Text- to- Speech Chrome Extension A Research Paper

[1]Sagar Dubey, [2]Yash Paithane, [3]Shreyash Shende,[4]Prof.Bharati Pawar

[1]Student, [2]Student, [3]Student,[4]Assistant Professor
[1]Computer Science and Engineering,
[1]PES College of Engineering, Chh.Sambhaji Nagar, India
[1]sagardubey1622@gmail.com, [2]yashpaithane07@gmail.com, [3]shreyashshende165@gmail.com
[4]bhartipawar15@gmail.com

***Abstract*—This exploration paper explores the development of a textbook- to- speech Chrome extension, fastening on the perpetration and evaluation of a hang- grounded approach for textbook- to- speech conversion. The paper delves into the fundamentals of textbook- to speech technology, analyses the chosen hang- grounded system, and discusses the specialized perpetration, stoner experience, availability considerations, and performance analysis of the extension. Text to Speech(TTS) conflation is a process of rephrasing natural language textbook into speech. Pieces of recorded speech induce synthesized speech and a database is maintained for storing this synthesized speech. A speech synthesizer's affair is determined through its resemblance to the person maximum and its capacity to be inferred. In recent times between the two main subsections machine literacy and deep literacy of Artificial Intelligence(AI), deep literacy has achieved huge success in the sphere of textbook to speech conflation. In this literature, a taxonomy is introduced which represents some of the deep literacy-grounded infrastructures and models popularly used in speech conflation. Different datasets that are used in TTS have also been bandied. Further, for assessing the quality of the synthesized speech some of the extensively used evaluation matrices are described. Eventually, the paper concludes with the challenges and unborn directions of the textbook- to- speech conflation system.**

I. INTRODUCTION

Currently, we're living in the digital period. Every staff associated with our life is getting digital. nearly every smartphone has a smart adjunct that can speak and communicate like a mortal. Speech recognition is one of the technologies used in those smart sidekicks. Text- to- Speech is a part of speech recognition. Text- to- speech(TTS) is a natural language modelling approach that converts textbook units into speech units for audio donation. There are multitudinous technologies used in Text- to- Speech. [1] numerous programming languages are used in these technologies. Python, a programming language substantially used in Text- to-Speech technology. There are numerous Python library e.g. TTS, pyttsx3, paddle speech. But everyone's performance is not the same. Our thesis is about measuring the effectiveness of colorful Text to- Speech technology in colorful aspects. Speech is the most important part of communication between mortal beings. Though there are different means to express our studies and feeling, speech is considered as the main medium for communication. Speech recognition is the process of making a machine fete the speech of different people grounded on certain words or expressions [2]. Variations in the pronunciation are relatively apparent in each existent's speech. The original form of the speech is a signal, and a signal is reused similar that all the information present in the signal is converted in to the textbook format. The point birth is the process of taking a signal and converting it to the needed format with certain sense. Indeed though speech is the easiest way of communication, there live some problems with speech recognition like the ignorance, pronunciation, broken words, stuttering issues etc.

All these have to be addressed while recycling a speech. Text summarization is one of the major generalities used in the field of attestation. Lengthy documents are delicate to read and understand as it consumes lot of time. Text summarization solves this problem by furnishing a docked summary of it with semantics.[3] In the proposed work a combination of speech to textbook conversion and textbook summarization is enforced. This mongrel system will prop operations that bear brief summary of lengthy speeches which is relatively useful for attestation. Text- to- speech(TTS) technology islands the gap between written textbook and spoken language, enabling druggies to convert digital textbook into audible speech. It has come an integral part of colorful operations, including assistive technologies for individualities with visual impairments, language literacy tools, and availability features in web cybersurfs and mobile bias.[4] TTS systems calculate on sophisticated algorithms that dissect textbook and induce corresponding speech sounds, replicating natural mortal speech patterns. crucial factors include Text analysis and processing Speech conflation Voice customization and selection The advancement of TTS technology has significantly bettered the quality and lightheartedness of synthesized speech, making it a precious tool for enhancing stoner experience and availability across different operations.

II. RELATED WORKS

Speech to textbook conversion finds operations in colorful scripts. An effective system to gain ignorance in English language that enhances the stoner's way of speech through correctness of pronunciation following the English phonetics [5]. A relative analysis mentioning the benefits and faults of the colorful sizes of vocabulary speech recognition systems. This work demonstrated the part of language model in perfecting the delicacy of speech to textbook conversion with different scripts with noises and broken words. [13] presented a multilingual speech- to textbook conversion system using Mel- frequence Cepstral Coefficient(MFCC) point birth fashion and Minimum Distance Classifier, Support Vector Machine(SVM) styles for speech bracket. [6] In a model to convert natural Bengali language textbook was proposed which used open-source frame Sphinx 4. Authors claim a normal of 71.7 % delicacy for their approach in the tested dataset. English textbook summarization grounded on association semantic rules. According to the author the new birth scheme proved to have better confluence and perfection performance in the birth process. LDA is the most accepted algorithm for textbook bracket grounded on a particular content. An enhancement of the same is proposed in a new similarity calculation system.[7] gave a brief preface to the orders of summarization ways pressing their advantages and downsides. This works gives perceptivity to the experimenters for opting specific styles grounded on their demand. The judgment selection process modelled as a multi-objective optimization problem was described in. The authors used mortal literacy optimization algorithm for this purpose.

In point birth grounded on neural networks was proposed which the authors claim to be more effective compared to the online extractive options. had proposed a fashion for error discovery of grapheme to- phoneme conversion in textbook- to- speech conflation.

According to them their approach gave better error correction rate which can prop the mortal commentator. From the literature that was reviewed it was relatively apparent the demand of speech to textbook conversion as well as the summarization of the same is a necessity and hence this exploration work.[8] introduced across-dimensional textbook summarization which uses the conception of dimensional selection and filtering. The system was experimented using the results of Multidimensional knowledge representation database. A textbook analyzer was developed. which was used to identify the structure of the textbook given as input. The authors claims the proposed system was suitable to give the results effectively which had used the automatic textbook categorization and textbook summarization. There exists different textbook summarization techniques. A detailed overview of the same is proposed. An analogous super stud. A modified approach of K Nearest Neighbor for achieving textbook summarization. The author riveted more on the trustability aspect.

A Vietnamese language grounded textbook summarization approach with three stages using graphs[9] The authors claims that the proposed approach was suitable to gather further meaningful textbook applicable to native speakers. handed a system to epitomize the videotape so as to same time and space Aswell as helps in archiving. An overview of textbook summarization riveting further on the ways to avoid redundancy was done in proposed a system for effective textbook reclamation grounded on the query. The authors used automatic textbook summarization approach for the same.[10].

What are TTS Software

Text- to- speech(TTS)1 or “ read audibly ” technology is an assistive technology that reads a digital textbook out loud. It converts published words on a computer or any other digital device similar as smart phones, Tablets and iPads into speech or voice. The scholars can hear the audio enumeration of the published textbook incontinently whether it's a single character or long paragraph.

III. METHODOLOGY

In this paper, we've proposed to assay the possibilities that are involved in the conversion process of the written language into speech(text-to- speech) using the programming sphere. In order to achieve this ideal, we initiated our work by examining the textbook-to- speech interfaces handed.

Hover-based Text-to-Speech Method

The hang- grounded textbook- to- speech system offers a stoner-friendly and intuitive approach to textbook- to- speech conversion. Unlike traditional TTS operations that bear unequivocal selection or commands, the hang- grounded system triggers speech conflation simply by swimming.[11] the mouse cursor over a textbook element. This flawless integration into the stoner's browsing experience enhances availability and convenience. The hang- grounded system leverages event listeners that descry mouse hang events on textbook rudiments. When a hang event is touched off, the extension activates the TTS machine, converting the textbook under the cursor into audible speech. This approach eliminates the need for druggies to manually elect textbook, making it particularly profitable for individualities with motor impairments or those seeking quick and royal textbook- to- speech conversion.

Selected Method for Text-to-Speech Conversion

For this Chrome extension, the favored system for textbook- to- speech conversion is a combination of the hang- grounded approach and the Speech conflation API handed by the cybersurfed. This approach offers several advantages flawless integration The hang- grounded system seamlessly integrates into the stoner's browsing experience, barring the need for redundant way or buttons. Availability The approach caters to individualities with visual impairments or motor difficulties by allowing them to pierce textbook-to- speech functionality without the need for homemade selection.

Cybersurfed comity The Speech Synthesis API is natively supported by utmost ultramodern web cybersurfs, icing cross- platform comity for the extension[12] This system provides a stoner-friendly and effective result for textbook- to- speech conversion within a Chrome extension terrain.

Importance of TTS Accessibility for Chrome Users

Chrome, being the most popular web cybersurfed encyclopedically, boasts a vast stoner base with different requirements and capacities. Integrating robust TTS functionality within Chrome enhances its availability for druggies with colorful disabilities. This includes individualities with visual impairments, dyslexia, or other literacy challenges who may find reading grueling or prefer audile literacy.[13] By making Chrome more accessible, we empower druggies with disabilities to navigate the digital world with lesser ease and independence.

Overview of Chrome's Built-in TTS Capabilities

Chrome has an erected- in TTS point that can read audibly web runners and named textbook. This point is accessible through the cybersurfer's settings, allowing druggies to customize voice settings, language selection, and playback speed. still, Chrome's erected-in TTS functionality lacks the inflexibility and customization options offered by devoted extensions.

IV. SPEECH RECOGNITION

Speech Recognition is the capability of machine/ program to identify words and expressions in spoken language and convert them into machine- readable format. Speech Recognition Systems can be classified on base of the following parameters. Speaker All speakers have a different kind of voice. The models hence are either designed for a specific speaker or an independent speaker. Oral Sound The way the speaker speaks also plays a part in speech recognition. Some models can fete either single utterances or separate utterance with a pause in between.[14]

Vocabulary The size of the vocabulary plays an important part in determining the complexity, performance, and perfection of the system.

1)Basic Speech Recognition Model: Each speech recognition system follow some standard way.

i)Pre-processing: The analog speech signal is converted into digital signals for after processing. This digital signal is moved to the first order pollutants to spectrally flatten the signals. This helps in adding the signal’s energy at an advanced frequence.

ii) Feature Extraction: This step finds the set of parameters of utterances that have a correlation with speech signals. These parameters, known as features, are reckoned through processing of the aural waveform. The main focus is to cipher a sequence of point vectors(applicable information) furnishing a compact representation of the given input signal. Generally used point birth ways are bandied below[15]

Linear Predictive Coding(LPC): The introductory idea is that the speech sample can be approached as a direct combination of once speech samples. The digitized signal is blocked into frames of N samples. also each sample frame is windowed to minimize signal discontinuities. Each framed window is also bus- identified. The last step is the LPC analysis, which converts each frame of autocorrelations into LPC parameter set.

Mel- frequence Cestrum-efficient(MFCC): It's a veritably important fashion and uses mortal audile perception system. MFCC applies certain way to the input signal Framing Speech surge- form is cropped to remove hindrance if present; Windowing minimizes the discontinuities in the signal; Discrete Fourier transfigure converts each frame from time sphere to frequence sphere; Mel Filter Bank Algorithm the signal is colluded against the Mel diapason to mimic mortal hail[16].

Dynamic Time wrapping: This algorithm is used for measuring the similarity between two- time series which may vary in speed, grounded on dynamic programming. It aims at aligning two sequences of point vectors(1 of each series) iteratively until an optimal match(according to a suitable criteria) between them is set up.[4]

iii)Acoustic Models It's the abecedarian part of Automated Speech Recognition(ASR) system where a connection between the aural information and phonetics is established. Training establishes a correlation between the introductory speech units and the aural compliances.

iv) Language Models This model induces the probability of a word circumstance after a word sequence. It contains the structural constraints available in the language to induce the chances of circumstance. The language model distinguishes word and expression that has an analogous sound.

v) Pattern classification It's the process of comparing the unknown pattern with being sound reference pattern and computing similarity between them. After completing the training of the system at the time of testing, patterns are classified to fete the speech. Different approaches for pattern matching are[17]

Template Based Approach: This approach has a collection of speech patterns which are stored as a reference representing dictionary words. Speech is honored by matching the uttered word with the reference template

Knowledge Based Approach: This approach takes set of features from the speech and also train the system to induce set of product rules automatically from the samples.

Neural Network Based Approach: This approach is able of working more complicated recognition task. The introductory idea is to collect and in-commercial knowledge from a variety of knowledge sources with the problem at hand[9].

Statistical Based Approach: In this approach, variations in speech are modelled statistically(e.g. HMM) using training styles.

Language Translation

In India, we've a variety of languages spoken. The 2001 Census recorded 30 languages which were spoken by further than a million native speakers and 122 which were spoken by further than 10,000 people, which is why it's veritably necessary to have operations and processes that can convert textbook from one language to another, keeping the saintship of the communication. Machine restatement(MT) is a field of Artificial Intelligence and Natural Language Processing which deals with restatement from one language to another using machine restatement system.[6]. The mortal restatement process may be described as decrypting the meaning of the source textbook, and Reencoding this meaning in the target language. Some of the machine restatement models are bandied below

i) Rule Grounded Machine translation(RBMT): Translations is generated on the base of morphological, syntactic, and semantic analysis of both the source and the target languages. Such a system correspond of collection of rules Grammar rules principally correspond of analysis of languages in terms of alphabet structures(syntax, semantic, morphology, part of speech trailing and orthographic features); bilingual or multilingual wordbook wordbook for looking up words during restatement while the software program allows effective and effective commerce of factors; and software programs to understand a process those rules.[18] are three types of rule- grounded model • Direct It's wordbook grounded. • Transfer It uses dictionaries and structural analysis into every SL input textbook after which it’s converted to intermediate representation. • Interlingual source language is converted into an intermediate language which is independent of any of the languages involved in the restatement.[4]

ii) Statistical machine translation(SMT) It's characterized by the use of machine literacy styles. SMT is a data- driven approach which uses resemblant aligned corpora and treats restatement as a fine logic problem. In that, every judgment in the target language is a restatement with probability from the source language. [4] The advanced the probability, the advanced is the delicacy of restatement and vice-versa. Basic SMT armature includes • Language model for calculating the probability of the target language • restatement model for calculating tentative probability of target language affair given source language input • Decoder model- gives the stylish restatement possible t by maximize the two-probability mentioned over.

iii) Example-based machine translation(EBMT): It's grounded on the idea of analogy. In this approach, the corpus that's used is one that contains textbooks that have formerly been restated. Given a judgment that's to be restated, rulings from this corpus are named that contain analogous sub-sentential factors.[1] The analogous rulings are also used to restate the sub-sentential factors of the original judgment into the target language, and these expressions are put together to form a complete restatement. The Analogy restatement uses three stages; matching, adaptation and recombination Matching- The SL input textbook is fractured, followed by hunt for exemplifications from database which nearly matches the input SL scrap string and the applicable fractions are picked. The TL fractions corresponding to the applicable fractions are uprooted.[2] adaptation- If the match is exact, the fractions are recombined to form TL affair, differently find the TL portion of the applicable match correspond to specific portion in SL and align them. • Recombination- Combination of applicable TL fractions in order to form legal grammatical target textbook. mongrel machine restatement The expansion of methodologies in the once decade and the preface of new operations for automated restatement have stressed the limitations of espousing one single approach to the problems of restatement. mongrel MT is a system of machine restatement that's characterized by the use of multiple machine restatement approaches within a single machine restatement system. It's a combination of RMBT and SMT system, and it makes the use of the advantages of both these styles. Statistical data is hence, put to use in generation of wordbook and syntax. Machines like IBM Watson inventor pall, Google Translate, and Microsoft translators are extensively used by colorful operations or work singly in helping effectively restated languages for better understanding.[19]

Technical Implementation Details The perpetration of the textbook- to- speech Chrome extension involved several crucial way :

1. Extension Manifest: The manifest train defines the extension's structure, warrants, and functionalities.

2. Content Script: The content script interacts with the web runner's content, listens for hang events on textbook rudiments, and triggers the TTS machine.

3. Speech Synthesis API: The Speech Synthesis API provides the medium to synthesize speech from textbook. It handles textbook analysis, phoneme conversion, and speech generation.

4.Event Handling: The extension uses event listeners to descry hang events on textbook rudiments within a web runner. When a hang event occurs, the TTS machine is actuated.

5.Voice Customization druggies can configure the extension to use different voices and acclimate speech settings, similar as rate and volume, to epitomize their experience. The extension's law utilizes JavaScript and HTML5 for the stoner interface and interacts with the Speech Synthesis API for audio generation.[20]

User Experience and Availability Considerations

The stoner experience of the textbook- to- speech Chrome extension was designed with availability and usability in mind. The extension aims to give a flawless and intuitive experience for druggies with colorful requirements and preferences. crucial considerations include:

Ease of Use: The hang- grounded system simplifies textbook- to- speech conversion, barring the need for homemade selection. Customization druggies can epitomize speech settings similar as voice, rate, and volume to match their preferences

Availability Features: The extension integrates with screen compendiums and other assistive technologies to enhance availability for druggies with visual impairments.

Usability: The extension's stoner interface is designed to be clear and intuitive, allowing druggies to fluently navigate its functionalities.

The extension provides options for customizing speech parameters, including voice selection, rate adaptation, and volume control, to insure an acclimatized experience for different druggies.

Evaluation and Performance Analysis

The evaluation of the textbook- to- speech Chrome extension involved comprehensive testing and analysis to assess its performance and effectiveness. crucial aspects of the evaluation include:

Speech Quality The lightheartedness, clarity, and delicacy of the synthesized speech were estimated to assess the quality of the TTS machine.

Performance The responsiveness and effectiveness of the extension were measured, considering factors similar as quiescence, resource consumption, and cybersurfed performance.

Availability The extension's availability features were tested to insure comity with screen compendiums and other assistive technologies

User Feedback: User feedback was collected through checks and interviews to understand their experience and identify areas for enhancement. The evaluation results demonstrated that the extension effectively delivered textbook- to- speech functionality with a stoner-friendly interface, icing availability and delivering high- quality synthesized speech.

V. *APPLICATIONS*

There's a wider range of operation for the TTS processor. In the near future it'll come a boon for the society. As it can be used by the impaired people living each over the world and can make them analogous to normal people i.e. removing the disability hedge for the people facing complexion in the society. Some of the major operations are mentioned below

- For the **Dyslexic** people, facing the problem in reading the textbook, TTS motor can be installed anywhere i.e. either in their systems or their mobile phone
- It's veritably important useful for the **dumb** peoples. As they face severe problems in making the other person who's willing to help him understand their need. So they can class fluently and in the advertisement window of TTS processor and the speech sound will make other's understand their need
- It can be used to educate the **pre-illiterate** children. As it can educate them how to gasp the words in their introductory stage of literacy.

VI. *EVALUATION METRICS*

For speech conflation, several evaluation criteria similar as Mean Opinion Scores(MOS), F0 Root- Mean- Square Error(RMSE), Voiced/ unspoken(V/ UV), Mel cepstral(MCEP) are popularly used and Table XIII provides a detailed description of these evaluation criteria . Among these MOS has been used in maximum papers for objective evaluation. In the sphere of speech conflation, speech quality is the most popular way to compare the performance of any algorithm, armature, or model. MOS calculates the average score of quality and for this reason, MOS has surfaced as the most common and popular figure for synthesized speech quality

VII. *DISCUSSION AND FUTURE RESEARCH DIRECTION*

SPSS and concatenative conflation are two primary models in TTS technology. Compared with the concatenative speech conflation system, SPSS systems have numerous strengths. They're much more flexible to transfigure the synthesized speech into different speech characteristics, feelings, and speaking ways. Three high factors that hamper the quality of synthetic speech are the quality of vocoders, aural modeling delicacy, and over smoothing.[10] In SPSS systems, RBMs have been accessibly used for modeling voice signals. For SPSS, the Gaussian admixture model performs less directly in modeling spectral envelopes' distribution over RBM.RBM replaces single Gaussian distributions by representing the spectral envelope distribution at each HMM state. This procedure significantly upgrades the conventional HMMbased speech conflation system's lightheartedness using Melcepstra and reduces the over-smoothing problem at conflation time. After that, DBN has been enforced for presenting the distribution of the spectral envelope at each HMM state and at the same time to include the dynamic features of spectral envelopes into RBM modeling

The experimental results and evaluation process shows that both DBN- HMM and RBM- HMM produce better spectral envelope parameter sequences over traditional Gaussian HMM with better conception power DBN- HMM and RBM- HMM perform most probably due to the use of Gaussian distribution. multitudinous deep literacy- grounded approaches have been suggested in recent times for converting textbook to speech conflation, still. Several ways are substantially used in interpreting textbook- to- speech, similar as CNN and RNN. The critical comparison between CNN and RNN occurs in the training period. Deep Convolutional TTS is only adequately trained at night(15 hours), while the quality of speech was nearly applicable. On the other hand, RNN generally takes a lot of time to train the model for several days or weeks, as it's suitable for an important machine and lower suitable for GPU resemblant calculation. numerous textbooks on speech conflation styles use CNN rather of RNN to resolve this problem. Due to elevated parallelizability,[16] CNN- grounded textbook to speech conflation styles works much faster than RNN- grounded ways. In recent times, the styles used for speech conflation that performs better than traditional approaches have been bettered.[20]

The AdVRNN now implements the variability of natural speech for aural modeling in speech conflation. still, the inimical Learning Scheme in AdVRNN training to break the problem of over-smoothing AdVRNN performs better than traditional speech conflation grounded on RNN. HMM- grounded SPSS has come more popular in the last many times. But the status of the synthesized speech does not reach the lightheartedness using HMM. So to upgrade the lightheartedness of synthesized speech,

DNN has been used. The aural modeling delicacy and the over-smoothing problem has been answered incompletely by LSTM. But it was set up that by using GAN, the over-smoothing problem can be answered completely. For reproducing the stochastic element in the glottal excitations, GAN performs better than DNN. also, the DNN- grounded GANs perform significantly better than deep CNN- grounded GAN. While several problems have been answered in recent times, there's still a great eventuality for development.

VIII. CONCLUSION

This exploration paper has presented the development and evaluation of a textbook- to- speech Chrome extension using a hang- grounded approach for textbook- to- speech conversion. The paper stressed the specialized perpetration details, stoner experience considerations, and performance analysis of the extension. The evaluation results demonstrated the effectiveness and usability of the hang- grounded system for enhancing availability and stoner convenience.

The exploration interest in speech conflation has been changing from clarity and intelligibility to expressiveness and lightheartedness. In the morning period of speech conflation, parametric conflation and concatenative- grounded styles have been substantially used, which hinders the lightheartedness of synthesized speech. Dissimilarly, deep literacy- grounded speech conflation styles concentrate on the lightheartedness of speech. In this paper, we've explored deep literacy- grounded speech conflation styles. We've inclined a taxonomy of speech conflation styles, showing a general block illustration of major infrastructures and models popularly used to synthesize speech and pressing their advantages and disadvantages. Different evaluation criteria and datasets with their advantages and disadvantages have been bandied. A comparison of the colorful trials was presented. Indeed if deep literacy grounded speech conflation styles have achieved tremendous success in recent times, there's still a big occasion to produce high-quality speech from textbook. With the current arrival of featherlight deep neural network infrastructures, speech can be synthesized automatically from the textbook in the near future.

REFERENCES

[1] L.R. Rabiner, B.-H.Juang, 1993 "Fundamentals Of Speech Recognition" (Prentice-Hall, Englewood Cliff)

[2] Olabe, J. C.; Santos, A.; Martinez, R.; Munoz, E.; Martinez, M.; Quilis, A.; Bernstein, J., "Real-Time Text- To-Speech Conversion System For Spanish," Acoustics, Speech, And Signal Processing, Ieee International Conference On Icassp '84. , Vol.9, No., Pp.85,87, Mar 1984..

[3] A Novel Model For Speech To Text Conversion ,Deepav.Jose, Alfateh Mustafa, Sharan R, Published In International Refereed Journal Of Engineering And Science (Irjes)Issn (Online) 2319-183x, (Print) 2319- 1821volume 3, Issue 1 (January 2014), Pp. 39-41..

[4] T. Dutoit, (1997). (An Introduction To Text-To-Speech Synthesis. Kluwer Academic Publishers, Dordrech/ Boston/London

[5] H. Hon, A. Acero, X. Huang, J. Liu, M. Plumpe. "Automatic Generation Of Synthesis Units For Trainable Text-To-Speech Systems," Proceedings Of Icassp 98 (Cd-Rom), 1998.

[6] Jose D V, Alfateh Mustafa, Sharan R, "A Novel Model for Speech to Text Conversion," International Refereed Journal of Engineering and Science (IRJES), vol 3, no. 1, 2014.

[7] K. M. Shivakumar, V. V. Jain and P. K. Priya, "A study on impact of language model in improving the accuracy of speech to text conversion system," 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, pp. 1148-1151, 2017.

[8] Y. H. Ghadage and S. D. Shelke, "Speech to text conversion for multilingual languages," 2016 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, pp. 0236-0240, 2016.

[9] Umar Nasib Abdullah, Kabir Humayun, Ahmed Ruhan, Uddin Jia., "A Real Time Speech to Text Conversion Technique for Bengali Language," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), pp. 1-4, 2018.

[10] L. Wan, "Extraction Algorithm of English Text Summarization for English Teaching," 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Xiamen, China, 2018, pp. 307-310.

[11] Saiyed S., Sajja P. S., "Review on text summarization evaluation methods," Indian Journal of Computer Science and Engineering, vol. 8, no. 4, pp. 497, 2017.

[12] R. Alguliyev, R. Aliguliyev and N. Isazade, "A sentence selection model and HLO algorithm for extractive text summarization," 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, pp. 1-4, 2016.

[13] Jain D. Bhatia and M. K. Thakur, "Extractive Text Summarization Using Word Vector Embedding," 2017 International Conference on Machine Learning and Data Science (MLDS), Noida, pp. 51-55, 2017.

[14] K. Vythelingum, Y. Estève and O. Rosee, "Error detection of grapheme-to-phoneme conversion in text-to- speech synthesis using speech signal and lexical context," 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, pp. 692-697, 2017.

[15] J. Zenkert, A. Klahold and M. Fathi, "Towards Extractive Text Summarization Using Multidimensional Knowledge Representation," 2018 IEEE International Conference on Electro/Information Technology (EIT), Rochester, MI, pp. 0826-0831, 2018.

[16] C. Lakshmi Devasena and M. Hemalatha, "Automatic Text categorization and summarization using rule reduction," IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012), Nagapattinam, Tamil Nadu, pp. 594-598, 2012.

[17] S. R. Rahimi, A. T. Mozhdehi and M. Abdolahi, "An overview on extractive text summarization," 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, pp. 0054-0062, 2017.

[18] V. Dalal and L. Malik, "A Survey of Extractive and Abstractive Text Summarization Techniques," 2013 6th International Conference on Emerging Trends in Engineering and Technology, Nagpur, pp. 109-110, 2013.

[19] A. Vimalaksha, S. Vinay, A. Prekash and N. S. Kumar, "Automated Summarization of Lecture Videos," 2018 IEEE Tenth International Conference on Technology for Education (T4E), Chennai, India, pp. 126-129, 2018.

[20] S. Biswas, R. Rautray, R. Dash and R. Dash, "Text Summarization: A Review," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), Changsha, China, pp. 231-235