International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **05** Issue **02** | **2021**

Voice Driven Web App

[1]Suma S, Ms. [2]Pallavi V Patil

[1]MCA Scholar, School of CS & IT, Dept. of MCA Jain (Deemed-to-be University)-560069
[2]Assistant Professor, School of CS & IT, Dept. of MCA Jain (Deemed-to-be University) – 560069

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

1. INTRODUCTION

The field of automatic speech recognition has witnessed a number of significant advances in the past 5 - 10 years, spurred on by advances in signal processing, algorithms, computational architectures, and hardware. These advances include the widespread adoption of a statistical pattern recognition paradigm, a data-driven approach which makes use of a rich set of speech utterances from a large population of speakers, the use of stochastic acoustic and language modeling, and the use of dynamic programming-based search methods.

A series of (D)ARPA projects have been a major driving force of the recent progress in research on large-vocabulary, continuous-speech recognition. Specifically, dictation of speech reading newspapers, such as north America business newspapers including the Wall Street Journal (WSJ), and conversational speech recognition using an Air Travel Information System (ATIS) task were actively investigated. More recent DARPA programs are the broadcast news dictation and natural conversational speech recognition using Switchboard and Call Home tasks. Research on human-computer dialogue systems, the Communicator program, has also started. Various other systems have been actively investigated in US, Europe and Japan stimulated by DARPA projects. Most of them can be classified into either dictation systems or human-computer dialogue systems.

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. 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.

2. PROBLEM STATEMENT

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."

3. 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)

4. 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 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.

5. METHODOLOGIES

Speech recognition involves three processes: extraction of acoustic indices from the speech signal, estimation of the probability that the observed index string was caused by a hypothesized utterance segment, and determination of the recognized utterance via a search among hypothesized alternatives.

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.

5.1 Speech to text

The Web Speech API is actually separated into two totally independent interfaces. We have Speech Recognition for understanding human voice and turning it into text (Speech -> Text) and Speech Synthesis 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

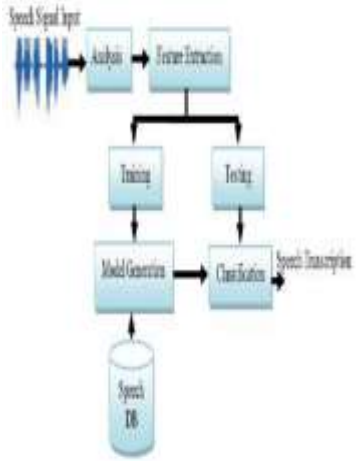
speaking and knew which words go together to form phrases that make sense. It also allows you to dictate special characters like full stops, question marks, and new lines.

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.

5.2 Text to speech

Speech Synthesis is actually very easy. The API is accessible through the speech Synthesis 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 Speech Synthesis Utterance class.

6. ARCHITECTURE



6.1 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. 3Saves notes to local Storage. Shows all notes and gives the option to listen to them via Speech Synthesis. 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: Accuracy (percentage error in converting spoken words to digital data) 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. speech and language processing tools and techniques will be critical in development.

7. FUTURE ENHANCEMENT

Speech recognition technology has made a remarkable progress in the past 5 - 10 years. Based on the progress, various application systems have been developed using dictation and spoken dialogue technology. One of the most important applications is information extraction and retrieval. Using the speech recognition technology, broadcast news can be automatically indexed, producing a wide range of capabilities for browsing news archives interactively. Since speech is the most natural and efficient communication method between humans, 19 automatic speech recognition will continue to find applications, such as meeting/conference summarization, automatic closed captioning, and interpreting telephony. It is expected that speech recognizer will become the main input device of the "wearable" computers that are now actively investigated. In order to materialize these applications, we have to solve many problems.

The most important issue is how to make the speech recognition systems robust against acoustic and lingustic variation in speech. In this context, a paradigm shitt from speech recognition to understanding where underlying messages of the speaker, that is, meaning/context that the speaker intended to convey are extracted, instead of transcribing all the spoken words, will be indispensable.

Speech recognition has achieved strong adoption in radiology over the past several years, as many hospitals and groups have sought to preserve the convenience and high value of narrative dictation while simultaneously streamlining their production process. The benefits have been clear and centered on improvements in reporting efficiency, namely, substantial report turnaround time reduction, cost savings and integration with picture archiving and communications systems (PACS) workflow

8. CONCLUSION

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.

9. REFERENCES

[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 https://tutorialzine.com/2017/08/converting-www.youtube.com(speech recognizer)
[6] Savage, J., Rivera, C., Aguilar, V., "Isolated word speech recognition using Vector Quantization Techniques and Artificial Neural Networks", 1991.
[7] Debyeche, M., Haton, J.P., Houacine, A., "Improved Vector Quantization Technique for Discrete HMM speech recognition system", International Arab Journal of information Technology, Vol. 4, No. 4, October 2007.
[8] Hatulan, R. J. F., Chan, A. J. L., Hilario, A. D., Lim, J. K. T., and Sybingco, E., "Speech to text converter for Filipino Language using Hybrid Artificial Neural Network and Hidden Markov Model", ECE Student Forum December 1, 2007 De La Salle University.
[9] Sendra, J. P., Iglesias, D. M., Maria, F. D., "Support Vector Machines For Continuous Speech Recognition", 14th European Signal Processing Conference 2006, Florence, Italy, Sept 2006.
[10] Jain, R. And Saxena, S. K., "Advanced Feature Extraction & Its Implementation In Speech Recognition System", IJSTM, Vol. 2 Issue 3, July 2011.
[11] Aggarwal, R.K. and Dave, M., "Acoustic Modelling Problem for Automatic Speech Recognition System: Conventional Methods (Part I)", International Journal of Speech Technology (2011) 14:297–308.
[12] Aggarwal, R. K. and Dave, M., "Acoustic modelling problem for automatic speech recognition system: advances and refinements (Part II)", International Journal of Speech Technology (2011) 14:309–320.
[13] Ostendorf, M., Digalakis, V., & Kimball, O. A. (1996). From HMM"s to segment models: a unified view of stochastic modeling for speech recognition. IEEE Transactions on Speech and Audio Processing, 4(5), 360–378.
[14] Yasuhisa Fujii, Y., Yamamoto, K., Nakagawa, S., "AUTOMATIC SPEECH RECOGNITION USING HIDDEN CONDITIONAL NEURAL FIELDS", ICASSP 2011: P-5036-5039.
[15] Mohamed, A. R., Dahl, G. E., and Hinton, G., "Acoustic Modelling using Deep Belief Networks", submitted to IEEE TRANS. On audio, speech, and language processing, 2010.
[16] Sorensen, J., and Allauzen, C., "Unary data structures for Language Models", INTERSPEECH 2011.
[17] Kain, A., Hosom, J. P., Ferguson, S. H., Bush, B., "Creating a speech corpus with semi-spontaneous, parallel conversational and clear speech", Tech Report: CSLU-11- 003, August 2011.
[18] Hamdani, G. D., Selouani, S. A., Boudraa, M., "ALGERIAN ARABIC SPEECH DATABASE (ALGASD): CORPUS DESIGN AND AUTOMATIC SPEECH RECOGNITION APPLICATION", The Arabian Journal for Science and Engineering, Volume 35, Number 2C, Dec 2010.
[19] NGUYEN Hong Quang, TRINH Van Loan, LE The Dat, "Automatic Speech Recognition for Vietnamese using HTK", 2004.
[20] Mathur, R., Babita, Kansal, A., "Domain specific speaker independent continuous speech recognizer using Julius", Proceedings of ASCNT – 2010, CDAC, Noida, India, pp. 55 – 60.
[21] Kumar, K. and Aggarwal, R. K., "Hindi Speech Recognition System Using HTK", International Journal of Computing and Business Research, ISSN (Online): 2229- 6166, Volume 2 Issue 2 May 2011.
[22] Gupta, R., and Sivakumar, G., "Speech Recognition for Hindi Language", IIT BOMBAY, 2006.
[23] Venkataramani, B., "SOPC-Based Speech-to-Text Conversion", 2006.

[24] Lee, K.S., "EMG-Based Speech Recognition Using Hidden Markov Models With Global Control Variables" IEEE Transactions on Biomedical Engineering, vol. 55, issue-3, pp: 930-940, March 2008.
[25] Rabiner, L. Juang, B. H., Yegnanarayana, B., "Fundamentals of Speech Recognition", Pearson Publishers, 2010.
[26] Garg, A., Nikita, Poonam, "Connected digits recognition using Distance calculation at each digit", IJCEM International Journal of Computational Engineering & Management, Vol. 14, October 2011, ISSN (Online): 2230-7893.
[27] Mishra, A. N., Biswas, A., Chandra, M., Sharan, S. N., "Robust Hindi connected digits recognition", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 4, No. 2, June, 2011.
[28] 2Syama, R. and Mary Idicula, S., "Speech Recognition for Malyalam Language", 2008.
[29] Kumar, R., Singh, C., Kaushik, S., "Isolated and Connected Word Recognition for Punjabi Language using Acoustic Template Matching Technique",2004.
[30] Thangarajan, R., Natarajan, A.M., Selvam, M., "Word and Triphone Based Approaches in Continuous Speech Recognition for Tamil Language", March 2008,
[31] Thangarajan, R., Natarajan, A.M., Selvam, M., "Syllable based Continuous Speech Recognition for Tamil", Jan 2008.
[32] Mohammad A. M. Abushariah, Moustafa Elshafei, Othman O. Khalifa, "Natural Speaker-Independent Arabic Speech Recognition System Based on Hidden Markov Models Using Sphinx Tools", May 2010.
[33] Ronzhin, A. I., Karpov, A. A., "Large Vocabulary Automatic speech recognition for Russian Language", 2004.
[34] Thang Tat Vu, Dung Tien Nguyen, Mai Chi Luong, JohnPaul Hosom, "Vietnamese Large Vocabulary continuous speech recognition", 2004.
[35] Huang Feng-Long, "An Effective approach for Chinese speech recognition on small size of vocabulary", Signal & Image Processing: An International Journal (SIPIJ) Vol.2, No.2, June 2011.
[36] Nadungodage, T. and Weerasinghe, R., "Continuous Sinhala Speech Recognizer", Conference on Human Language Technology for Development, Alexandria, Egypt, 2-5 May 2011.
[37] Raza, A., Hussain, S., Sarfraz, H., Ullah, I., and Sarfraz, Z., "An ASR System for Spontaneous Urdu Speech " in Proceedings of O-COCOSDA'09 and IEEE Xplore, 2009.
[38] Vipperla, R., Bozonnet, S., Wang, D., Evans, N. "Robust speech recognition in multi-source noise environments using convolutive non-negative matrix factorization", CHIME Workshop on Machine Listening in Multisource Environments, Sept 2011.
[39] Gemmeke, J. F., Segbroeck, M. V., Wang, Y., Cranen, B., Hamme, H. V., "Automatic speech recognition using missing data techniques: Handling of real-world data", 2011.
[40] Paul, D., and Parekh, R., "Automatic Speech Recognition of Isolated Words Using Neural Networks", Vol. 3 No. 6, IJEST-2011.
[41] Kanokphara, S., Tesprasit, V., Thongprasirt, R., "Pronunciation Variation Speech Recognition without Dictionary Modification On Sparse Database",2002.
[42] Potamianos, A., and Rose, R.C., "On Combining Frequency Warping and Spectral Shaping for HMM based Speech Recognition", IEEE international conference on acoustics, Speech, & Signal Processing, April1997.
[43] Rabiner, L. R., Wilpon, J. G., Rosenberg, A. E., "A Voice Controlled Repertory-Dialer System", The Bell System Technical Journal Vol. 59, No. 7, 1980.
[44] Aldefeld, B. Rabiner, L.R. Rosenberg, A.E. Wilpon, J.G., "Automated Directory Listing Retrieval System based on Isolated Word Recognition", Vol 68, issue 11, Nov 1980.
[45] Myers, C. S. And Rabiner, L. R., "Automated Directory Listing Retrieval System based on recognition of connected letter strings, Journal of the Acoustical Society of America, Vol. 71, No. 3, Mar 1982.
[46] Kawahara, T., "New Transcription System using ASR in Japanese Parliament", Academic Center for Computing and Media Studies, Kyoto University, 2010'