Voice Based Email System
for the
Visually Impaired

Submitted March 2021 in fulfillment of
HCI Mini Project
Under the guidance of Dr. V K Sambhe

Anirudh Khatry (171080012)
Kritika Ravishankar (171081063)

Department of Computer Engineering Information Technology
Veermata Jijabai Technological Institute, Mumbai - 400019
(Autonomous Institute Affiliated to University of Mumbai)

Table of contents

Introduction

We have seen that the introduction of the Internet has revolutionized many fields. The Internet has made life of people so easy that people today have access to any information they want easily. Communication is one of the main fields highly changed by the Internet.

E-mails are the most dependable way of communication over the Internet, for sending and receiving some important information. But there is a certain norm for humans to access the Internet and the norm is you must be able to see.

But there are also differently abled people in our society who are not gifted with what you have. There are some visually impaired people or blind people who can’t see things and thus can’t see the computer screen or keyboard.

A survey has shown that there are more than 240 million visually impaired people around the globe. That is, around 240 million people are unaware of how to use the Internet or E-mail. The only way by which a visually challenged person can send an Email is, they have to speak the entire content of the mail to another person(not visually challenged) and then that third person will compose the mail and send on the behalf of the visually challenged person. But this is not a right way to deal with the problem. It is very unlikely that every time a visually impaired person can find someone for help.

Existing systems

The most common mail services that we use in our day to day life cannot be used by visually challenged people. This is because they do not provide any facility so that the person in front can hear out the content of the screen. As they cannot visualize what is already present on screen they cannot make out where to click in order to perform the required operations. For a visually challenged person using a computer for the first time is not that convenient as it is for a normal user even though it is user friendly.

However, there is a bulk of information available on technological advances for visually impaired people. This includes development of text to Braille systems, screen magnifiers and screen readers. Although there are many screen readers available then also these people face some minor difficulties. Screen readers read out whatever content is there on the screen and to perform those actions the person will have to use keyboard shortcuts as mouse location cannot be traced by the screen readers. This means two things; one that the user cannot make use of the mouse pointer as it is completely inconvenient if the pointer location cannot be traced and second that the user should be well versed with the keyboard as to where each and every key is located. A user who is new to computers can therefore not use this service as they are not aware of the key locations.

Another drawback that sets in is that screen readers read out the content in sequential manner and therefore users can make out the contents of the screen only if they are in basic HTML format. Thus the new advanced web pages which do not follow this paradigm in order to make the website more user-friendly only create extra hassles for these people. Also audio screen readers(ASR) have a noisy audio interface.In case of noisy environment performance of ASR degrades.Also , These available systems require use of a keyboard which is very difficult for blind people to recognize and remember the characters of the keyboard.

All these are some drawbacks of the current system which we will overcome in the system we are developing.

Proposed System

This project proposes a python based application, designed specifically for visually impaired people. It provides a voice based mailing service wherein they could read and send mails on their own, without any guidance through their gmail accounts by merely providing voice inputs for the same.

The different use cases of the system are:

- Read an email
- Compose a new mail
- Go to sent mails
- Go to inbox
- Go to deleted mails

The VMAIL system has a very intuitive and user-friendly interface and can be used easily by partially as well as completely visually impaired people to access mails easily. Hence dependence of visually challenged on other individuals for their activities associated with mail can be condensed.

The application will use IVR- Interactive voice response, thus sanctioning everyone to control their mail accounts using their voice only and to be able to read, send, and perform all the other useful functionalities. The system will ask the user with voice commands to perform a certain action and the user will respond to it. The main advantage of this system is that the use of the keyboard is completely eliminated , the user will have to respond through voice only.

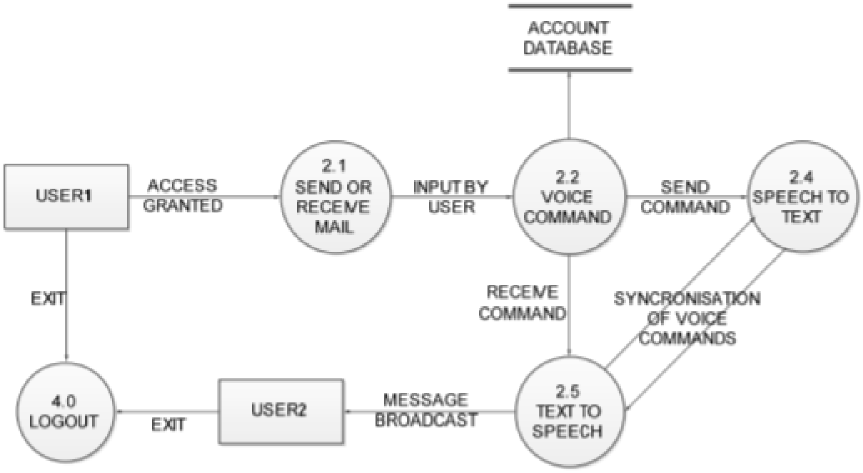


Fig 1. System Architecture

Implementation

In this system mainly three types of technologies are used namely:

1. Interactive voice response (IVR)
2. Speech to text (STT)
3. Text to speech (TTS)

● Interactive voice response:

Interactive voice response (IVR) is a technology that allows a computer to interact with humans through the use of voice and DTMF tones input via a keypad. In telecommunications, IVR allows customers to interact with a company's host system via a keyboard or by speech recognition, after which services can be inquired about through the IVR dialogue. IVR systems can respond with pre-recorded or dynamically generated audio to further direct users on how to proceed.

The purpose of an IVR is to take input, process it, and return a result. The term voice response unit (VRU) is sometimes used as well. DTMF decoding and speech recognition are used to interpret the user's response to voice prompts. IVR allows the user to interact with an email host system via a system keyboard, after that users can easily service their own enquiries by listening to the IVR dialogue. IVR systems generally respond with pre-recorded Audio voice to further assist users on how to proceed.

Other technologies include using text-to-speech (TTS) to speak complex and dynamic information, such as e-mails, news reports or weather information. IVR technology is mainly being used for hands-free operation. TTS is computer generated synthesized speech that is no longer the robotic voice traditionally associated with computers. Real voices create the speech in fragments that are spliced together (concatenated) and smoothed before being played to the caller.

● Speech Recognition (Speech to text)

Speech recognition is the interdisciplinary subfield of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

Speech recognition works using algorithms through acoustic and language modeling. Acoustic modeling represents the relationship between linguistic units of speech and audio signals; language modeling matches sounds with word sequences to help distinguish between words that sound similar.Often, hidden Markov models are used as well to recognize temporal patterns in speech to improve accuracy within the system.

The pros of speech recognition software are it is easy to use and readily available. Speech recognition software is now frequently installed in computers and mobile devices, allowing for easy access. The downside of speech recognition includes its inability to capture words due to variations of pronunciation, its lack of support for most languages outside of English and its inability to sort through background noise. These factors can lead to inaccuracies.

Speech recognition performance is measured by accuracy and speed. Accuracy is measured with word error rate. WER works at the word level and identifies inaccuracies in transcription. A variety of factors can affect computer speech recognition performance, including pronunciation, accent, pitch, volume and background noise.

Speech recognition is used to characterize the broader operation of deriving content from speech which is known as speech understanding. In our case , we use speech to text in order to comprehend the commands given by the user and accordingly interpret them via IVR and perform the required operation.

- ## Speech synthesis (Text to speech)

Speech synthesis is the synthetic production of speech. A automatic data handing out system used for this purpose is called a speech synthesizer, and may be enforced in software packages and hardware products. A text-to-speech (TTS) system converts language text into speech. Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database.

Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood clearly. An intelligible text to speech program permits individuals with reading disabilities to concentrate on written words on a computing device. Several computer operating systems have enclosed speech synthesizers .

The text to speech system consists of 2 parts:-front-end and a back-end. The front-end consist of 2 major tasks. Firstly, it disciple unprocessed text containing symbols like numbers and abstraction into the equivalent of written out words. This method is commonly known as text, standardization, or processing. Front end then assigns spoken transcriptions to every word, and divides and marks the text into speech units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end—often referred to as the synthesizer—then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody (pitch contour, phoneme durations), which is then imposed on the output speech.

Text-to-speech (TTS) is a type of speech synthesis application that is used to create a spoken sound version of the text in a computer document, such as a help file or a Web page. TTS can enable the reading of computer display information for the visually challenged person, or may simply be used to augment the reading of a text message.

Current TTS applications include voice-enabled e-mail and spoken prompts in voice response systems. TTS is often used with voice recognition programs.In our case, TTS is used to read out the different mails and provide feedback as to what is there on the screen to the user.

Tools and Technologies used

Frontend : HTML, CSS, Javascript, Bootstrap

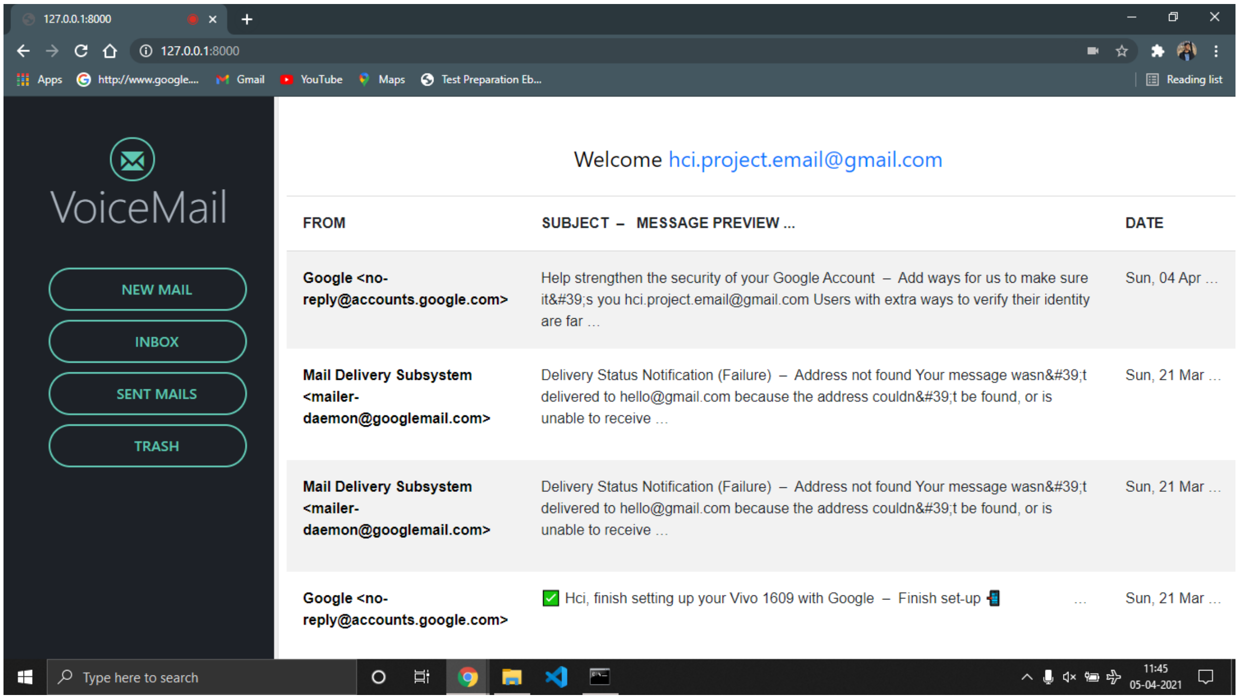
Backend:
Django -- a high-level Python web framework that enables rapid development of secure and maintainable websites.

APIS used:

1. Gmail API - The Gmail API is used to interact with users' Gmail inboxes and settings, and supports several popular programming languages, such as Java, JavaScript, and Python.
2. gtts (text to speech)- gTTS (Google Text-to-Speech), a Python library and CLI tool to interface with Google Translate's text-to-speech API. Writes spoken mp3 data to a file, a file-like object (bytestring) for further audio manipulation, or stdout. It features flexible pre-processing and tokenizing.
3. Speech to text- Accurately convert speech into text using an API powered by Google's AI technologies.

User Interface

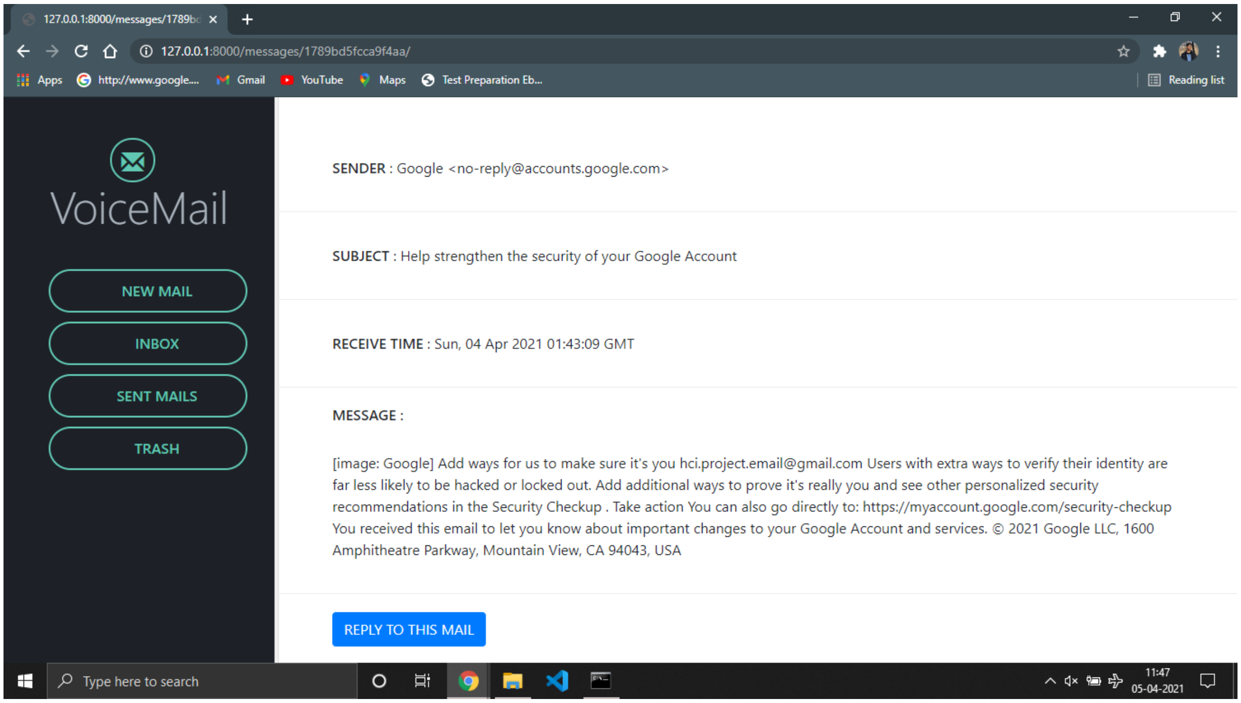
Inbox Screen



Welcome hci.project.email@gmail.com

FROM	SUBJECT – MESSAGE PREVIEW ...	DATE
Google <no-reply@accounts.google.com>	Help strengthen the security of your Google Account – Add ways for us to make sure it&#39;s you hci.project.email@gmail.com Users with extra ways to verify their identity are far ...	Sun, 04 Apr ...
Mail Delivery Subsystem <mailer-daemon@googlemail.com>	Delivery Status Notification (Failure) – Address not found Your message wasn&#39;t delivered to hello@gmail.com because the address couldn&#39;t be found, or is unable to receive ...	Sun, 21 Mar ...
Mail Delivery Subsystem <mailer-daemon@googlemail.com>	Delivery Status Notification (Failure) – Address not found Your message wasn&#39;t delivered to hello@gmail.com because the address couldn&#39;t be found, or is unable to receive ...	Sun, 21 Mar ...
Google <no-reply@accounts.google.com>	✅ Hci, finish setting up your Vivo 1609 with Google – Finish set-up 📲 ...	Sun, 21 Mar ...

Open Mail Screen



SENDER : Google <no-reply@accounts.google.com>

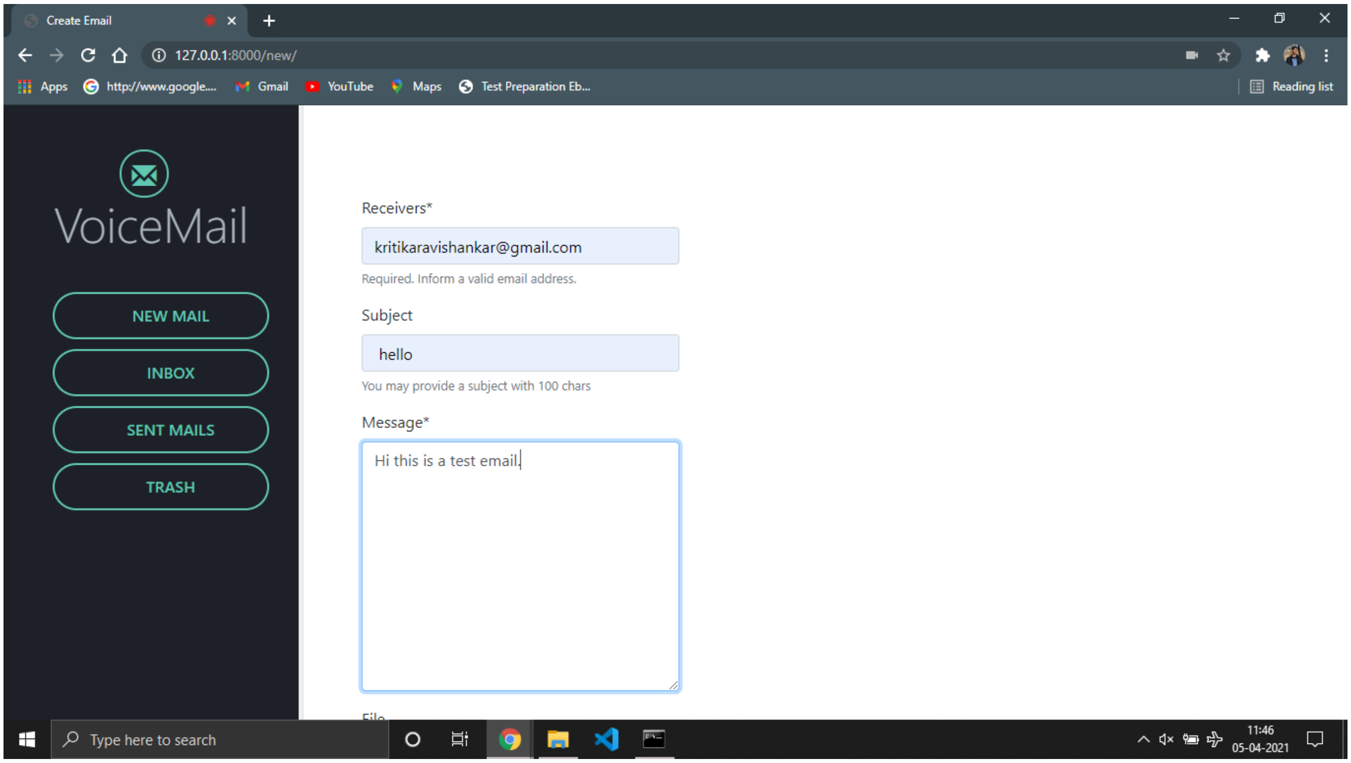
SUBJECT : Help strengthen the security of your Google Account

RECEIVE TIME : Sun, 04 Apr 2021 01:43:09 GMT

MESSAGE :

[image: Google] Add ways for us to make sure it's you hci.project.email@gmail.com Users with extra ways to verify their identity are far less likely to be hacked or locked out. Add additional ways to prove it's really you and see other personalized security recommendations in the Security Checkup . Take action You can also go directly to: https://myaccount.google.com/security-checkup You received this email to let you know about important changes to your Google Account and services. © 2021 Google LLC, 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA

Compose mail Screen



Receivers*

kritikaravishankar@gmail.com

Required. Inform a valid email address.

Subject

hello

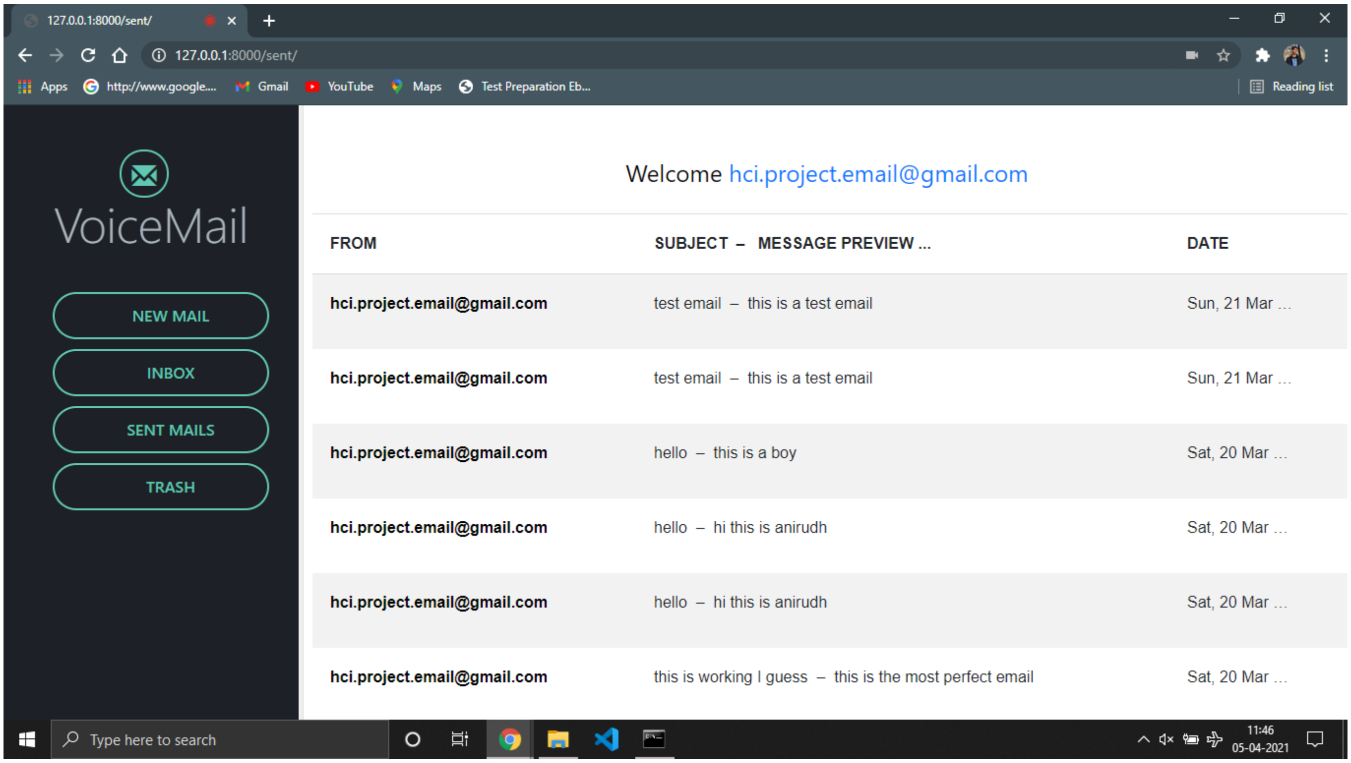
You may provide a subject with 100 chars

Message*

Hi this is a test email.

File

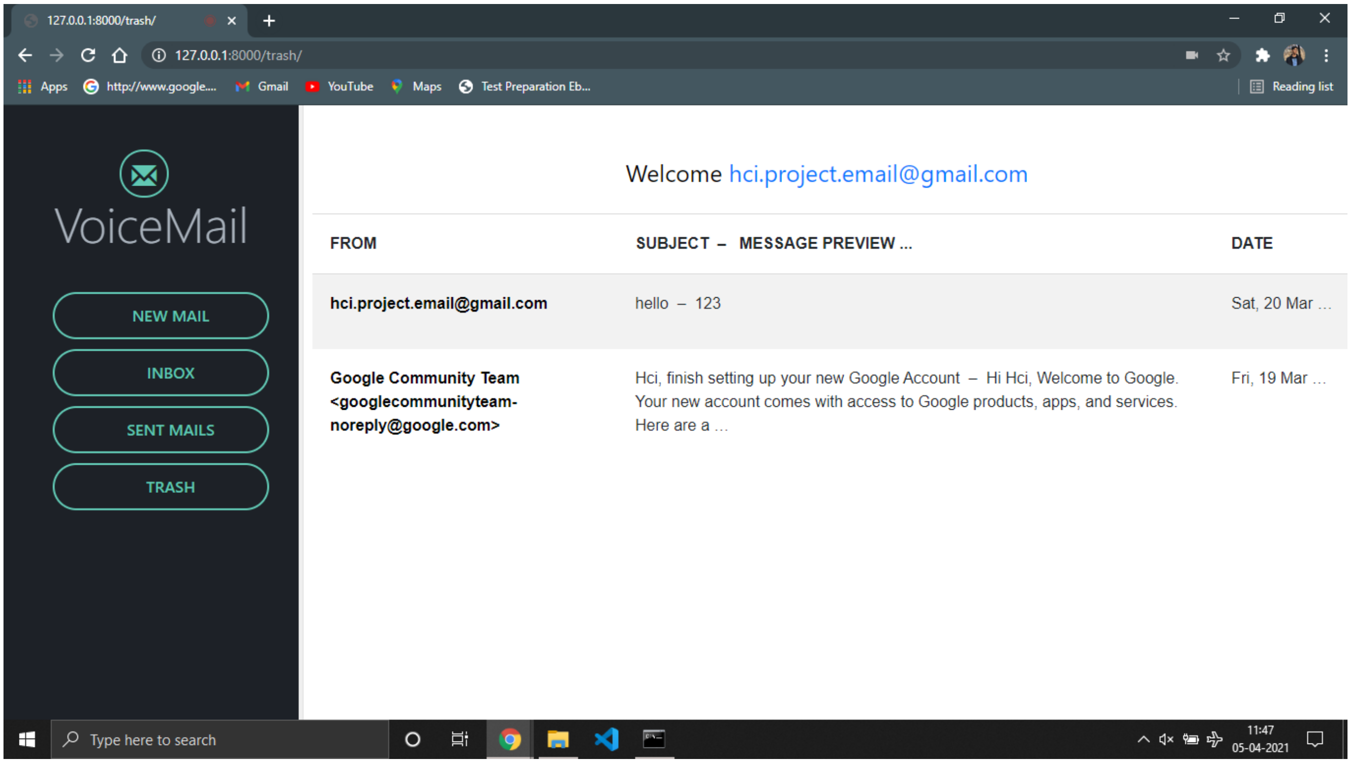
Sent Mail Screen



Welcome hci.project.email@gmail.com

FROM	SUBJECT – MESSAGE PREVIEW ...	DATE
hci.project.email@gmail.com	test email – this is a test email	Sun, 21 Mar ...
hci.project.email@gmail.com	test email – this is a test email	Sun, 21 Mar ...
hci.project.email@gmail.com	hello – this is a boy	Sat, 20 Mar ...
hci.project.email@gmail.com	hello – hi this is anirudh	Sat, 20 Mar ...
hci.project.email@gmail.com	hello – hi this is anirudh	Sat, 20 Mar ...
hci.project.email@gmail.com	this is working I guess – this is the most perfect email	Sat, 20 Mar ...

Trash Mail Screen



Welcome hci.project.email@gmail.com

FROM	SUBJECT – MESSAGE PREVIEW ...	DATE
hci.project.email@gmail.com	hello – 123	Sat, 20 Mar ...
Google Community Team <googlecommunityteam-noreply@google.com>	Hci, finish setting up your new Google Account – Hi Hci, Welcome to Google. Your new account comes with access to Google products, apps, and services. Here are a ...	Fri, 19 Mar ...

HCI Objectives achieved

1.Strive for consistency
By utilizing familiar icons, colors, menu hierarchy, call-to-actions, and user flows when designing similar situations and sequence of actions. Standardizing the way information is conveyed ensures users are able to apply knowledge from one click to another; without the need to learn new representations for the same actions.

2.Offer Informative feedback

The user should know where they are at and what is going on at all times. For every action there should be appropriate, human-readable feedback within a reasonable amount of time.
Our project reads out the mail as well as all the associated details like receiver,subject and attachments before confirming if the user wants to send it.

3.Aesthetic Design
Aesthetics in designing HCI systems have been mainly studied as a source for decoration or visualizing information.Since this is a system for the visually impaired,the major focus is on the usability and UX. Thus we keep the UI consistent and minimalistic with:

- Simple plain light reflective/absorbing colors
- Visual Balance – Symmetry / Asymmetry
- Low visual noise – No clutter or crowding

4.Cater to Diverse Usability
Our application caters to not only the visually impaired but also is accessible to a normal person who may use his voice or use the mouse to navigate through the application. It is suitable for the expert and novice users. Also the application can be accessed through both desktop and mobile devices and hence is very convenient to use from a wide variety of devices by a wide variety of users.

5.Reduce short term memory load
Human attention is limited and we are only capable of maintaining around five items in our short-term memory at one time. Therefore, interfaces should be as simple as possible with proper information hierarchy, and choosing recognition over recall. Recognizing something is always easier than recall because recognition involves perceiving cues that help us reach into our vast memory and allowing relevant information to surface. For example, we often find the format of multiple choice questions easier than short answer questions on a test because it only requires us to recognize the answer rather than recall it from our memory.
Our application makes use of the standard gmail format and recognisable icons which make navigation very easy. Also it supports standard and simple voice commands which are used to navigate the application.

Conclusion and Future Scope

This proposed system helps in overcoming some drawbacks that were earlier faced by the blind people in accessing emails.We have eliminated the concept of using keyboard shortcuts along with screen readers which will help reducing the cognitive load of remembering keyboard shortcuts

This project is proposed for the betterment of society.This project aims to help the visually impaired people to be a part of growing digital India by using the internet and also aims to make the lives of such people quite easy. Also, the success of this project will encourage developers to build something more useful for visually impaired or illiterate people, who also deserve an equal standard in the society. It has been observed that nearly 60% of the total blind population across the world is present in India.

Voice could be extended to image attachments and other options such as indentation, fonts etc., that are available with normal email.

References

- 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).
- Ummuhanysifa U.,Nizar Banu P K , “Voice Based Search Engine and Web page Reader”. In the International Journal of Computational Engineering Research (IJCER). Pages 1-5.
- G. Shoba, G. Anusha, V. Jeevitha, R. Shanmathi. “AN Interactive Email for Visually Impaired”. In International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 2014 on Pages 5089-5092.
- The Radicati website. [Online]. Available: http://www.radicati.com/wp/wp content/uploads/2014/01/EmailStatistics-Report-2014-2018-Executive-Summary.pdf.