SPEECH / TEXT NOTE APP USING JAVA SCRIPT AND HTML

Suma S
Student, MCA, Jain University, Bangalore, Karnataka, India

ABSTRACT

Speech recognition is a vast research field for researchers in modern era. Earlier, the human language was processed by the computer system for speech recognition. Thus, the main objective is to develop recognition system which improves human to human communication by enabling human-machine communication by processing of text or speech. Various applications of speech recognition systems are present and these all includes various research challenges.

Speech to text or text to speech is a part of Natural Language Processing which is a subfield of Artificial Intelligence. In Speech Recognition, spoken words/sentences are translated into text by computer. It is also known as Speech to Text (STT). Speech / Text note app could be very useful in number of applications. Especially in personal assistant bot, dictation, voice command based control system, audio transcriptions, quick notes with audio support, voice based authentication, etc.

Web Speech API , It's a very powerful browser interface that allows you to record human speech and convert it into text. We will also use it to do the opposite - reading out strings in a human-like voice. Speech / Text note app can be classified into two main areas, dictation and human-computer dialogue systems.

Features

- **Text input:** *Users are provided with a text box where they can enter the required text in the software.*

- **Speech rate:** *Users can even alter the speech speed for application to read out text by choosing the appropriate rate provided by the software.*

INTRODUCTION

The science that is most directly related to processing of human language is natural language processing. The dealing of this science directly to the natural language makes it different from other processing related activity in the field of application: the human language. NLP and Understanding is the state of art that is quite demanding these days. The research in this field has been started 50 years ago, but because of limitations of resources that are required in processing the speech, it was not implemented then in commercial applications. In today world the computer dependency expand the field of speech processing. In various science fi movies like Star Wars, Sneakers, Star Trek, Red Dwarf, UFO, Blade runner and many more the concept of natural language processing has been used. The

science that is most directly related to processing of human language is natural language processing. The dealing of this science directly to the natural language makes it different from other processing related activity in the field of application: the human language. NLP and Understanding is the state of art that is quite demanding these days. The research in this field has been started 50 years ago, but because of limitations of resources that are required in processing the speech, it was not implemented then in commercial applications. In today world the computer dependency expand the field of speech processing. In various science fi movies like Star Wars, Sneakers, Star Trek, Red Dwarf, UFO, Blade runner and many more the concept of natural language processing has been used. any more the concept of natural language processing has been used. any more the concept of natural language

EPRA International Journal of Research and Development (IJRD)

processing has been used. any more the concept of natural language processing has been used. any more the concept of natural language processing has been used. any more the concept of natural language processing has been used. any more the concept of natural language processing has been used. **v** any more the concept of natural language processing has been usedThe science that is most directly related to processing of human language is natural language processing. The dealing of this science directly to the natural language makes it different from other processing related activity in the field of application: the human language. NLP and Understanding is the state of art that is quite demanding these days. The research in this field has been started 50 years ago, but because of limitations of resources that are required in processing the speech, it was not implemented then in commercial applications. In today world the computer dependency expand the field of speech processing. In various science fi movies like Star Wars, Sneakers, Star Trek, Red Dwarf, UFO, Blade runner and many more the concept of natural language processing has been used.Speech recognition allows you to provide input to an application with your voice. Just like clicking with your mouse, typing on your keyboard, or pressing a key on the phone keypad provides input to an application, speech recognition allows you to provide input by talking. In the desktop world, you need a microphone to be able to do this. In the VoiceXML world, all you need is a telephone. he science that is most directly related to processing of human language is natural language processing. The dealing of this science directly to the natural language makes it different from other processing related activity in the field of application: the human language. NLP and Understanding is the state of art that is quite demanding these days. The research in this field has been started 50 years ago, but because of limitations of resources that are required in processing the speech, it was not implemented then in commercial applications. In today world the computer dependency expand the field of speech processing. In various science fi movies like Star Wars, Sneakers, Star Trek, Red Dwarf, UFO, Blade runner and many more the concept of natural language processing has been used. he science that is most directly related to processing of human language is natural language processing. The dealing of this science directly to the natural language makes it different from other processing related activity in the field of application: the human language. NLP and Understanding is the state of art that is quite demanding these days. The research in this field has been started 50 years ago, but because of limitations of resources that are required in

processing the speech, it was not implemented then in commercial applications. In today world the computer dependency expand the field of speech processing. In various science fi movies like Star Wars, Sneakers, Star Trek, Red Dwarf, UFO, Blade runner and many more the concept of natural language processing has been used. Speech recognition systems that require a user to train the system to his/her voice are known as speaker-dependent systems. If you are familiar with desktop dictation systems, most are speaker dependent. Because they operate on very large vocabularies, dictation systems perform much better when the speaker has spent the time to train the system to his/her voice.

Speech recognition systems that do not require a user to train the system are known as speaker independent systems. Speech recognition in the Voice XML world must be speaker independent. Think of how many users (hundreds, maybe thousands) may be calling into your web site. You cannot require that each caller train the system to his or her voice. The speech recognition system in a voice-enabled web application MUST successfully process the speech of many different callers without having to understand the individual voice characteristics of each caller.

Recognition accuracy is an important measure for all speech recognition applications. It is tied to grammar design and to the acoustic environment of the user. You need to measure the recognition accuracy for your application, and may want to adjust your application and its grammars based on the results obtained when you test your application with typical users.

EXISTING SYSTEM

A critical machine learning based review is defined which addresses the various challenging tasks of **speech recognition system** in NLP. In the **existing systems**, the **recognition** rate is very less and the noise ration during the **recognition** process creates a problem.

The performance of the audio input system degrades due to noise from the outer sources. Accuracy and reliability of the system is affected by the unwanted input and low output result. The fault tolerance capacity lacks in this case. User responsiveness is also one of the challenges, it happens when the resources are not ready and user starts to speak the command and then it leads to problem of synchronizing the data with multiple applications (media, phone, navigation)

PROPOSED SYSTEM

Speech recognition is a thriving domain with many important applications. It's easy to predict that speech recognition research will continue as well as

EPRA International Journal of Research and Development (IJRD)

important practical applications will be created. Accurate speech recognition is not so hard problem so it should be solved in a foreseeable future. And it's not about AI because it's obvious that most of the speech recognition issues are not caused by the lack of understanding but rather a lack of good algorithms.

Noises, accents and so on are just purely technical problems which will be eventually solved. Researches often consider speech recognition in a noisy environment as a standalone problem with a practical goal to build an application that works. At the same time our knowledge about speech fundamentally improves from day to day and the goals are more and more ambitious. Recent BABEL programs aims to improve support for non-English languages for example and it's planned that we will have quite good step forward in a next few years. Some leading researchers are working on language-independent speech recognition. The accuracy on the standard test sets also improves from year to year. And voice applications are already in every smartphone.

Like computers started to play chess better than human speech recognition soon will be done better by computers too. Importantly, that will add some important knowledge about nature as a whole and human brain in particular. So speech recognition is an important step to our exploration of the nature laws.

LITERATURE REVIEW

Anupam Choudhary et. al. (2012) described the speech recognition process using the approach of AI. The recognition method used is language mode, trigram model and acoustic model. No GUI is used, acoustic model interface with the telephony system to manage spoken dialogues by the speaker.

Alexandre Trilla (2012) worked on the approach of Automatic Speech Recognition using NLP technique. It depicts the production of sound from the text i.e. text to speech synthesis and vice versa i.e. known as automatic speech recognition.

D D Doye et. al. (2015) they worked on the approach of new non linear time alignment model rather than DTW algorithm. They worked for finding suitable time alignment algorithm for the Marathi language. They took 46 monosyllabic confusing alphabets and 46 confusing names for their work. They main feature used in this research were Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC) and Linear Prediction Coefficient (LPC)

Dr. Kavita R. et. al. (2014) They proposed a work on digitizing the audio into samples by using the concept of sampling. The MFCC feature is used for extraction process. These coefficients are used for matching the Tamil database through the DTW approach. This main focus of this paper is security of extracting and matching by using the DTW and mathematical approaches.

Elyes et. al. (2014) The hybrid approach of SVM/HMM is used for Arabic ASR on triphones modeling. They used the Arabic speech recognition system that is based on triphones are 64.68 % with HMMs, 72.39 % with MLP/HMM and 74.01 % for SVM/HMM hybrid model.

SPEECH TO TEXT

The Web Speech API is actually separated into two totally independent interfaces. We have SpeechRecognition for understanding human voice and turning it into text (Speech -> Text) and SpeechSynthesis for reading strings out loud in a computer generated voice (Text -> Speech). We'll start with the former.

The Speech Recognition API is surprisingly accurate for a free browser feature. It recognized correctly almost all of my speaking and knew which words go together to form phrases that make sense. It also allows you to dictate special characters like full stops, question marks, and new lines.

The first thing we need to do is check if the user has access to the API and show an appropriate error message. Unfortunately, the speech-to-text API is supported only in Chrome and Firefox (with a flag), so a lot of people will probably see that message.

TEXT TO SPEECH

Speech Synthesys is actually very easy. The API is accessible through the speechSynthesis object and there are a couple of methods for playing, pausing and other audio related stuff. It also has a couple of cool options that change the pitch, rate, and even the voice of the reader.

All we will actually need for our demo is the speak() method. It expects one argument, an instance of the beautifully named SpeechSynthesisUtterance class.

PROBLEM DEFINITION

General speech recognition systems are still largely speaker/situation dependent. Often seemingly subtle changes can affect the ability of the system to recognize commands. Differences, such as the choice of microphone or the number of people in a room, can interfere with some voice systems. It can be difficult to create or even predict the environment in which a VR experience will be run, making it difficult to train the voice system under the same circumstances in which it will need to perform.

Good speaker-independent recognition is achievable, however, if some restrictions are placed

on the voice commands. Restrictions can be either a small vocabulary or a limited, well-defined grammar. The latter is often an option in applications designed to emulate military communications, which are often "by the book."
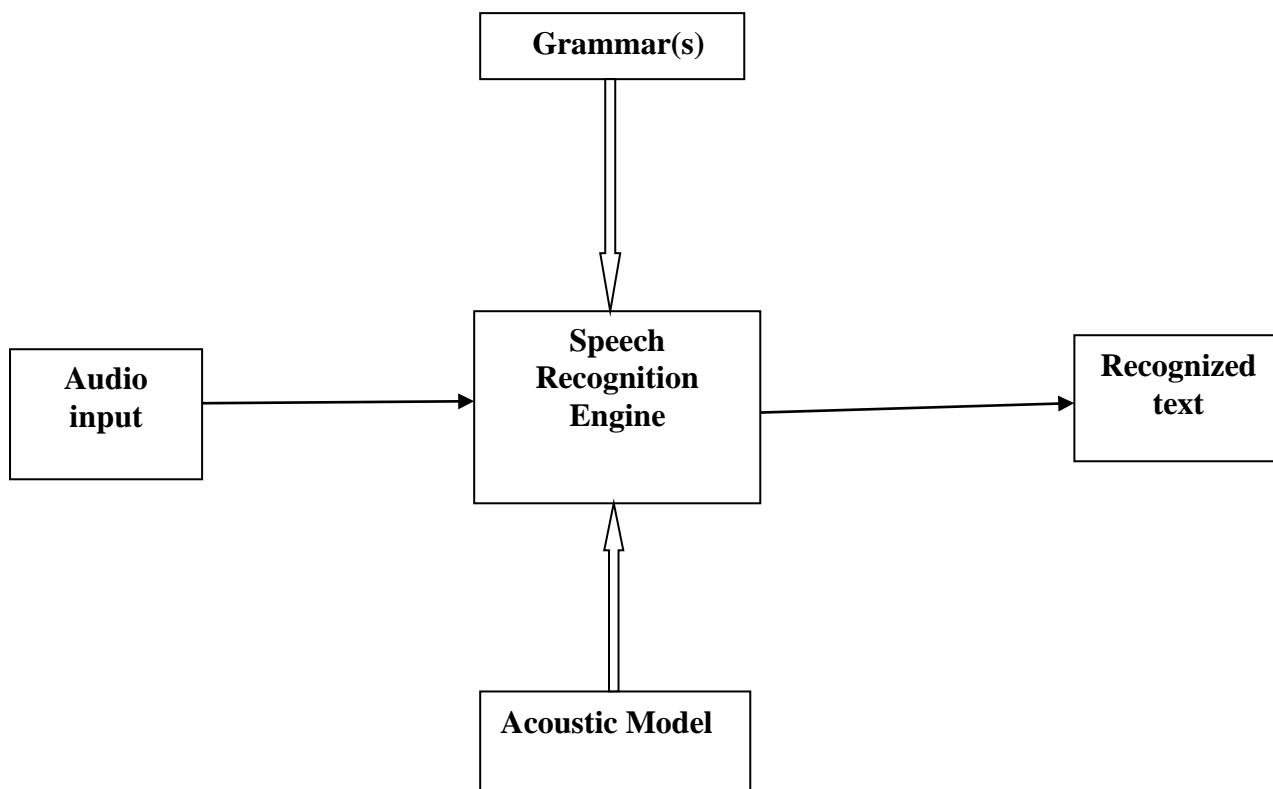
ARCHITECTURE
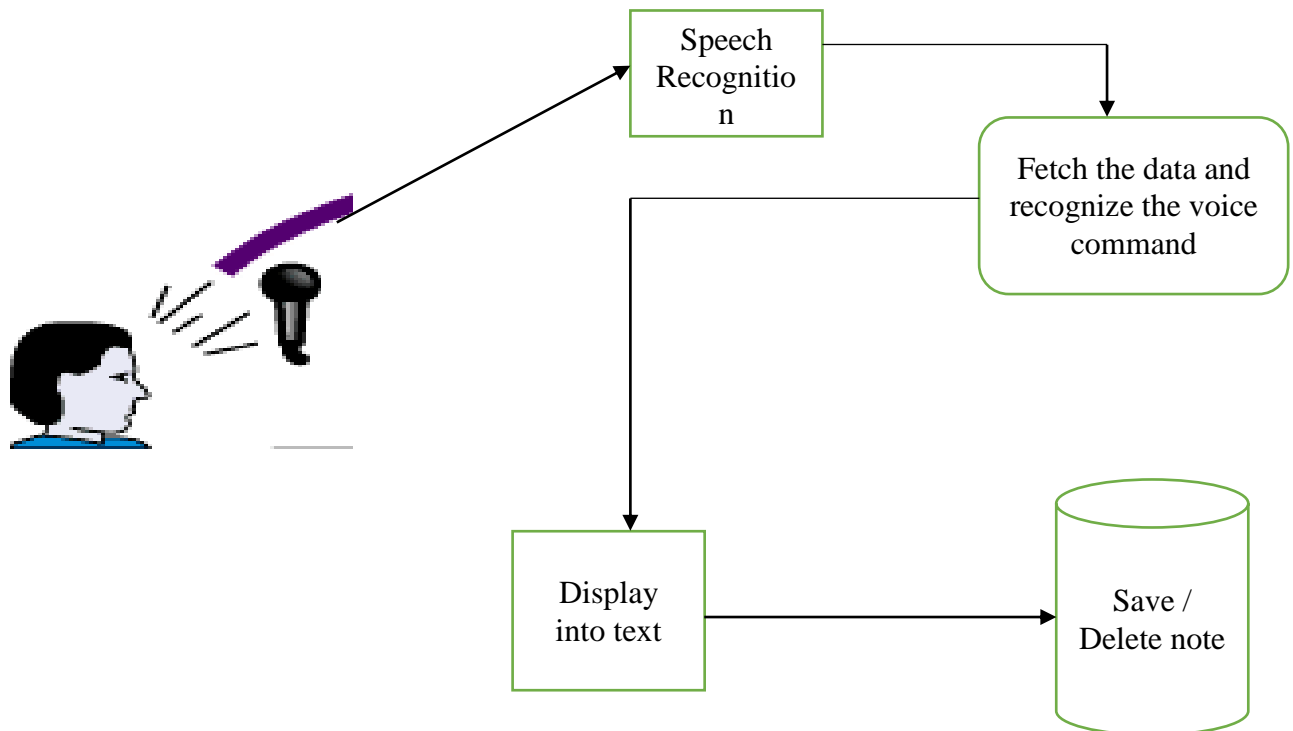
The App

To showcase the ability of the API we are going to build a simple voice-powered note app. It does 3 things:

- Takes notes by using voice-to-text or traditional keyboard input.
- Saves notes to localStorage.
- Shows all notes and gives the option to listen to them via Speech Synthesis.

EPRA International Journal of Research and Development (IJRD)

FLOW CHART

```
                        ┌──────────────┐
                        │  Grammar(s)  │
                        └──────┬───────┘
                               │
                               ▼
┌──────────┐          ┌──────────────┐          ┌──────────────┐
│  Audio   │─────────▶│    Speech    │─────────▶│  Recognized  │
│  input   │          │  Recognition │          │     text     │
└──────────┘          │    Engine    │          └──────────────┘
                      └──────▲───────┘
                             │
                      ┌──────┴───────┐
                      │ Acoustic Model│
                      └──────────────┘
```

Text to speech, abbreviated as TTS, is a form of speech synthesis that converts text into spoken voice output. Text to speech systems were first developed to aid the visually impaired by offering a computer-generated spoken voice that would "read" text to the user.

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech.

Speech to text - Speech recognition (SR) is the inter-disciplinary sub-field 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). It incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields.

Speech-to-text software is a type of software that effectively takes audio content and transcribes it into written words in a word processor or other display destination. This type of speech recognition software is extremely valuable to anyone who needs to generate a lot of written content without a lot of manual typing.

```
┌──────────────┐       ┌──────────────┐        ┌──────────────┐
│  Input from  │ ────▶ │   Speech     │ ─────▶ │   Speech     │
│  Micro phone │       │  Generation  │        │  Processing  │
└──────────────┘       └──────────────┘        └──────────────┘
                                                        │
                                                        ▼
                                                ┌──────────────┐
                                                │    Text      │
                                                │   Output     │
                                                └──────────────┘
```
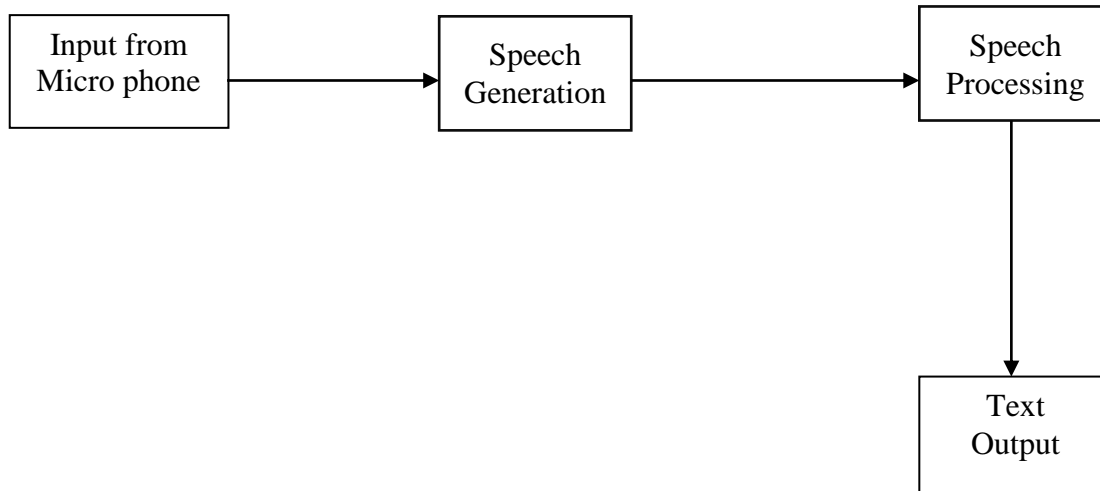
Speech recognition involves recording spoken words using either a microphone or telephone. The audio is then converted into a set of words stored digitally in the speech recognition devices.

Any speech recognition program is evaluated using two factors:

1. Accuracy (percentage error in converting spoken words to digital data)

2. Speed (extent to which the program can keep up with a human speaker)

Speech recognition technology has a long list of applications. Speech recognition software programs are used for general dictation, transcribing, using a computer hands-free, medical transcription, automated customer service etc.

CONCLUSION AND FUTURE ENHANCEMENT

A good way and process for the recognition of speech is to find a best way which can minimize the error rate during recognition. This paper defined the various recognition techniques and methods used in the current era with their pros and cons. Thus our literature indicated that efforts can be made to propose a novel approach for the recognition process which will produce better results as compare to the existing methodologies. For this better results, database of the speech signals should be last so that texting can be performed on large database. Furthermore in future research can be made when people interact with complex media indicate that

speech and language processing tools and techniques will be critical in development.

REFERENCES AND BIBILOGRAPHY

1. *Anupam Choudhary, Ravi Kshirsagar, 2012 Process Speech Recognition System using Artificial Intelligence Technique In International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-5.*
2. *Alexandre Trilla, 2012 Natural Language Processing in Text to Speech synthesis and Automatic Speech Recognition In IEEE, VOL.4*
3. *Dr. Kavitha, Nachammai, Ranjani, Shifali., 2014 Speech Based Voice Recognition System for Natural Language Processing In International Journal of Computer Science and Information Technologies, Vol. 5*
4. *Elyes Zarrouk, Yassine Ben Ayed, Faiez Gargouri, 2014 Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study, International Journal of Speech Technology ,Volume 17, Issue 3, pp 223-233.*
5. *Converting from Speech to Text with JavaScript*
6. *https://tutorialzine.com/2017/08/converting-from-speech-to-text-with-javascript*
7. *www.youtube.com(speech recognizer)*