INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 06 ISSUE: 06 | JUNE - 2022 IMPACT FACTOR: 7.185 ISSN: 2582-3930

Speech to Text Conversion Using Python

1.Abdul Afzal Pasha A, Student, Department of Master of Computer Applications, University B.D.T College of Engineering, Davanagere, Karnataka, India.

2.Prof.MD.Irshad Hussain B, Associate Professor, Department of Master of Computer Application, University B.D.T College of Engineering, Davanagere, Karnataka, India.

***Abstract*—** Speech, is the most powerful way of communication with which human beings express their thoughts and feelings through different languages. The features of speech differs with each language. However, even while communicating in the same language, the pace and the dialect varies with each person. This creates difficulty in understanding the conveyed message for some people. Sometimes lengthy speeches are also quite difficult to follow due to reasons such as different pronunciation, pace and so on. Speech recognition which is an inter disciplinary field of computational linguistics aids in developing technologies that empowers the recognition and translation of speech into text. Text summarization extracts the utmost important information from a source which is a text and provides the adequate summary of the same. The research work presented in this paper describes an easy and effective method for speech recognition. The speech is converted to the corresponding text and produces summarized text. This has various applications like lecture notes creation, summarizing catalogues for lengthy documents and so on. Extensive experimentation is performed to validate the efficiency of the proposed method.

1.INTRODUCTION

Speech is the most important part of communication between human beings. Though there are different means to express our thoughts and feeling, speech is considered as the main medium for communication. Speech recognition is the process of making a machine recognize the speech of different people based on certain words or phrases. Variations in the pronunciation are quite evident in each individual's speech.

The original form of the speech is a signal, and a signal is processed such that all the information present in the signal is converted in to the text format. The feature extraction is the process of taking a signal and converting it to the required format with certain logic. Even though speech is the easiest way of communication, there exist some problems with speech recognition like the fluency, pronunciation, broken words, stuttering issues etc. All these have to be addressed while processing a speech. Text summarization is one of the major concepts used in the field of documentation. Lengthy documents are difficult to read and understand as it consumes lot of time. Text summarization solves this problem by providing a shortened summary of it with semantics. In the proposed work a combination of speech to text conversion and text summarization is implemented. This hybrid method will aid applications that require brief summary of lengthy speeches which is quite useful for documentation. The flow diagram of the proposed approach is mentioned in Figure 1,in which the speech recognition and text summarization is given as two different modules. The combination of these two modules aids any application in which summarization is required. The first and foremost step to work with NLP (Natural Language Processing) is to extract the features from the speech which has some values. If a word or a sentence is

recognized as meaningless, then it becomes an obstacle to summarization process. Even the punctuation plays a vital role in summarization as semantics is important while summarizing the text.

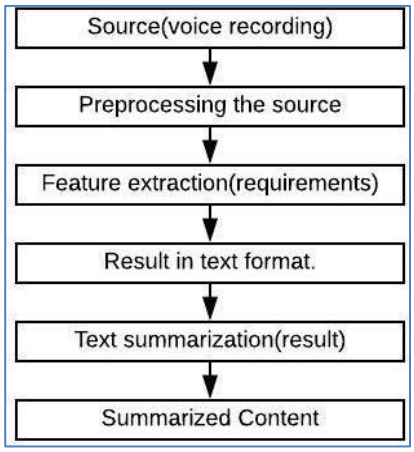

Figure 1. Speech recognition and text summarization process flow

2. LITERATURE SURVEY

Speech to text conversion finds applications in various scenarios. An effective method to gain fluency in English language that enhances the user's way of speech through correctness of pronunciation following the English phonetics was developed by Jose et al. [1]. A comparative analysis mentioning the benefits and demerits of the various sizes of vocabulary speech recognition systems was done by Shivakumar et al. [2]. This work demonstrated the role of language model in improving the accuracy of speech to text conversion with different scenarios with noises and broken words.

Yogita et al. [3] presented a multilingual speech-to text conversion system using Mel-Frequency Cepstral Coefficient (MFCC) feature extraction technique and Minimum Distance Classifier, Support Vector Machine (SVM) methods for speech classification. In [4] a model to convert natural Bengali language to text was proposed which used open source framework Sphinx 4. Authors claim an average of 71.7% accuracy for their approach in the tested dataset. English text summarisation based on association semantic rules is proposed by Wan [5].According to the author the new extraction scheme proved to have better convergence and precision performance in the extraction process. LDA is the most accepted algorithm for text classification based on a particular topic. An improvement of the same is proposed in a novel similarity computation method.

Saiyed and Sajja [6] gave an brief introduction to the categories of summarization techniques highlighting their advantages and drawbacks. This works gives insights to the researchers for selecting specific methods based on requirement. The sentence selection process modelled as a multi-objective optimization problem was described in [7]. The authors used human learning optimization algorithm for this purpose. In [8] feature extraction based on neural networks was proposed which the authors claim to be more effective compared to the online extractive options. Vythelingum et al.[9] had proposed a technique for error detection of grapheme to-phoneme conversion in text-to-speech synthesis. According to them their approach gave better error correction rate which can aid the human annotator. From the literature that was reviewed it was quite evident the requirement of speech to text conversion as well as the summarization of the same is a necessity and hence this research work. Zenkert at el. [10] introduced a cross-dimensional text summarization which uses the concept of dimensional selection and filtering. The method was experimented using the results of Multidimensional knowledge representation database. A text analyzer was developed by Devasena and Hemalatha [11] which was used to identify the structure of the text given as input.The authors claims the proposed system was able to give the results effectively which had used the automatic text categorisation and text summarisation. There exists different text summarization techniques.a detailed overview of the same is proposed in[12] by Rahimi et al. A similar study was done by Dalal and Malik also [13].

A modified approach of K Nearest Neighbor for achieving text summarization was done by Jo [14]. The author focussed more on the reliability aspect. A Vietnamese language based text summarization approach with three stages using graphs was proposed by Tran and Nguyen in [15].

The authors claims that the proposed approach was able to gather more meaningful text relevant to native speakers. Vimalaksha et al. [16] provided a method to summarize the video so as to same time and space as well as helps in archiving. An overview of text summarization focussing more on the techniques to avoid redundancy was done in [17]. Matsubayashi et al. [18] proposed a system for effective text retrieval based on the query. The authors used automatic text summarisation approach for the same. The rest of the paper is organised as follows. Section 3 gives the details of the

proposed model; Section 4 mentions the results obtained followed by Conclusions and Future scope in Section 5.

3. TECHNOLOGY OVERVIEW

The speech from the source is recorded using a microphone and the feature is extracted in text format using Google Application Programming Interface (API). However, the text extracted using the Google API does not include period (.) at the end of the sentence. This can lead to confusion in the termination of the statements. In order to avoid this, in the proposed approach a custom code has been written to provide a period after a pause of 2e+6 µs or more. This makes the sentence clearer and it is pre-processed to add period (.) and question mark (?). In order to proceed with the concept of adding a period to the extracted text, 2e+6 µs has been considered as the minimum pause time. If there is a pause for more than the said time also, the system will wait for the user input due to validation.

Since period plays a vital role in the completion of a sentence, a new sentence will be started with the concept of conjunction in the absence of period. This problem is eliminated in the proposed model by the use of temporary storage. Therefore, whenever there is a pause, the period will be added to the text and will be temporarily stored in the temporary variable. If the next sentence begins with a conjunction, the temporary variable will be cleared and the sentence will be appended to previous sentence using conjunction. Conversely to the conjunction, if the sentence begins with a subject then the temporary variable value is used and the period will be appended to the sentence. Wh-questions are expected to end with a question mark(?). Hence, whenever the sentence begins with the wh-statement, the temporary variable will hold (?) on a pause of 2e+6 or more. In case the next sentence begins with the question tag statement then the value in the temporary variable will not be used. If the sentence begins with a new subject then question mark will be appended to the end of the sentence.

The proposed method summarizes the extracted text according to the rank of the sentences which can be determined through the frequency of occurence of words. The sentence tokenize and word tokenize techniques from the packages of python NLTK are used to find the frequency of words. When the text is extracted from the input using Google API, the sentences and words in the text are obtained using sentence tokenize and word tokenize respectively. The input given by the user as speech will be converted to signal. And the signals will be converted to text format in collaboration with the Google API. In order to process the generated text with the proposed model, word tokenize and sentence tokenize is used. The complete set of a sentence is given as inputs to the sentence tokenize, every sentence is separated with the occurrence of the dot. All the sentences are given as inputs to the word tokenize, each word is separated with the occurence of the space.

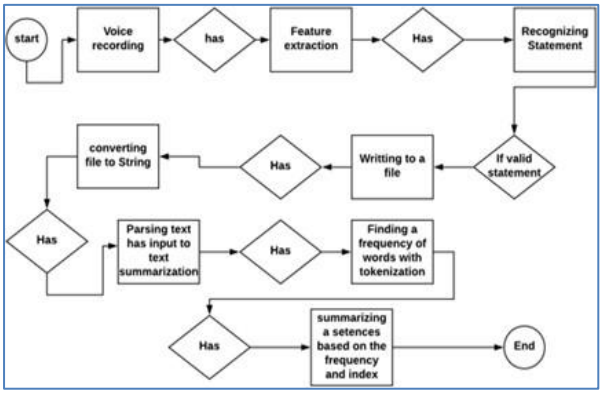

Figure 2 : Flow chart of model

When a text with proper format is used for summarizing, it is less complex to process as it is in the exact format and is often precise and clear. But this is not the case when a speech is taken as the input. Here the speech has to be converted to text and then it should be summarized. The problems to be tackled here are the occurrence of repeated words, broken words, different dialect and synonyms used to convey the message etc. Therefore, to overcome such problems, the words with less importance is eliminated. For this, a minimum and maximum range is set for the occurrence of any specific word. Even though the sentence and word frequency are used, to find the important sentence in the whole content, a ranking model is applied.

Finally, after the words are tokenized, the frequency of every word is calculated by the by the summarization algorithm in proposed model. The weight of the sentence is found with the consideration of the frequency of words. The index is ranked according to the weight of the sentence and with the identification of the index, the sentence is summarized. Python largest function is used to rank the sentence based on the weight of the sentence. The text will be summarized based on the weight of the sentences. The flow chart of implementation procedure of the proposed model is as shown in Figure 2.

4.METHODOLOGY

The algorithm for the proposed method is given below.
Step 1 − START
Step 2 − declare microphone as a source
Step 3 − declare three lists audio Recorded,
Text Format Of Record, temporary List
Step 4 -- while sentence! = exit exit
Audio Recorded = listen(source)
Extracted Text = recognize google (audio Recorded)

If (pause && next sentence of extractedText starts with subject) sentence=".”+sentence
else if(pause && next sentence of extractedText starts with conjunction) sentence=”,”+sentence
end while
[The while loop is exited with text “exit exit”]

Step 5 − declare webSpoken as file webSpoken=sentence
Step 6− declare two lists sentences, words Step 7 - sentences=sent_tokenizer(webspoken)
words =word_tokenize(sentences)
compute_frequencies(words)
Step 8 – initalizeminCut=0.1 and maxCut=0.9
If word frequency>maxCut and frequency<minCut remove the word
While ranking index! (sorted) ranking=nlargest(Sorted list of sentence)
Step 9 – Print the sentence in the order of ranking Step 10 – STOP

5.RESULTS

The recorded speech can be converted to text with the help of Google API. It is difficult to separate the text into sentence which is generated using Google API, because the extrated text does not have a period(.). To make the sentences distinct, in the proposed model, a period is appended at the end of the sentence when there is a pause. If the sentence is a wh-sentence, a question mark(?) is appended to the end of the sentence. This makes it easier to tokenize the sentences, as python string tokenization uses period to differentiate sentences. If the sentence has a pause and if it begins another sentence with the conjunction, a comma(,) is appended to the end of the sentence. This makes it easier to tokenize the sentences, as python string tokenization uses period to differentiate sentences. The proposed model considers the punctuations (‘.’, ‘,’ and ‘?’) in the recognized text. The proposed model recognition is faster when compared to the basic model (sentences without ‘.’, ‘,’ and ‘?’) recognition. The basic model summarizes the recognized text without any pre-processing. But in the proposed approach, pre-processing is used to add a period(.) at the end of each sentence to indicate the termination of a sentence. In python sentence tokenization, sentences are tokenized based on the presence of period. Though there are many punctuation marks that can be included in a sentence, the focus in the proposed model is only on period and question mark. Table 1 shows the time taken to recognize sentences with and without period and question mark respectively. Based on the recognition time, we can say that the sentences which includes period and question mark are recognized faster than the sentences without it.

CONCLUSION

Speech recognition and text summarization are two vast areas to be explored. The proposed research work aims to reduce the time and effort of manual documentation of lengthy speeches in an event. Speech recognition and text summarization can ease the work of documentation. Even for the verification of the summarized content, the system can be automated to read out the summarized content with the help of text to speech conversion. As of now, speech summarization for sentences terminating with a full stop or containing a small pause shown by comma is experimented. The future work is to include all punctuation marks in the recognized speech which helps in improving the text summarization performance. This model can be used where ever there is a requirement of summarising lengthy lectures into precise documents as the automated system will convert the speech to text and also summarise the content. It can be of great help for students to archive lecture notes from classes, conferences or seminars.

REFERENCES

[1] W. Astuti and E. B. WahyuRiyandwita, "Intelligent automatic starting engine based on voice recognition system," 2016 IEEE Student Conference on Research and Development (SCOReD),2016,pp.1-5,doi:10.1109/SCORED.2016.7810061.

[2] Shah, Z. A. Zaidi, B. S. Chowdhry and J. Daudpoto, "Real time face detection/monitor using raspberry pi and MATLAB," 2016 IEEE 10th International Conference on Application of Information and Communication

Technologies (AICT), 2016, pp. 1-4, doi: 10.1109/ICAICT.2016.7991743.
[3] C. Patil, Y. Marathe, K. Amoghimath and S. S. David, "Low Cost Black Box for Cars," 2013 Texas Instruments India Educators' Conference, 2013, pp. 49-55, doi: 10.1109/TIIEC.2013.16.

[4] S. Chaklader, J. Alam, M. Islam and A. S. Sabbir, "Black Box: An emergency rescue dispatch system for road vehicles for instant notification of road accidents and post crash analysis," 2014 International Conference on Informatics, Electronics & Vision (ICIEV), 2014, pp. 1-6, doi: 10.1109/ICIEV.2014.6850749.

[5] J. A. Lopez Leyva and V. D. AjasTerriquez, "Car Black Box System (CBBS) Using FPGA for Determine the Car orientation: Preliminary Results," 2014 International Conference on Mechatronics, Electronics and Automotive Engineering, 2014, pp. 125-128, doi: 10.1109/ICMEAE.2014.20.

[6] Daniel Hefenbrock, "Accelerating Viola-Jones face detection to FPGA-level using GPUs," Proceedings of the 2010 IEEE, 18th Annual International Symposium on Field-Programmable Custom Computing Machines, 2010, pp.11-18.

[7] C. Gao and S.-L. Lu, "Novel fpga based haar classifier face detection algorithm acceleration," in Field Programmable Logic and Applications, 2008. FPL 2008. International Conference on, Sept. 2008, pp. 373–378.

[8] V. Nair, P.-O.Laprise, and J. J. Clark, "An fpga-based people detection system," EURASIP J. Appl. Signal Process., vol. 2005, pp. 1047–1061, 2005.

[9] H. H. B. Aziz, N. H. A. Aziz and K. A. Othman, "Mobile phone car ignition system using EmbededBlue 506 Bluetooth technology," 2011 IEEE Control and System Graduate Research Colloquium, 2011, pp. 70-76, doi: 10.1109/ICSGRC.2011.5991832.

[10]J. Karim, W. M. A. B. W. Amat and A. H. A. Razak, "Car Ignition System via Mobile Phone," 2009 International Conference on Future Computer and Communication, 2009, pp. 474-476, doi: 10.1109/ICFCC.2009.116.

[11] 11. J.J.Patoliya, M.M. Desai, "Face Detection based ATM
Security System using Embedded Linux Platform ", 2nd International Conference for Convergence in Technology (I2CT), 2017.

[12] 12. M.Karovaliyaa, S.Karediab, S.Ozac, Dr.D.R.Kalbande,
"Enhanced security for ATM machine with OTP and Facial recognition features", International Conference onAdvanced Computing Tech- nologies and Applications(ICACTA), 2015.

13. Sivakumar T. 1 , G. Askok 2 , k. S. Venuprathap, "Design
and Implementation of Security Based ATM theft Monitoring system", International Journal of Engineering Inventions, Volume 3, Issue 1, 2013.

14C. Bhosale, P. Dere, C. Jadhav, "ATM security using face
and fingerprint recognition", Interna- tional Journal of Research in Engineering, Technology and Science, Volume
VII, Special Issue, Feb 2017.

15. ManojV , M. Sankar R , Sasipriya S , U. Devi E, Devika T ,
"Multi Authentication ATM Theft Prevention Using iBeacon", International Research Journal of Engineering and
Technology (IRJET).
16. L. Wang,H. Ji, Y. Shi, " Face recognition using maximum
local fisher discriminant analysis",18th IEEE International
Conference on Image Processing, 2011.

17. K.Shailaja and Dr.B.Anuradha, "Effective Face Recognition using Deep Learning based Linear.