Grenze International Journal of Engineering and Technology, June Issue

Canarese Real Time Text to Speech System using Concatenation Technique

Anusha K S[1], Chandrika Prasad[2],Jagadish S Kallimani[3] and Jayalakshmi D S[4]
[1]Ramaiah Institute of Technology, Dept of CSE, Bangalore, India
[2-4]Ramaiah Institute of Technology, Bangalore, India
Email: ksanusha16@gmail.com, {chandrika, Jagadish.k, jayalakshmids}@msrit.org

***Abstract*— In order to share information and to communicate with each other, translation from one language into another is important. Kannada, a Dravidian language spoken in South India, is one of the most widely used Dravidian language. People who only speak Kannada but are interested in the documents or any other relevant information. In order to communicate with all users, in this paper implements a translation of text into the Kannada speaking language for messages, documents, and images. For real-time translation, the suggested translator makes use of a Yandex API and Google API.**

Linguistics relates to the scientific study of languages. Text to speech (TTS) is a method to generate audio versions of texts that seem natural and intelligible. TTS is used in numerous services, including audio navigation, news casting, tour interpretation, etc. An effective text-to-speech algorithm produces speech signals with good prosodic being natural. By offering customers audio content, Kannada TTS can also be employed in industries like education, tourism, and business. Aphasia is an issue brought on by language deficiencies brought on by traumatic brain injury. A specific learning disorder is dyslexia. In our chaotic environment, the TTS systems aid people in saving time. It primarily aids older individuals with failing vision and those who are visually challenged in reading Kannada. People who lack literacy can rely on this system to help them become independent, which is crucial in government organizations. Additionally, it aids those with little linguistic proficiency in Kannada. The most used environment is MATLAB, and accuracy is measured by mean opinion score (MOS).

***Index Terms*— Application Program Interface (API), MATLAB (Matrix Laboratory), Mean Opinion Score (MOS), Text to Speech (TTS).**

I. INTRODUCTION

A. General introduction

One of the most comfortable and natural ways for people to engage is through voice communication. Computer programmers that can speak naturally to people have grown in popularity as this sector of technology has progressed. Linguistics is the academic study of languages. Language form, meaning, and context are some of the accepted concepts in linguistics. Aphasia is a condition brought on by language deficiencies brought on by brain injury. A stroke, a concussion, or damage to a specific section of the brain can cause speech problems. Patients with aphasia are usually treated with speech therapy. A specific learning disorder is dyslexia. Children who have dyslexia have trouble reading accurately and fluently. Additionally, they could struggle with spelling, writing, and comprehension of text. Although it is unclear what proportion of kids have it, it is the most prevalent learning

Grenze ID: 01.GIJET.9.2.133

problem. According to some experts, the percentage is between five percent and ten percent. Some estimates put the number of people who make a reading-related hint at up to 17%. As seen by users' growing reliance on speech recognition systems like Siri, Google Voice Searching, and others, automatic sound synthesis is one of the technological innovations that has received the greatest praise. Various techniques can be used to do speech recognition, based on the desired outcomes. To analyses, model, and recognize speech in the form of individual words, continuous speech, and unplanned language, a variety of techniques can be utilized, including Hidden Markov Models, Gaussian Mixture Models, and deep neural network models. Several voice synthesis API's, including the Microsoft Win 32 voice API, Speech Synthesizer API's, win 32 Speech Application Programming Interface, and speech database, are only available on the Windows platform. These techniques produce artificial speech that does not seems like Kannada because of the Kannada speech Synthesis. It is therefore simpler to produce the Kannada speech, but accent is issue.

II. Literature Survey

A. Introduction

The idea of discourse and speech acknowledgement first came around in the 1940s. The speech acknowledgment programmer was first implemented in 1952 at the Chime Labs and was intended to recognize digits in quiet environments. Considering that this was the beginning of speech recognition technology, it was for the initial criteria for conversation detection that automation and understanding speculative modelling was done in the 1940s and 1950s. The simple acoustic phonetic feature of speech sounds was supported in the 1960s by our propensity to be able to recognize short vocabularies of unique phrases. The text-to-speech system synthesizer in this project converts entered language texts into generated spoken speech and reads it out to does for-designed algorithms.
The result has been updated and adjusted to further improve speech understanding. The Mat Development's part, Sphinx and .net framework, and a number of other frameworks are just a few of the ones used to create these apps. Consequently, creators of this emerging technology are aware of how difficult it is to understand it. When they go into this subject of technology, it's quite easy to understand each speech synthesis as well as discourse acknowledgment. Everywhere when compared to conventional human conversation and computerized speech, it is incredibly strong and unostentatious, which developers must recognize and use accordingly. This project was created using the essential job operations that we often utilize daily, and it is regularly updated in addition based on user input.

B.Overview of Regional Language

Canarese or Kannada is most famous language in the southern part of India. Kannada is the state language of Karnataka. Like other Dravidian regional languages like Telugu, Tamil, and Malayalam, it has a common origin. The people who speak Kannada is known as Kannadigas. Kannada is known as the old language with its literature. In Indian languages Kannada literature is awarded with more outstanding performance in the literature contribution. Kannada language has its own script. Kannada stands as queen among all other languages. The total population of Karnataka is 6.41 cores out of that 4.4 cores people talk proper Karnataka. From BC so much of modifications are being done to Kannada language. The modifications are classified into 4 parts: Purva Halegannada, Halegannada, Nadugannada, Hosagannada. Kannada Alphabet consists of total of 49 pronunciation letters are divided into three categories as follows: 1. Vowels/Swaragalu: There are 13 vowels, each of which is a separate letter, and there are two different sorts of swaras (vowels) based on how long they take to pronounce. They are Deerga Swara, a freely existing independent vowel that can be pronounced in two matras, and Hrasva Swara, a freely existing independent vowel that can be uttered in one matra.2. Vyanjanagalu/Consonants: There are 34 consonants, and each one depends on a vowel to take an independent form. Two categories can be made of these the two Vargeeyas 3. Yogavaahakagalu: Visarga and Anuswaras are the two Yogavaahakagalu.
In a table 1 and table 2, table 3 below mentioned shows how the each vowels, classified and unclassified consonats has can unicode and its equivalent decimal value respectivley.
In the fig.1 explains about the processs for the TTS system. Text is shown after input from the keyboard. The text is divided into words, and then those words are further divided into the alphabet. Each alphabet has been assigned the unique unicode , for each unicode conatins the speech corpus. In order to obtain continuous speech, the speech corpus is combined to form the real time speech.
In the table 4, it contains the existing papers and the methodology used in those papers and the dataset, accuracy and remarks based on the analyses made by reading those papers. The different methodology used like Syllabification Synthesis, Articulator Synthesis, Direct Concatenation, Neural Architecture Search (NAS), Semi supervised end to end ASR, Unit selection Concatenation method using MATLAB and different methodology as

per the author interest. In table 5 each method with their performance analysis average is calculated. By using the table 6, each technique used for TTS system in literature survey as given method number with their respective accuracy the graph is plotted is given in fig 2. Most of the Text to speech systems are developed for therapy sessions for blind or people with less reading skills or confusion. In the literature survey real time applications are built on MATLAB environment. The mismatch of speech waves can be reduced by setting up the start and end limit for each waveform. In all the Text to speech models try to avoid mixing the speech for the perfect speech pronunciation. Julius is the real time text to speech system is built for Japanese the output is very intelligent and natural. This Julius is the first TTS system is constructed. TTS consumes more CPU cycles, because the output of each model needs to evaluate in a sequence. The speech corpus contains the voice record of both male and female. Datasets vary from small to large database from 50-100 words or else full textbook.

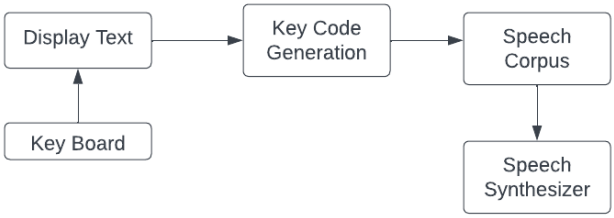


Figure 1.Block Diagram of TTS

TABLE I. UNICODE FOR INDEPENDENT VOWELS

Alphabets	Unicode	Decimal Equivalent
ಅ	0C85	3205
ಆ	0C86	3206
ಇ	0C87	3207
ಈ	0C88	3208
ಉ	0C89	3209
ಊ	0C8A	3210
ಋ	0C8B	3211
ಎ	0C8E	3214
ಏ	0C8F	3215
ಐ	0C90	3216
ಒ	0C92	3218
ಓ	0C93	3219
ಔ	0C94	3220

TABLE II. UNICODE FOR CLASSIFIED CONSONANTS

Alphabets	Unicode	Decimal Equivalent
ಕ	0C95	3221
ಖ	0C96	3222
ಗ	0C97	3223
ಘ	0C98	3224
ಙ	0C99	3225
ಚ	0C9A	3226
ಛ	0C9B	3227

ಜ	0C9C	3228
ಝ	0C9D	3229
ಞ	0C9E	3230
ಟ	0C9F	3231
ಠ	0CA0	3232
ಡ	0CA1	3233
ಢ	0CA2	3234
ಣ	0CA3	3235
ತ	0CA4	3236
ಥ	0CA5	3237
ದ	0CA6	3238
ಧ	0CA7	3239
ನ	0CA8	3240
ಪ	0CAA	3242
ಫ	0CAB	3243
ಬ	0CAC	3244
ಭ	0CAD	3245
ಮ	0CAE	3246

TABLE III. UNICODE FOR UNCLASSIFIED CONSONANTS

Alphabets	Unicode	Decimal Equivalent
ಯ	0CAF	3247
ರ	0CB0	3248
ಲ	0CB2	3250
ಳ	0CB3	3251
ವ	0CB5	3253
ಶ	0CB6	3254
ಷ	0CB7	3255
ಸ	0CB8	3256
ಹ	0CB9	3257

TABLE IV. PERFORMANCE OF EXISTING TEXT TO SPEECH MODULES

Paper Number	Methodology	Dataset	Accuracy	Remarks
[1]	Syllabification Synthesis	Ten DS,50-500words	Computational time: 65%	Time Consuming because each word is divided into symbols and then the speech is generated

[1]	Articulator Synthesis	Ten DS ,50-500 words	Computational time:88%	Produces the clear and distinct speech recognition
[2]	Direct Concatenation	“Suvarna Sampada “the 1st Sem BA textbook	----------	Need to train and test model for more datasets
[3]	Neural Architecture Search (NAS)	LJS speech dataset	16x fewer MACs, 6.5x inference, and 15x compression ratio	For more effective TTS, NAS must be paired with various types of trimming and quantization.
[4]	End to end semi-supervised ASR	Small pair of LibriSpeech Corpus	10%-to-8%-character error rate Word mistake rate: 20.6% to 18.0%	Need to train whole dataset of LibriSpeech corpus, and also try with the unsupervised leaning approach for the model
[5]	Transfer learning technique	20000+utterances of Read speech	78.2% is the average preferred score.TTS from Google and Nuance received ratings of 13.1% and 5.1%, respectively.	Considered best TTS model according to the author but the recording the speech from the unknown speaker makes it difficult.
[6]	Concatenation method using MATLAB	General Kannada sentences	93% of natural speech and 95% of accuracy	It can’t be used for the real time application. Previously trained dataset has better performance than the real time application
[7]	Unit selection Concatenation method using MATLAB	Randomly Chosen words	100% for the only Randomly Chosen words	The words are read one by one .it cause the time delay in the real time applications

[8]	Artificial Neural Networks	200 epochs	Accuracy increased 2% than Mel Frequency Cepstral	Further experiment can be done with prosody features with spectral features
[9]	Statistical computational approach	Input text is given by virtual keyboard	----------	It can be implemented for the further languages to get the different sounds
[11]	Hidden Marko Modeling	115 commonly used phrases	Julius tool works better than HTK	It hasn't explored for the fullest. Accuracy decreases with the noise
[12]	Hidden Marko Tool kit (HTK)	110 different Kannada words	Average accuracy of 98.5% for small to medium vocabulary	Performance needs to be analyzed based on the sub words, work need to be extended for large vocabulary
[13]	Convolution Neural Network	10-word sentence with 100 embedding's	Accuracy level around 75%	Although CNN is a relatively quick method, creating a large dataset is the most challenging aspect.
[14]	Deep Learning Approach	34,800 photocopies—600 photo copy for each of the 58 braille characters in Hindi.	34800 photos were included in the dataset, with a 96.47% accuracy goal.	The accuracy of deep Learning was found more than SVM.it can be future implemented in real time applications
[15]	Unified Sequence to Sequence	940000 mandarin utterances	Mean opinion score for pipeline based front end is 4.37 it is close to the human recording	It makes use of the continuous polyphones in the test dataset, and if the phoneme from the previous step was already inaccurate, it generates an incorrect phoneme.
[16]	Hidden Marko Model	Input text to Unicode	increased seven percent compared	A low sampling wavelength is employed for the

		Transformation Format (UTR-16)	to the standard method.	training procedure to speed up calculation.
[17]	Concatenate approach	Test dictionary is 4910 syllables	MOS score is 3.99	Speech signals are recorded in less noisy environment, the prosody analysis can be done through machine
[18]	Bidirectional Kalaman Filter in MATLAB	TIDIGIT database	Overall word accuracy 90%, scenario system accuracy is 80%	Kalman filter is used to remove the noise, but it takes more time to filter the noise it increases the computation time
[19]	Open Source Multi speaker speech	IIT-H,DeitY,CMU Wilderness dataset	MOS score is >3.6	The attempts to prepare for current languages will be enhanced. A small quantity of data from a language of interest is used as adaption data.
[20]	K nearest neighbor technique	3 lakh image samples	Recognition accuracy is 84.728%	The model can be further implemented for compound characters like ottaaksharas and gunitaksharas and again this can extend to the words

TABLE VI. METHOD WISE ACCURACY

Technique method	Accuracy
Method 1	65%
Method 2	88%
Method 3	18%-20.6%
Method 4	97%
Method 5	75%
Method 6	98.5%
Method 7	96.47%
Method 8	85%

TABLE V. PERFORMANCE COMPARISON OF TTS MODELS

Method Number	Technique used	Accuracy /MoS
Method 1	Syllabification Synthesis	65%
Method 2	Articulator Synthesis	88%
Method 3	Semi Supervised learning	20.6 -18.0%
Method 4	Concatenation	97%, MOS:3.99
Method 5	Artificial Neural Networks	75%
Method 6	Hidden Marko Modelling	98.5%
Method 7	Deep learning Approach	96.47%
Method 8	Unified Sequence to Sequence	MOS:4.37
Method 9	Bidirectional Kalaman Filter in MATLAB	85%

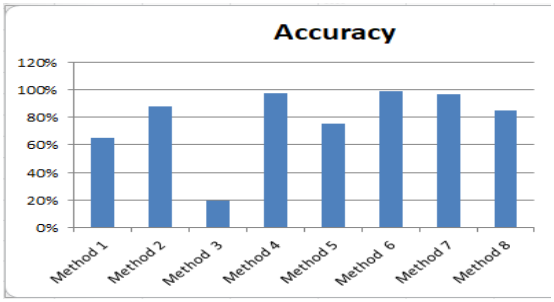


Figure 2. Accuracy of all the technique models

III. OBJECTIVES OF TTS

Whether in oral or written form, languages are the most ancient form of human interaction. There are several thousands of languages in this world The number of languages in the globe is in the thousands. Travel-related enterprises need translates to offer translation services to tourists, and in this article, we concentrate on understanding diverse users. Some central government documents, documents pertaining to land, and other documents written in English cannot be comprehended by native speakers. Since some users cannot speak, write, or understand English. So even those who are unable to read, write, or speak English can use our translator. The translator may both produce voice using TTS Synthesizer and OCR to take text from the document and convert it to the chosen language (in this case, Kannada). The speech that has been converted can be distributed to others. Colleges are developing a variety of text-based learning aids in digital format to assist kids with these language-based disabilities. They're able to have the recognized character or text read aloud to them so that students can learn a language more quickly and with more ease. TTS is frequently used in services like audio navigating, news casting, tourist translation, etc. to synthesize natural and understandable speech from audio input. An efficient TTS algorithm produces voice signals with good prosodic naturalness.

IV. SOFTWARE

A. Environment Setup

Majority of TTS systems are built by using MATLAB programming for implementing the TTS system. The model developed by the MATLAB is reliable and it is very quick while producing the speech. In this paper Kannada TTS is developed by using the Anaconda as the environment code is written in python. The developed tool is web application it is built on python module and the speech Synthesize API.TTs system implemented by python because it includes all the modules as a package. In proposed TTs is built by Google translator humans are used

to apps developed by Google like Maps, Google search by using both text and speech as an input. So, it easy for the user to use the tools built by Google, the speech synthesized by the Google API is natural and intelligent with no human errors. For launching the web application flask is used, pandas are used for getting the text data or speech data analysis, NumPy is used for computing or manipulating the text or speech before producing the output speech. The language deter is the main module in our system, in this implementation text to speech is explicitly built for Kannada and some other packages like Play sound is used to save the output from the model in the .mp3 format.

B. High Level Design

In figure 3, the high-level model of the Text to speech system is divided into 5 parts: 1. Text processing, 2. Grammar generation, 3. Speech database, 4. Concatenation of .wav, 5. Synthesized(output) speech.

1. Text Processing: The input is taken in the different forms like directly from keyboard or in the form of document in .txt format or pdf. To read the letters from the pdf is difficult, the OCR (optical character reader) is used to digitalize the text. The several steps involved in that from text recognition, segmentation and output is saved in the .txt file.
2. Grammar generation: The output from the text processing is taken in this module as an input. A whole document or sentence is read one by one word. The sentences are segmented based on the grammatical rules. Kannada follows left to right. This step is crucial step, if any incorrect from this step leads to the incorrect speech synthesis.
3. Speech database: The collection of speech database depends on how we construct the TTS model. Some models manually recorded each voice for the corresponding word, few authors uses both male and female voice for recording. Other examples where author uses the speech database from someone else to implement TTS. The most widely used method is by using the Google API and speech Synthesis API, it is cost efficient and user friendly.
4. Concatenation of .wav: When the word is read from the text document it is segmented based on the grammar. Each word is further segmented into smallest unit until it finds the corresponding .wav file from the speech database. Then the each .wav file is concatenated to in such a way that it reproduces the same sentences from the text in the speech form.
5. Synthesized speech: It is the final output from the model, it resembles like the human like speech, reading the document with intelligence.

C. Low Level Design

First, pre trained character embedding mapping transforms the unprocessed input text into a series of n-dimensional vectors n can be decided as per the user algorithm. Generated from a Word2Vector representation of the character integrated to rule. CRF (Conditional Random fields) is used to segment the sentences, A vector of K real numbers is transformed into a probability distribution with K potential outcomes by SoftMax. The main module is intended to get more information from the auxiliary module, which serves as a feature extractor to extract POS representation from raw text inputs. Phoneme, voice, and syntax sequences are the output from the auxiliary section, and these are used as language attributes for the back-end TTS.

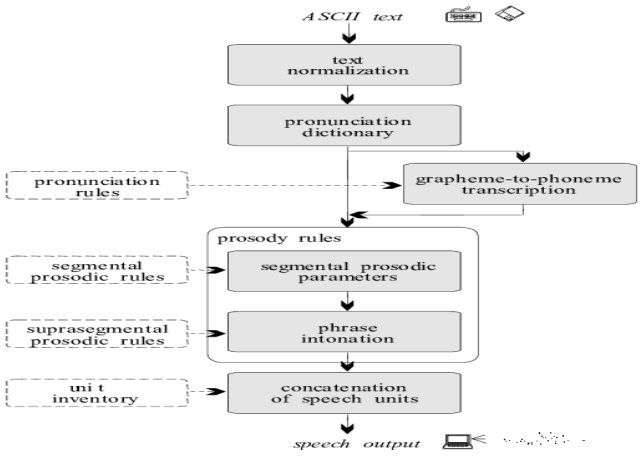


Figure 3. High level design of Kannada TTS

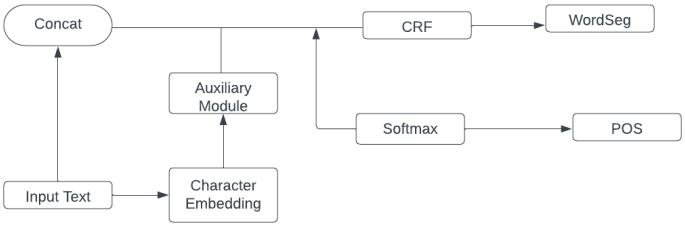
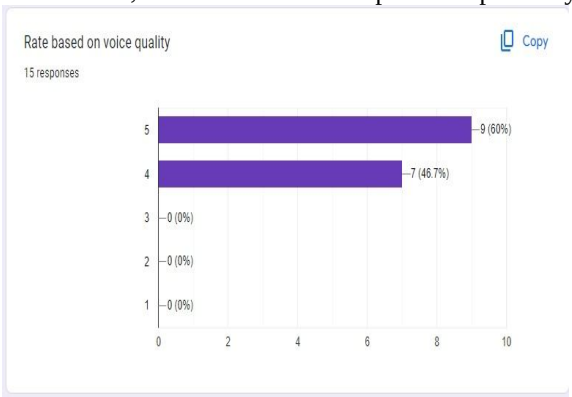
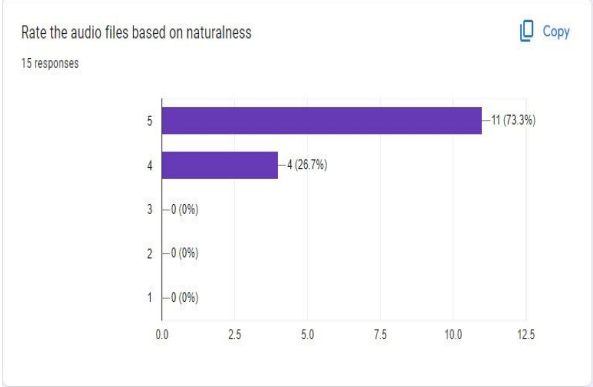


Figure 4. Low level Design of Kannada TTS

V. CONCLUSIONS

The results from the developed system are in the form of audio so the drive link is provided below to check the output for each training model:
https://drive.google.com/drive/folders/1MQTP87ipBrGS6Zmg5bq8BK4VW8nlSW0w?usp=share_link.
Here, we're attempting to provide a tool for those who only speak Kannada yet are interested in learning about documents or other information relating to land. The TTS conversion method has been very useful in increasing human independence. The literature survey is done based on the different methods and methodologies used for constructing TTS by different authors. Despite the advancement of technologies, these tools are now able to identify and convert text. The tool is developed by referring the existing models that helps the user to convert the inputted raw text data into the synthesized speech. The input given for the tool can be in different forms like image in the form PNG and document in the form of .txt or pdf. A fast-developing area of computer technology, text to speech synthesis is becoming more and more crucial to how we interact with systems and interfaces on various platforms. The many steps and procedures involved in text to speech synthesis have been recognized. The TTS systems assist users in our busy world save time. It primarily aids older individuals with failing vision and those who are visually impaired in reading Kannada People who lack literacy can depend on this framework to help them become independent, which is crucial in government organizations. Additionally, it aids those who are unfamiliar with the Kannada language. The survey is conducted for the developed model https://forms.gle/pnRP3y8dCAuUkDvA7 , feedback from the responses is plated by graph.





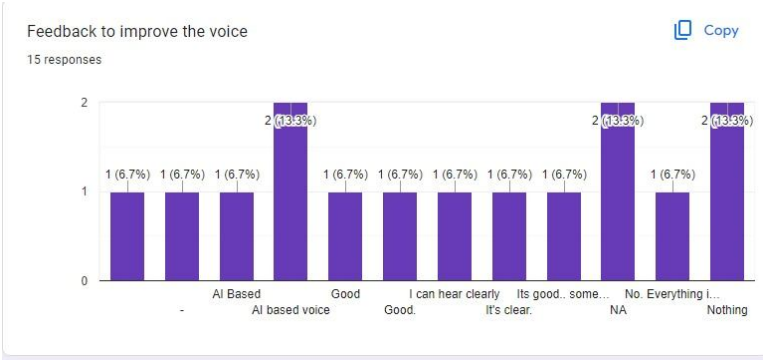


Figure 5. Survey results from developed TTS system

REFERENCES

[1] Alexandra, E., and P. Shyamala Bharathi. "Design and Implementation of Text to speech synthesizer using Syllabification synthesis algorithm and comparing with Articulator Synthesis algorithm." In 2022 International Conference on Business Analytics for Technology and Security (ICBATS), pp. 1-6. IEEE, 2022.

[2] Dhananjaya, M. S., B. Niranjana Krupa, and R. Sushma. "Kannada text to speech conversion: a novel approach." In 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), pp. 168-172. IEEE, 2016.

[3] Luo, Renqian, Xu Tan, Rui Wang, Tao Qin, Jinzhu Li, Sheng Zhao, Enhong Chen, and Tie-Yan Liu. "Lightspeech: Lightweight and fast text to speech with neural architecture search." In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5699-5703. IEEE, 2021.

[4] Karita, Shigeki, Shinji Watanabe, Tomoharu Iwata, Marc Delcroix, Atsunori Ogawa, and Tomohiro Nakatani. "Semi-supervised end-to-end speech recognition using text-to-speech and autoencoders." In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6166-6170. IEEE, 2019.

[5] Kumar, KK Anil, HR Shiva Kumar, Ramakrishnan Angarai Ganesan, and K. P. Jnanesh. "Efficient Human-Quality Kannada TTS using Transfer Learning on NVIDIA's Tacotron2." In 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp. 01-06. IEEE, 2021.

[6] Joshi, Anusha, Deepa Chabbi, M. Suman, and Suprita Kulkarni. "Text to speech system for Kannada language." In 2015 international conference on communications and signal processing (ICCSP), pp. 1901-1904. IEEE, 2015.

[7] Udayashankara, V., and Swapna Havalgi. "Speech therapy system to Kannada language." 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP). IEEE, 2016.

[8] Mothukuri, Siva Krishna P., Pradyoth Hegde, Nagaratna B. Chittaragi, and Shashidhar G. Koolagudi. "Kannada Dialect Classification using Artificial Neural Networks." In 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), pp. 1-5. IEEE, 2020.

[9] Reddy, Mallamma V., and M. Hanumanthappa. "English to Kannada/Telugu name transliteration in Clir: a statistical approach." International Journal of Machine Intelligence 3, no. 4 (2011).

[10] Jayawardhana, Pabasara, Achala Aponso, Naomi Krishnarajah, and Amila Rathnayake. "An intelligent approach of text-to-speech synthesizers for english and sinhala languages." In 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), pp. 229-234. IEEE, 2019.

[11] Sharma, Roshan S., Sri Harsha Paladugu, K. Jeeva Priya, and Deepa Gupta. "Speech recognition in kannada using htk and julius: A comparative study." In 2019 international conference on communication and signal processing (iccsp), pp. 0068-0072. IEEE, 2019.

[12] Ananthakrishna, T., M. Maithri, and Kumara Shama. "Kannada word recognition system using HTK." In 2015 annual IEEE India conference (INDICON), pp. 1-5. IEEE, 2015.

[13] Aishwarya, Jaya, Poornima Panduranga Kundapur, Sampath Kumar, and K. S. Hareesha. "Kannada speech recognition system for Aphasic people." In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1753-1756. IEEE, 2018.

[14] Kaur, Parmesh, Sahana Ramu, Sheetal Panchakshari, and Niranjana Krupa. "Conversion of Hindi Braille to speech using image and speech processing." In 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), pp. 1-6. IEEE, 2020.

[15] Pan, Junjie, Xiang Yin, Zhiling Zhang, Shichao Liu, Yang Zhang, Zejun Ma, and Yuxuan Wang. "A unified sequence-to-sequence front-end model for mandarin text-to-speech synthesis." In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6689-6693. IEEE, 2020.

[16] Jayakumari, J., and A. Femina Jalin. "An improved text to speech technique for tamil language using hidden Markov model." In 2019 7th International Conference on Smart Computing & Communications (ICSCC), pp. 1-5. IEEE, 2019.

[17] Ramteke, G. D., and R. J. Ramteke. "Text-To-Speech Synthesizer for English, Hindi and Marathi Spoken Signals‖." at British Journal of Applied Science & Technology (2016): 2231-0843.

[18] Sharma, Neha, and Shipra Sardana. "A real time speech to text conversion system using bidirectional Kalman filter in Matlab." In 2016 International conference on advances in computing, communications and informatics (ICACCI), pp. 2353-2357. IEEE, 2016.
[19] He, Fei, Shan Hui Cathy Chu, Oddur Kjartansson, Clara E. Rivera, Anna Katanova, Alexander Gutkin, Isin Demirsahin et al. "Open-source multi-speaker speech corpora for building gujarati, kannada, malayalam, marathi, tamil and telugu speech synthesis systems." (2020).
[20] Sneha, U. B., A. Soundarya, K. Srilalitha, Chandravva Hebbi, and H. R. Mamatha. "Image to Speech Converter–A Case Study on Handwritten Kannada Characters." In 2018 international conference on advances in computing, communications and informatics (ICACCI), pp. 881-887. IEEE, 2018.