategies to recover wonders, for instance, Sentence Boundaries, Filler words and Disfluencies suggested as fundamental Metadata are discussed in and depict the technique that normally incorporates information about the region of sentence points of confinement and discourse Disfluencies to propel discourse acknowledgment yield.

Clarissa a voice empowered browser that is conveyed on the international space station (ISS). The principle segments of the Clarissa framework are speech acknowledgment module a classifier for executing the open microphone acknowledges/dismiss choice, a semantic investigation and an exchange director. Essentially centers on expressions.

For assembling prosody model for every expressive state, an end pitch and a delta pitch for every syllable are anticipated from a lot of highlights accumulated from the text content. The articulation labeled units are then pooled through the neutral data, in a TTS framework, such paralinguistic occasions proficiently give signs with regards to the condition of an exchange, and Markup determining these occasions is an advantageous path for a designer to accomplish these kinds of occasions in the sound originating from the TTS engine.

Principle highlights of are smooth and regular sounding speech can be integrated, the voice qualities can be transformed, it is "trainable. Restrictions of the fundamental framework is that synthesized speech is "buzz" since it depends on a coding system, it has been overwhelmed by top notch vocoder and shrouded semi-Markov model based acoustic demonstrating. Speech synthesis comprises of three classifications: Articulation Synthesis, Concatenation Synthesis, and Formant Synthesis. Principally centers around formant synthesis, exhibit of phoneme of syllable with formants frequency is given as information, recurrence of given information is handled, on teamed up with Thai-Tonal-Accent Rules convert given formants recurrence organization to wave design, so sound yield through soundcard.

Here, the review of methods need for proposed system. For this model need speech to text, text to speech, Mel Frequency Cepstral Coefficients (MFCCs).

The content to speech framework [5] is utilized to make an interpretation of textual content into voice based stream which is partitioned into two sections such as head and tail end. The head end has basically two issues to address. Firstly which

converts over the content consist of images like the numbers written in textual form? This technique is known as the normalization of text. The head end at that point allots phonetic recreation to every single word. The way toward allocating transcripts to words is known as content to phoneme translations. The tail end usually implied to as the synthesizer at that point changes over the symbolic linguistic representation into sound.

Automatic Speech Recognition (ASR) or Computer Speech Recognition (CSR) is the discourse to content acknowledgment [6]. It is the process where the computer understands the voice and playing out any necessary undertaking or the capacity to coordinate a voice against a given input or acquired vocabulary. The undertaking is getting a PC for understanding spoken language. By “understand" we intend to respond properly and by converting the data voice into another understandable means of communication for example content. The Speech acknowledgment is along these lines at times alluded to as voice synthesis or the process of speech to text (STT). Such a speech acknowledgment framework comprises of a mic, where the user can speak; the mechanism which will analysis and identify the speech; a PC to take and decipher the speech; a great quality soundcard for input and additionally yield; a legitimate and great elocution.

The initial phase in any automatic speech recognition framework is to separate highlights for example recognize the parts of the audio sign that are useful for distinguishing the phonetic substance and disposing of the various hardened which conveys data like emotion, background noise and so forth. The central matter to comprehend about speech it is subsequently the sounds that is produced by a human are separated by the state of the vocal tract consist of teeth, tongue and so on. This shape figures out what sound turns out. In the event that we can decide the shape precisely, this should give us an exact portrayal of the phoneme being created. The state of vocal tract shows itself in the wrap. This page will give a short instructional exercise on MFCCs Mel Frequency Cepstral Coefficients (MFCCs) is an element generally utilized in programmed speech and speaker recognition.

V. CONCLUSION AND FUTURE WORK

The analyzed research works help people with visual impairment to access the email in problem freeway which is the most all-inclusive type of contact in today’s world. The proposed framework helps to lessen the hurdle, for example, memorizing and utilizing the mouse clicks and keyboard shortcuts that were used by the person with visually debilitated while receiving the email. Thus we proposed a voice based authentication instead of the traditional username and password.

This framework will coordinate the necessary activity and the aftereffect of the activity. Considering all these executing strategies this framework gets easy to use, secure as well as interactive. The framework growing presently is depend just on personal machine. With the use of technically advanced smartphones, such systems and applications has a chance to be implemented as an App in smartphones. Thus there is a scope to implement the framework in various other languages rather than implementing it only in English language.

REFERENCES

[1] Jagtap Nilesh, Pawan Alai, Chavhan Swapnil and Bendre M.R.. “Voice Based System in Desktop and Mobile Devices for Blind People”. In International Journal of Emerging Technology and Advanced Engineering (IJETAE), 2014 on Pages 404-407 (Volume 4, issue 2).

[2] Ummuhany sifa U., Nizar Banu P K , “Voice Based Search Engine and Web page Reader”. In Internationa Journal of Computational Engineering Research (IJCER). Pages 1-5.

[3] Preeti Saini, Parneet Kaur “Automatic Speech Reorganization: A Review”, International journal of Engineering Trends and Technology-Volume 4 Issue 2-2013.

[4] JishaGopinath, Aravind S, PoojaChandran, Saranya SS “Text-to-Speech Conversion System using OCR”, International Journal of Emerging Technology and Advanced Engineering website: www.ijetae.com(ISSN 2250-2459, ISO 9001:2008 certified journal, Volume 5, Issue 1, January 2015)

[5] Mel Frequency Cepstral Coefficient(MFCC) tutorial

[6] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[7] Speaker/Voice-Command Recognition in MATLAB http://matlab-recognition-code.com.