

PROFESSIONAL TRAINING REPORT
at
Sathyabama Institute of Science and Technology
(Deemed to be University)

Submitted in partial fulfillment of the requirements for the award
of Bachelor of Technology Degree in Information Technology

By
Catherine Jenifer. R

38120022



DEPARTMENT OF INFORMATION TECHNOLOGY
SCHOOL OF COMPUTING
SATHYABAMA INSTITUTE OF SCIENCE AND
TECHNOLOGY
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119. TAMILNADU

DECEMBER 2020



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" Grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119.

Phone: 044 - 2450 3150 / 51 / 52 / 54 / 55 Fax: 044 - 2450 2344

www.sathyabama.ac.in



SCHOOL OF COMPUTING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the Bonafide work of CATHERINE JENIFER.R (Reg.no. 38120022) who carried out the project entitled "VOICE TO TEXT CONVERTER" under our supervision from June 2020 to December 2020.

Internal Guide

Dr. Y. Bevish Jinila M.E., Ph.D.,

Head of the Department

Dr. R. Subhashini M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I Catherine Jenifer. R hereby declare that the Project Report entitled “**VOICE TO TEXT CONVERTER**” done by me under the guidance of Dr. Y. Bevish Jinila M.E., Ph.D., at Sathyabama Institute of Science and Technology (Deemed to be University), Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600119, is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Information Technology.

DATE: 21/01/2021

PLACE: CHENNAI
CANDIDATE

SIGNATURE OF THE

ACKNOWLEDGEMENT

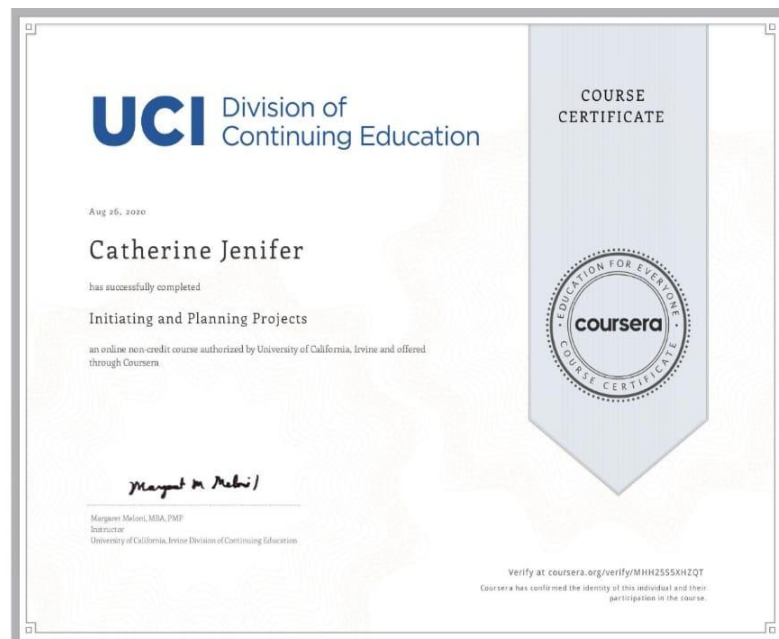
I am pleased to acknowledge our sincere thanks to Board of management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey our thanks to **Dr. T. SASIKALA M.E., Ph.D., Dean, School of Computing** and **Dr. R. SUBHASHINI M.E., Ph.D., Head of the Department, Department of Information Technology** for providing us the necessary support and details at the right time during the progressive reviews.

I would like to express our sincere and deep sense of gratitude to our Project Guide **Dr. Y. Bevish Jlnila M.E., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express our thanks to all Teaching and Non-teaching staff members of the Department of **INFORMATION TECHNOLOGY** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE





RETECH
SOLUTIONS PVT. LTD.
certified company

ABSTRACT

VOICE TO TEXT CONVERTER

Speech to text converting is one form of the fast-growing engineering technologies. Nearly 20% of the people in the world are suffering from various disabilities. Many of them are blind or unable to use their hands effectively. They can use this application to easily communicate with others with the help of computers. I have developed a speech-to-text input method for web systems. The system is provided as a JavaScript library and dynamic HTML documents. Web developers can embed it on their web page by inserting only one line in the header field of an HTML document. This project helps the disabled people and also helps people to cope with the speed of the real world.

TO WHOMSOEVER IT MAY CONCERN
This is to certify that CATHERINE JENIFER B student of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY, BTECH/IT has undergone the Internship Training through online in our concern entitled **PYTHON** from 12th Aug 2020 to 26th Aug 2020 in relevant departments related to their academic studies.

During the above period, the performance was good & we wish great success in all your future endeavours.

Technology drives the world. We are the people who are

J. meru

Authorised Signature



No-31, 1st Floor, Alagesan Street, Tambaram West, Chennai-45.

Office : 044-42077204 | +91-8122266653

www.retechsolutions.in

CHAPTER NO	TITLE NAME	PAGE NO
	ABSTRACT	6
	LIST OF FIGURES	8
	LIST OF ABBREVIATION	9
1	INTRODUCTION	10
	1.1 GENERAL	10
	1.2 OUTLINE OF THE PROJECT	10
	1.3 AVAILABLE TECHNOLOGY FOR SPEECH RECOGNITION	10
	1.4 CLASSIFICATION OF VOICE	12
2	RECOGNISING SYSTEM	
	AIM AND SCOPE	13
	2.1 AIM	13
	2.2 PROBLEM STATEMENT	13
	2.3 SCOPE	13
3	SYSTEM ANALYSIS AND DESIGN	14
	3.1 GENERAL	14
		14
	3.2 PROGRAMMING LANGUAGES USED	15
		16
	3.3 SYSTEM REQUIREMENTS	
	3.4 PROJECT DESCRIPTION	
4	RESULT AND DISCUSSION	17
	4.1 ADVANTAGES AND DISADVANTAGES	17
	4.2 RESULTS	18
5	CONCLUSION AND FUTURE WORK	18
	5.1 CONCLUSION	18
	5.2 FUTURE WORK	19
	REFERENCE	20
	APPENDIX	20
	A. SCREENSHOTS	22
	B. SOURCE CODE	

FIGURE NO	FIGURE NAME	PAGE NO
3.1	DATA FLOW	15
3.2	ARCHITECTURE BLOCK DIAGRAM	16
4.1	OUTPUT 1	20
4.2	OUTPUT 2	20
4.3	OUTPUT 3	21
4.4	OUTPUT 4	21

LIST OF ABBREVIATIONS

ABBREVIATIONS

HTML

CSS

JS

EXPANSION

HYPERTEXT MARK-UP LANGUAGE

CASCADING STYLE SHEETS

JAVA SCRIPT

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Speech recognition is a feature that gives us the ability to perform tasks using our spoken words as input. Speech recognition is gradually becoming a part of our lives in the form of voice assistants such as Alexa, Google Assistant, and Siri. Whether it's dictating words to your device to compose a document, doing a web search using voice, or controlling your computer using speech — speech to text conversion is making our life faster and comfortable. It has the potential to replace traditional forms of human to machine interface input devices, such as keyboards. A future where humans are able to interact with machines just by using their speech and bodily movements is not very far.

1.2 OUTLINE OF THE PROJECT

Human interact with each other in several ways such as facial expression, eye contact, gesture, mainly speech. The speech is primary mode of communication among human being and also, the most natural and efficient form of exchanging information among human in speech. Speech-to-text conversion (STT) system is widely used in many application areas. In the educational field, STT or speech recognition system is the most effective on deaf or dumb students.

1.3 AVAILABLE TECHNOLOGY FOR SPEECH RECOGNITION

As part of a program of research on speech-to-speech translation, we review some of the available technologies for speech recognition, the first component in any voice-based MT system.

Microsoft Speech API

Microsoft Speech API (SAPI) allows access to Windows' built-in speech recognition and speech synthesis components. The API was released as part of the OS from Windows 98 forward. The most recent release, Microsoft Speech API 5.4, supports a small number of languages: American English, British English, Spanish, French, German, simplified Chinese, and traditional Chinese. Because it is a native Windows API, SAPI isn't easy to use unless you're an experienced C++ developer.

Microsoft Server-Related Technologies

The Microsoft Speech Platform provides access to speech recognition and synthesis components that encourage the development of complex voice/telephony server applications. This technology supports 26 different languages, although it primarily just recognizes isolated words stored in a predefined grammar ([http://msdn.microsoft.com/en-us/library/ hh361571\(v=office.14\).aspx](http://msdn.microsoft.com/en-us/library/ hh361571(v=office.14).aspx)). Microsoft also provides the Microsoft Unified Communications API (UCMA 3.0), a target for server application development that requires integration with technologies such as voice over IP, instant messages, voice call, or video call. The UCMA API allows easy integration with Microsoft Lync and enables developers to create middle-layer applications.

Google Web Speech API

In early 2013, Google released Chrome version 25, which included support for speech recognition in several different languages via the Web Speech API. This new API is a JavaScript library that lets developers easily integrate sophisticated continuous speech recognition feature such as voice dictation in their Web applications. However, the features built using this technology can only be used in the Chrome browser; other browsers don't support the same JavaScript library.

1.4 CLASSIFICATION OF VOICE RECOGNISING SYSTEM

Speech recognition system can be classified in several different types by describing the type of speech utterance, type of speaker model and type of vocabulary that they have the ability to recognize. The challenges are briefly explained below:

A.Types of speech utterance

Speech recognition are classified according to what type of utterance they have ability to recognize. They are classified as:

- 1) Isolated word:** Isolated word recognizer usually requires each spoken word to have quiet (lack of an audio signal) on both side of the sample window. It accepts single word at a time.
- 2) Connected word:** It is similar to isolated word, but it allows separate utterances to „run-together“ which contains a minimum pause in between them.
- 3) Continuous Speech:** it allows the users to speak naturally and in parallel the computer will determine the content.
- 4) Spontaneous Speech:** It is the type of speech which is natural sounding and is not rehearsed.

B. Types of speaker model

Speech recognition system is broadly into two main categories based on speaker models namely speaker dependent and speaker independent Journal of Applied and Fundamental Sciences

- 1) Speaker dependent models:** These systems are designed for a specific speaker. They are easier to develop and more accurate but they are not so flexible.
- 2) Speaker independent models:** These systems are designed for variety of speaker. These systems are difficult to develop and less accurate but they are very much flexible.

C. Types of vocabulary

The vocabulary size of speech recognition system affects the processing requirements, accuracy and complexity of the system. In voice recognition system: speech-to-text the types of vocabularies can be classified as follows:

- 1) Small vocabulary:** single letter.
- 2) Medium vocabulary:** two or three letter words.
- 3) Large vocabulary:** more letter words.

CHAPTER 2

AIM AND SCOPE

2.1 AIM

Speech recognition technology is one from the fast-growing engineering technologies. It has a number of applications in different areas and provides potential benefits. Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition systems in those particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input. This project is designed and developed keeping that factor into mind, and a little effort is made to achieve this aim. My project is capable to recognize the speech and convert the input audio into text.

2.2 PROBLEM STATEMENT

There is a lot of open-source development happening in this field with newer use cases being envisioned for proper adoption. Lack of standardization of speech recognition libraries and browsers needing to seek user permission for listening to microphone input due to privacy concern is also holding it back.

2.3 SCOPE

Speech recognition can be implemented in the browser using JavaScript Web Speech API. The Web Speech API enables the web app to accept speech as input through the device's microphone and convert the speech into text by matching the words in the speech against the words in its vocabulary.

CHAPTER 3

SYSTEM ANALYSIS

3.1 GENERAL

3.1.1 USER INTERFACE

The user interface (UI) is the point of human-computer interaction and communication in a device. This can include display screens, keyboards, a mouse and the appearance of a desktop. It is also the way through which a user interacts with an application or a website

3.2 PROGRAMMING LANGUAGES USED

HTML5

Hypertext Markup Language is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets and scripting languages such as JavaScript.

CSS3

Cascading Style Sheets (CSS) is a stylesheet language used to describe the presentation of a document written in HTML or XML (including XML dialects such as SVG, MathML or XHTML). CSS describes how elements should be rendered on screen, on paper, in speech, or on other media.

JAVASCRIPT

JavaScript (JS) is a lightweight, interpreted, or just-in-time compiled programming language with first-class functions. While it is most well-known as the scripting language for Web pages, many non-browser environments also use it, such as Node.js, Apache CouchDB and Adobe Acrobat. JavaScript, often abbreviated as JS, is a high-level, interpreted scripting language that conforms to the ECMAScript specification. JavaScript has curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions.

3.3 SYSTEM REQUIREMENTS

- Operating System: Windows/Mac
- RAM: 4GB
- Processor: 64x 1.0Ghz processor
- ROM: 8GB

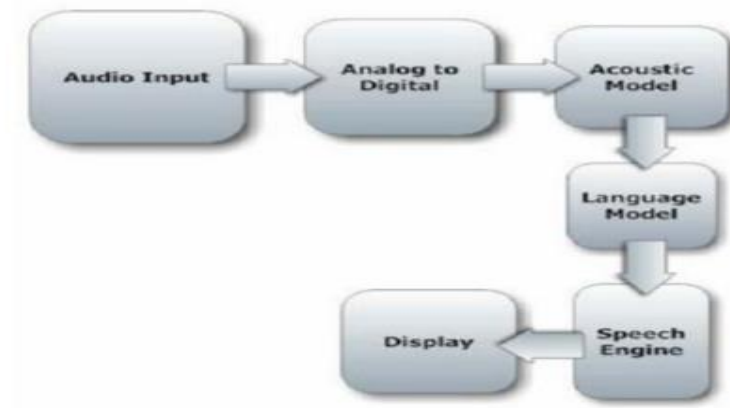


Fig3.1 DATA FLOW

3.4 PROJECT DESCRIPTION

This feature checks for words and phrases in the speech input and provides the identified words as output text. Speech recognition can be implemented in the browser using JavaScript Web Speech API. The Web Speech API enables the web app to accept speech as input through the device's microphone and convert the speech into text by matching the words in the speech against the words in its vocabulary.

The speech recognition feature in its current form is free to use, highly developed, and gives reasonably accurate results. It needs better adaptation and more devices and browsers to support it for wider acceptance. There is a lot of open-source development happening in this field with newer use cases being envisioned for proper adoption. Lack of standardization of speech recognition libraries and browsers needing to seek user permission for listening to microphone input due to privacy concern is also holding it back.

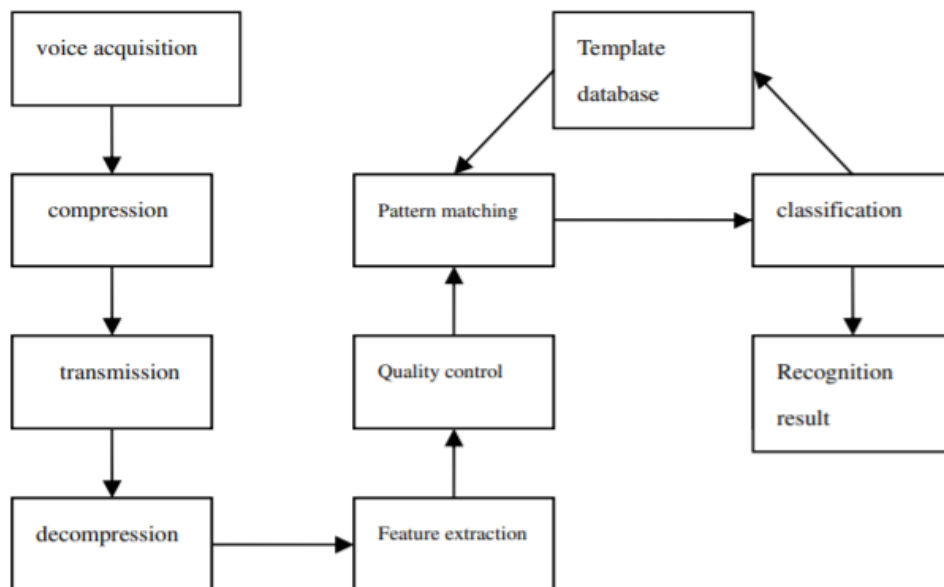


Fig 3.2 ARCHITECTURE BLOCK DIAGRAM

CHAPTER 4

RESULTS AND DISCUSSION

4.1 ADVANTAGES AND DISADVANTAGES

4.1.1 ADVANTAGES

- Converting Voice to Text are hard to ignore.
- Improved information Accuracy.
- Enhanced Focus.
- Able to write the text through both keyboard and voice input.
- Requires less consumption of time in writing text.
- Provide significant help for the people with disabilities.
- Lower operational costs.

4.1.2 DISADVANTAGE

- Low accuracy.
- Not good in the noisy environment.
- Lack of Misinterpretation.
- Voice recognition software won't always put your words on the screen completely accurately.
- Time Costs and Productivity.
- Accents and Speech Recognition.
- Background Noise Interference.
- Physical Side Effects.

4.2 RESULTS

The authentication procedure requests from the user to pronounce a random sequence of digits. After capturing speech and extracting voice features, individual voice characteristics are generated by registration algorithm. The central process unit decides whether the received features match the stored voiceprint of the customer who claims to be, and accordingly grants authentication. In this work, the architecture of a sopc-based voiceprint identification system is presented.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

In an era where voice assistants are more popular than ever, an API like this gives you a quick shortcut to building bots that understand and speak human language. Adding voice control to your apps can also be a great form of accessibility enhancement. Users with visual impairment can benefit from both speech-to-text and text-to-speech user interfaces.

5.2 FUTURE WORK

This work can be taken into more detail and more work can be done on the project in order to bring modifications and additional features. The current software doesn't support a large vocabulary, the work will be done in order to accumulate a greater number of samples and increase the efficiency of the software. The current version of the software supports only few areas of the notepad but more areas can be covered and effort will be made in this regard.

REFERENCES

1. <https://www.w3schools.com/html/default.asp>
2. https://www.w3schools.com/html/html_css.asp

APPENDIX

A. SCREENSHOTS

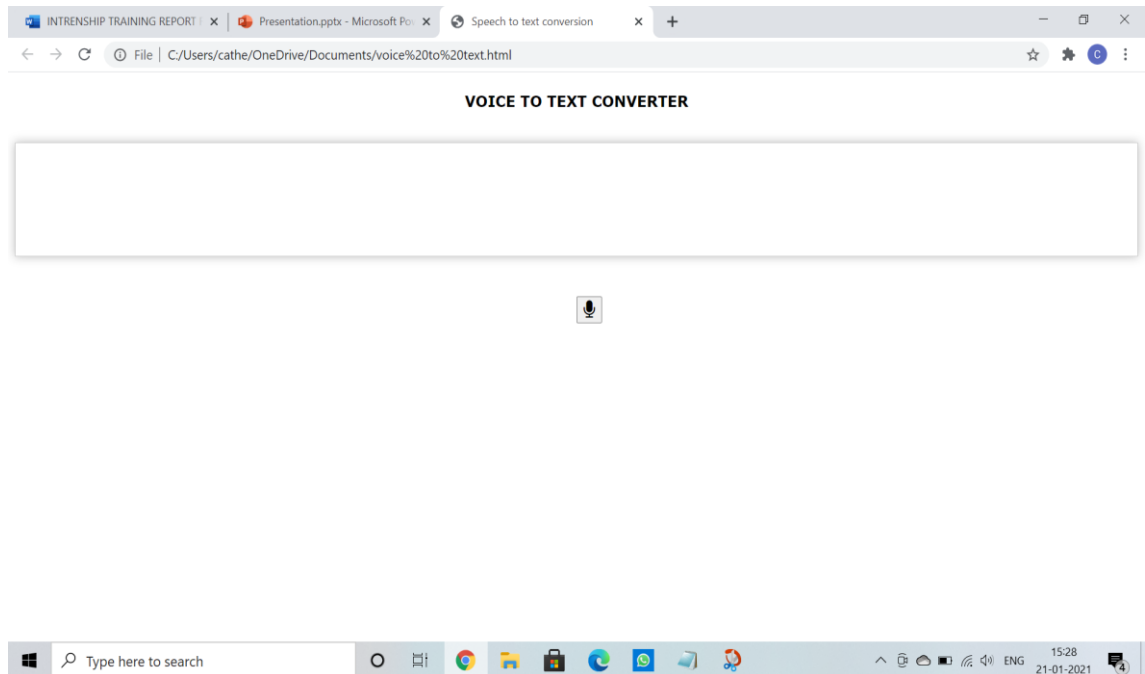


Fig 4.1 OUTPUT 1

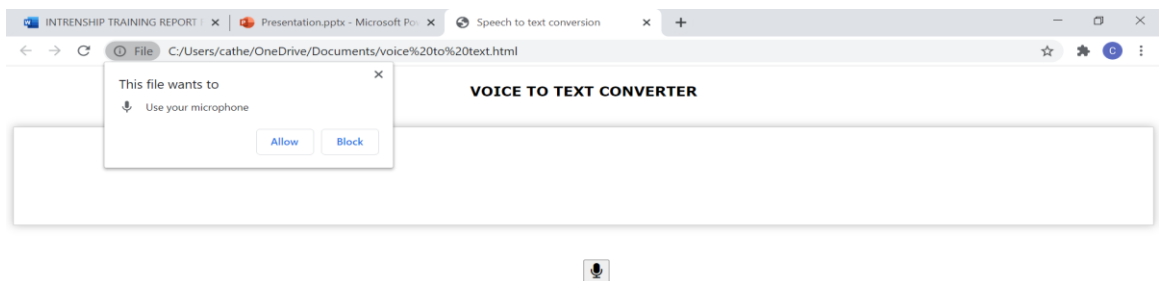


Fig4.2

OUTPUT 2

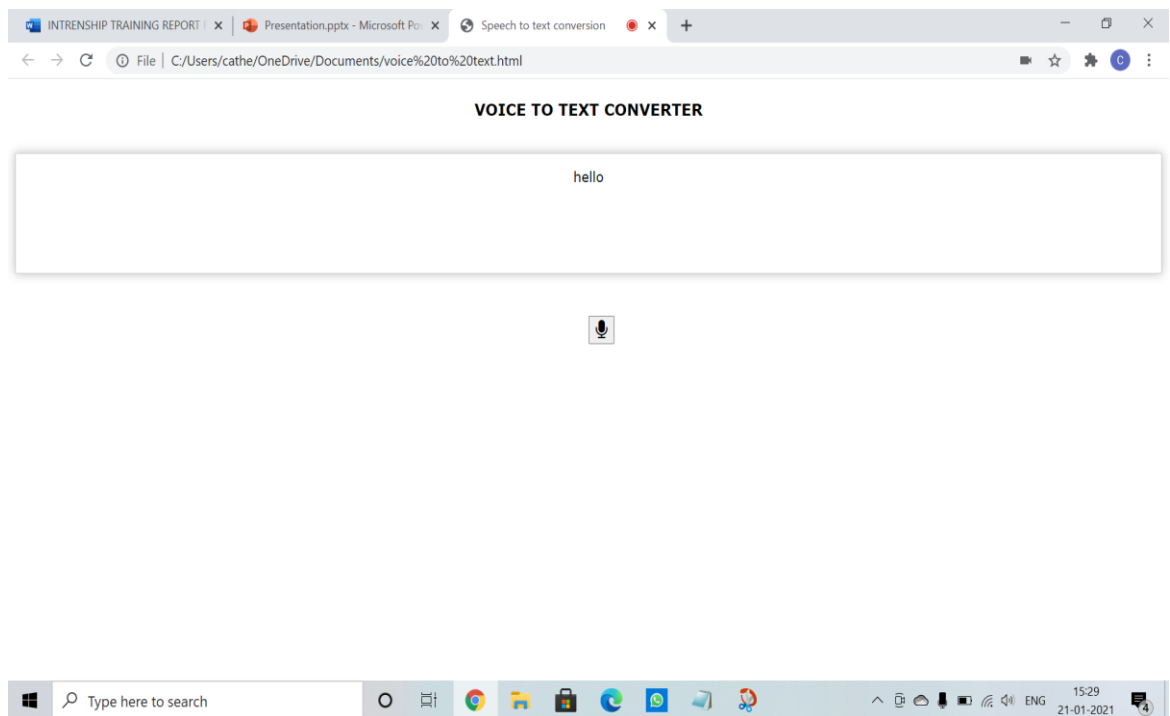


Fig4.3 OUTPUT 3

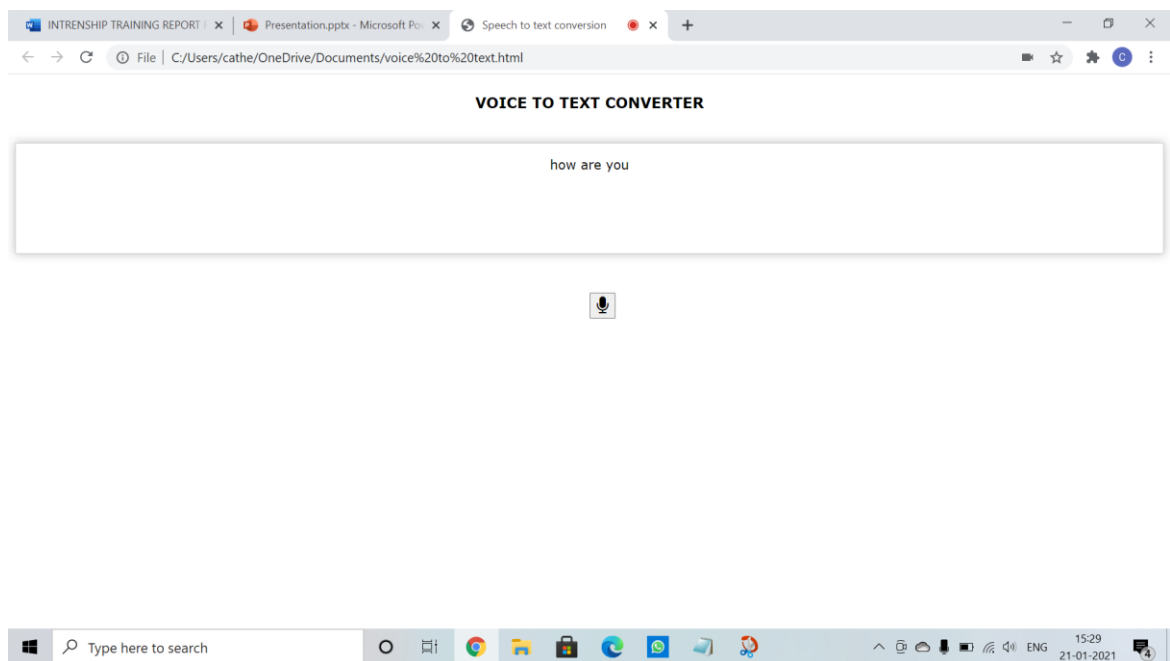


Fig4.4 OUTPUT 4

B. SOURCE CODE

```
<!DOCTYPE html>
<html>
<head>
  <title>Speech to text conversion</title>
  <link rel="stylesheet" type="text/css"
href="https://cdnjs.cloudflare.com/ajax/libs/font-
awesome/4.6.1/css/font-awesome.min.css"/>
  <style type="text/css">
    body
    {
      font-family:verdana;
      text-align:center;
      background-position: center center;
    }
    #result
    {
      height: 100px;
      border: 1px solid #ccc;
      padding: 10px;
      box-shadow:0 0 10px 0 #bbb;
      margin-top:35px;
      margin-bottom: 35px;
      font-size: 14px;
      line-height: 25px;
    }
    button
```

```

{
    font-size: 20px;
    position: absolute;
    top: 240px;
    left: 50%;
}
</style>
</head>
<body>
    <h4 align="center">VOICE TO TEXT CONVERTER</h4>
    <div id="result"></div>

    <button onclick="startConverting();" <i class="fa fa-
microphone"></i></button>

    <script type="text/javascript">
        var r=document.getElementById('result');

function startConverting()
{
    if('webkitSpeechRecognition' in window)
    {

        var speechRecognizer=new webkitSpeechRecognition();
        speechRecognizer.continuous=true;
        speechRecognizer.interimResults=true;
        speechRecognizer.lang='en-IN';
        speechRecognizer.start();
    }
}
    </script>

```

```

var finalTranscripts="";
    speechRecognizer.onresult=function(event)
    {
        var interimTranscripts=' ';
        for(var i=event.resultIndex; i<event.results.length; i++)
        {
            var transcript=event.results[i][0].transcript;
            transcript.replace("\n","<br>");
            if(event.results[i].isFinal)
            {
                finalTranscripts+=transcript;
            }
            else
            {
                interimTranscripts+=transcript;
            }
        }
        r.innerHTML=finalTranscripts+'<span
style="color:#999">'+interimTranscripts+'</span>';
    };
    speechRecognizer.onerror=function(event){
    };
}

else
{
    r.innerHTML='your browser is not supported';
}
}

```

</script>

</body>

</html>