International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

A Novel Text to Speech Technique for Tamil Language using Hidden Markov Models (HMM)

A. Femina Jalin, J. Jaya Kumari

Abstract: ***Application of digital signal processing in speech processing plays a major part in our everyday life. Text to speech system lets people to see and read out loud consecutively. Text-to-speech synthesizers use synthesis techniques that require good quality speech. Text to speech conversion (TTS) can apply to many applications such as automation, audio recording and audio-based assistance system. Text to speech conversion can be applied for various multinational language as well as for a number of local languages. An efficient text to speech conversion for Tamil language with extreme accuracy is proposed in this work. Multi feature, with a Hidden Markov Model (HMM) predictor is used to convert text to speech efficiently. By using the proposed method, the precision of the framework is enhanced by a factor of 6% when contrasted with the traditional system.***

Index Terms: Digital Signal Processing (DSP), Hidden Markov Model (HMM), Mean square error (MSE), Tamil Unicode, Text to speech conversion, (TTS).

I. INTRODUCTION

Text to speech conversion uses sequence of text as input and generate speech signal as output. It is generally utilized in speech generated devices for blind individuals. Any text to speech framework incorporates of two noteworthy components. Beginning with the output, it requires some sort of sound-generating component whose capacity is closely resembling to human vocal tract. Efficient text to speech algorithm generates speech signal with natural prosodic with good speaking style. A module whose info is the content or other semantic data to be talked and whose yield drives the sound-producing instrument. In present day innovation, both of these parts are programming. It can be executed so that they can keep running on numerous sorts of equipment stages [1].

Manuscript published on 30 August 2019.
*Correspondence Author(s)
A. Femina Jalin, Student, Department of Electronics and Communication Engineering, Noorul Islam Centre for Higher Education, India
J. Jaya Kumari, Professor, Department of Electronics and Communication Engineering, Mar Baselios College of Engineering and Technology, (Kerala), India.

The initial phase in a text to speech transformation framework changes over such data into a standard content arrangement. The initial phase in a text-to-speech conversion framework changes over such data into standard text format. Every language with an alphabetic composition has specific procedures to convert spelling into sound. The most fundamental attributes of the speech combination system remain instinctual nature and coherence. Naturalness is how intently the yield produces sound like human speech [2], [3], [4], however coherence is simplicity with that the yield remains comprehended. Perfect speech synthesizer is together characteristic and coherent. Speech synthesis frameworks for the most part attempt to expand the two qualities. The two important latest methods to generate man-made speech waveforms are concatenative speech synthesis and formant speech synthesis. Concatenative speech synthesis relies upon the connection of parts of the recorded speech. It conveys the bettert widely recognized regular natural-sounding synthesized speech. The contrasts between characteristic varieties in speech and the idea of the computerized strategies for sectioning the waveforms once in a while result in capable of being heard bug in the yield. There are three fundamental sub-types of concatenation union. Unit selection, Diphone synthesis, and Domain-specific synthesis are the fundamental sub-sorts of concatenative union. Numerous frameworks in light of formant blend innovation produce counterfeit, mechanical sounding speech that could never be mixed up for human speech. In any case, most extreme expectation isn't generally the objective of a speech combination framework, and formant blend frameworks have focal points over concatenative frameworks. Formant-mixed speech can be constantly coherent, even at high speeds, avoiding the acoustic glitches that customarily torment concatenative systems. Numerous frameworks in view of formant union innovation create counterfeit, automated sounding speech that could never be mixed up for human speech. Be that as it may, greatest expectation isn't generally the objective of a discourse union framework, and formant union frameworks have focal points over concatenative frameworks. In any case, greatest expectation isn't generally the objective of a speech combination framework, and formant synthesis framework have points of interest over concatenative frameworks.

Formant-combined speech can be dependably clear, even at high speeds, keeping away from the acoustic glitches that ordinarily torment concatenative frameworks [8]. Fast incorporated speech is utilized by the outwardly weakened to rapidly explore PCs utilizing a screen peruser. Formant synthesizers are normally littler projects than concatenative frameworks since they don't have a database of speech tests. They can along these lines be utilized in installed frameworks, where memory and chip control are particularly restricted. Since formant-based frameworks have finish control of all parts of the yield speech, a wide assortment of prosodies and pitches can be yield, passing on not simply inquiries and proclamations, but rather an assortment of feelings and manners of speaking. The better quality speech signal could be obtained by the use of unit selection synthesis and Harmonic plus Noise (HNM). HNM models are parametric models which serves to alter prosodic highlights. Here, the speech flag is achieved in the process of quasiperiodic components called clamor. HMM is one of the accurate parametric blends that comprise of dualistic stages; First is to train, and second one is to synthetic speech. In the Training stage, features are extracted, and retained in a component vector. In synthesis stage, highlight vectors are changed through to sound flag with the assistance of channel. The incorporation of HNM and HMM has the benefit of diminish advancement duration and expense [9]. Proposed method uses Hidden Markov model to convert the text to speech in an efficient manner. Hidden Markov model is pre-owned as a predictor to correct the pronunciation speed to get better quality. The accuracy of the system is improved well when compared to the conventional techniques.

II. RELATED WORK

A boundless text-to-speech system depended upon by means of make a speech signal, identifying with the given substance in a vernacular, is entirely reasonable to human group of spectators. Straightforwardly, unit selection-based synthesis (USS) and quantifiable parametric mix methodology are utilized in this work. A connection synthesizer was made for the dialect Tamil, using 12 hours of speech data, and exhibited that the syllable is the better sub word unit. HMM based voice developed utilizing a comparative proportion of discourse data defeats Fest Vox established voice. The intention is on the grounds that there is no sonic-glitches present in the speech synthesized. Another favored point of view here is the circumstance that the impression gauge is basically tinier when appeared differently in relation to any of the Fest Vox-based voice. In any case, the weights are, (I) buzziness in the orchestrated speech and (ii) nonappearance of speaker character. [10]. This paper uses inflection model, which is one of the fundamental prosodic parameters to accomplish the quality with respect to desire for the discourse in Tamil TTS. The enthusiastic evaluation of the proposed procedure exhibits the basic change in the quality to the extent desire for the conveyed speech. This technique is executed with assistance of neural system to change over content to discourse signal. Normally, positional, legitimate, phonological and articulatory features are used to set up the structure. The speech generation framework must be totally analyzed and associated with remove articulatory features. It is a dreary system, which requires progressively manual effort. A high dimensional component vector is similarly another essential. The issue is understood by using the articulation limits as a component close by positional, applicable and phonological features to set up the structure at the principal arrange. The central favored angle of the proposed presentation is, the annihilation of creation constraints for feature extraction. [11] In this work about the affiliation effort on structure building text to speech (TTS) systems for 13 Indian vernaculars. As Indian dialects are syllable-arranged, a syllable-based framework is delivered. As nature of speech union is of premier interest, unit-decision synthesizers are amassed. Building TTS systems for low-resource dialects requires that the data be unequivocally assembled a clarified as the database must be worked from the scratch. Various criteria need to tended to while building the database, to be explicit, speaker decision, rhetoric assortment, perfect substance assurance, treatment of out of vocabulary words and so forth. The various properties of the voice that impact talk association quality are first dismembered. Next the layout of the corpus of all of the Indian dialects is composed. The assembled data is set apart at the syllable level using a self-loader naming instrument.
The headway of syllable-based TTS es transversely more than 13 Indian dialects certifies the noteworthiness of syllable as a key unit for mix. As syllable encompasses co-verbalization, the prosody change required was on a very basic level less. It was similarly observed in the midst of substance parsing that a vital subset of the standards was typical over each and every Indian dialects. Whimsies were managed as exclusions. Given the normal name set it should possible to gather synthesizers for new dialects with by no effort [12]. The comparable to, data are utilized to create statistical models of HMM-based content to-speech (HMM-TTS). It is possible to get data from a wide scope of sources yet going along with them prompts a non-homogeneous or different dataset. This paper portrays the use of average voice models (AVMs) and a novel utilization of cluster adaptive training (CAT) with different setting subordinate decision trees to make HMM-TTS voices using arranged data: discourse data recorded in studios mixed with discourse data got from the web. Getting ready AVM and CAT models on various data yields ideal quality discourse over planning on eminent studio data alone. Tests moreover exhibit that CAT produces higher quality voices than AVMs paying little heed to the proportion of change data. All in all, it is exhibited that it is valuable to demonstrate the data using various setting gathering decision trees. [12]. A viable cross-language transformation structure for creating prosodic models for Chinese vernacular text- to-speech system is proposed. In this structure,Chinese prosodic models are balanced from a present Mandarin talking rate subordinate different leveled prosodic model. The reason of the framework relies upon the cross-local resemblances among Mandarin and other Chinese language to the extent syntactic and prosodic structures. Two key issues are tended to in this examination: One issue identifies with the use of cross-local resemblances in the arrangement and change of the vernacular talking rate-subordinate different leveled prosodic model.

Retrieval Number: I8589078919/19©BEIESP
DOI: 10.35940/ijitee.I8589.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The other issue identifies with the data pitiful condition brought about by the deficiency of a modification corpus covering crucial semantic settings and prosodic events and furthermore a wide talking rate run. This issue is enlightened by using the fundamental most noteworthy a posteriori methodology that logically deals with the vernacular talking rate-subordinate different leveled prosodic model parameters into decision trees to empower parameter estimations. The reasonability of the proposed methodology was surveyed by tests on two Chinese vernacular: Min and Hakka. Target and unique evaluations demonstrated that the prosodic features made by the tongue talking rate-subordinate dynamic prosodic models were extremely ordinary in various talking rates stretching out from 3.3 to 6.7 syllables for consistently. These results certify that the proposed cross-vernacular modification framework is ground-breaking and promising. [14]. In this paper, a structural maximum a posteriori (SMAP) speaker alteration approach to manage adjusting the speaking rate (SR)- subordinate different leveled prosodic model (SR-HPM) of a present SR-controlled Mandarin text-to speech framework to another, speaker's data for making another voice is inspected. One is the little SR consideration of the modification data and is handled by using the present SR-HPM that was set up from a speech corpus of wide SR incorporation as a helpful prior. Another is the data insufficiency issue coming about due to the far reaching number of parameters of the SR-HPM to be adjusted. It is appreciated by dynamically dealing with the SR-HPM parameters into decision trees so as to be profitably adjusted by the SMAP procedure. The viability of the proposed methodology is surveyed on speech databases of five new speakers. Both objective and theoretical appraisals show that the proposed system not simply performs better than anything the best likelihood based strategy in the watched SR extent of the goal speaker's HPMs for different talking styles. Talking style ID for the target speaker is moreover required. Another is the improvement of an online learning variation of the proposed SMAP-based speaker alteration method. This may be fundamentally useful for people to recognize singular TTS structures. The use of on the web or group getting ready estimations will be considered. [15].

III. PROPOSED SYSTEM

Proposed framework has two fundamental sections, that are training and another testing as shown in Fig 1. The initial phase in this suggested framework is to change over given Tamil content to relating Unicode. Unicode is a figuring industry standard for encoding, representation, and handling of text expressed in the greater part of the world's written work frameworks. Fig 2. Shows the Tamil unicode chart which is used in the world wide. Unicode can be executed by various character encodings. The Unicode standard characterizes UTF-8, UTF-16, and UTF-32, and a few different encodings are being used. Each unique code has corresponding voice samples [6],[10],[11]. All Unicode of Tamil language and their corresponding voice signals are stored in the database. From the given Unicode corresponding voice sample is selected from the database based on some conditions which are explained by below flow chart fig 4. In the training phase the system is trained by using frequency feature which is extracted after collecting speech sample from the database. In testing phase, waveform will be generated based on the Unicode which is obtained from the user given input. Brief explanation about proposed method is explained below.

A. Block Diagram

The Block graph for the proposed framework is appeared in Fig 1. It is splitted into two major sections. They are testing and training. Training portion consist of data collection feature extraction and learning. Testing process is carried out by performing text processing and feature extraction. Testing operation is performed individually on trained HMM model.

B. Tamil Unicode

Tamil Unicode is the representation of Tamil characters. Unicode sample and their corresponding waveform is information, yet in addition is vastly improved in the inconspicuous SR ranges. A multi-speaker prosody exhibiting study will be coordinated to create different SR-Represented in figure 2. In figure 2, First segment indicates the Unicode. Second column specifies the corresponding Tamil characters. Third and fourth columns specify the waveform and frequency responses. Normally the characters in the document will be stored in hex value for all the letters in the Tamil language [1]. Tamil Unicode starts from 0BB0 and ends at 0BFF. The frequency response for some words and their corresponding waveform can be shown in figure 6.

C. Segment Characters

Instead of storing the waveform individually, the corresponding waveform of the characters should be stored with all combination of characters. Based on the amplitude and nearby local information, we can split the characters or pronunciation individually. This process is also called as silence detection. When corresponding combination of Unicode is getting as input, the algorithm will be used to segregate the character set.

D. Feature Extraction

Feature extraction is one of the major steps in text to speech conversion. By using primary step which is explained in the first section can appropriately create the waveform

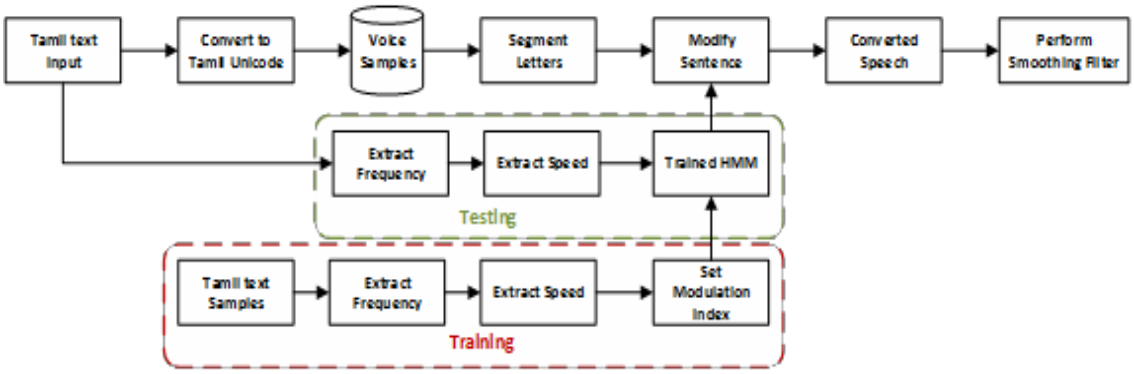


Fig 1. Block Diagram of Proposed Method

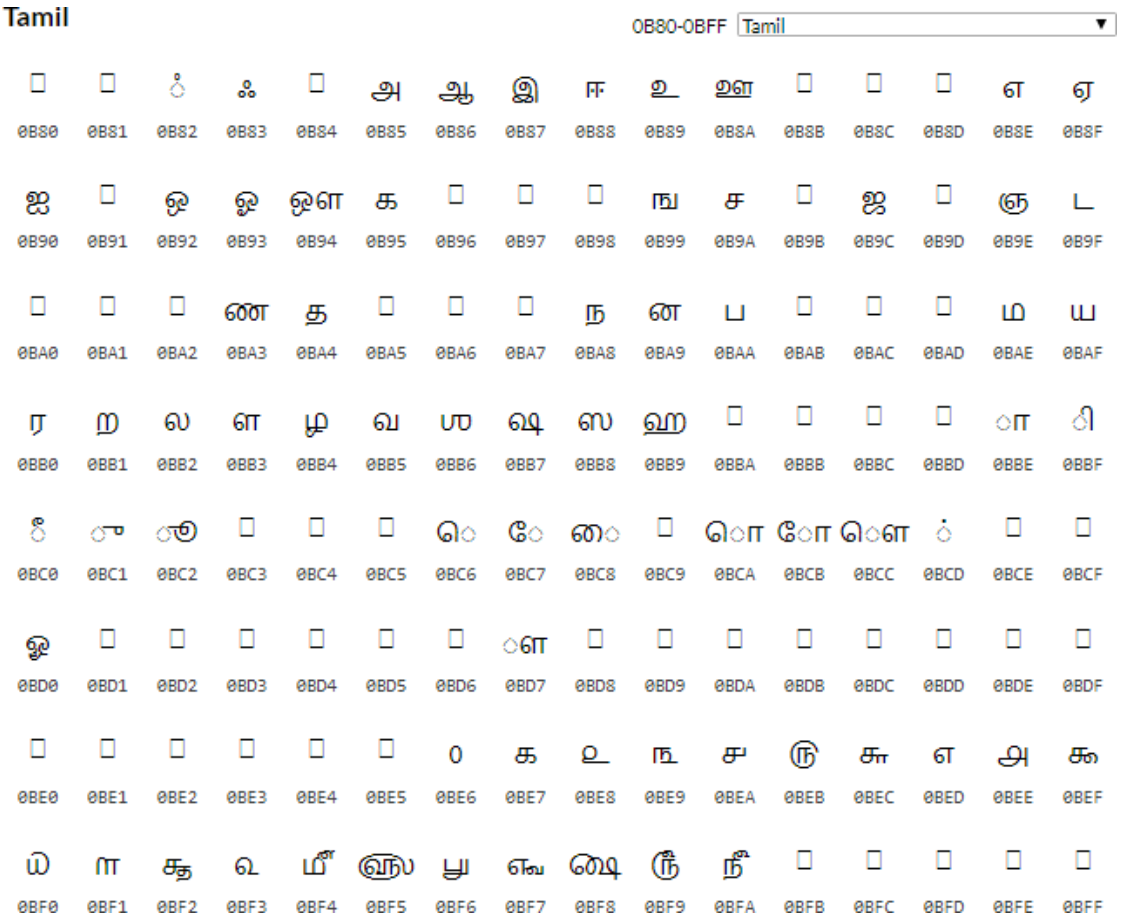


Fig 2. Tamil Unicode table for all Characters

The quality of the system should be further improved by using hidden Markov model which can predict corresponding speed and pitch for the roughly generated waveform. Majorly frequency and speed of the corresponding waveform is estimated and new speed will be used to modify the sentences. The frequency response for some words and their corresponding waveform can be shown in figure 5.Figure 3 shows the tamil uincode table and corresponding speech waveform.

Retrieval Number: I8589078919/19©BEIESP
DOI: 10.35940/ijitee.I8589.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

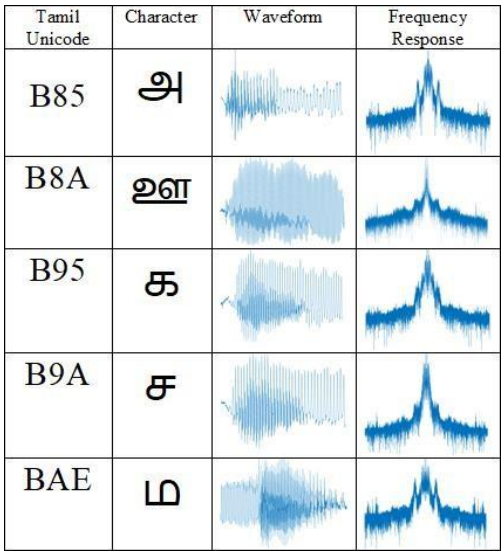


Fig 3. Tamil Unicode table and corresponding speech waveform

E.Hidden Markov Model (HMM)

The Hidden Markov Model (HMM) is generally a straightforward method to demonstrate successive information. Markov model is a stochastic model used to show haphazardly framework.Broadly utilized in temporal pattern recognition. For illustration; speech, handwriting, gesture recognition, robotics, biological sequences and lately in energy disaggregation. There are two factors in HMM. They are observed variables and hidden variables. variables are utilized to display states (ON, OFF, reserve and so forth).

1.The finite set of hidden states SS (e.g. ON, stand-by, OFF, etc.) of an appliance, S = {s1, s2....,sN}.

2.The finite set of MM observable symbol YY per states (power consumption) observed in each state, Y = {y1, y2,....,yM},Y={y1,y2,....,yM}. The observable symbol Y can be discrete or a continuous set.

3.The emission matrixB={bj(k)},B={bj(k)} representing the probability of emission of symbol k ϵ Y when system state is st==j such that: bj(k) = p(yt = k |st=j)bj(k)=p(yt=k|st=j)

4.And the initial state probability distribution {π}={πi} indicating the probability of each state of the hidden variable at t=1 such that, πi=P(q1=si),1≤i≥N.

The matrix B is an NxM. The emission probability can be a discrete distribution or a continuous distribution. In case of discrete emission, multinomial distribution is used and for continuous emission, multivariate Gaussian distribution is used.

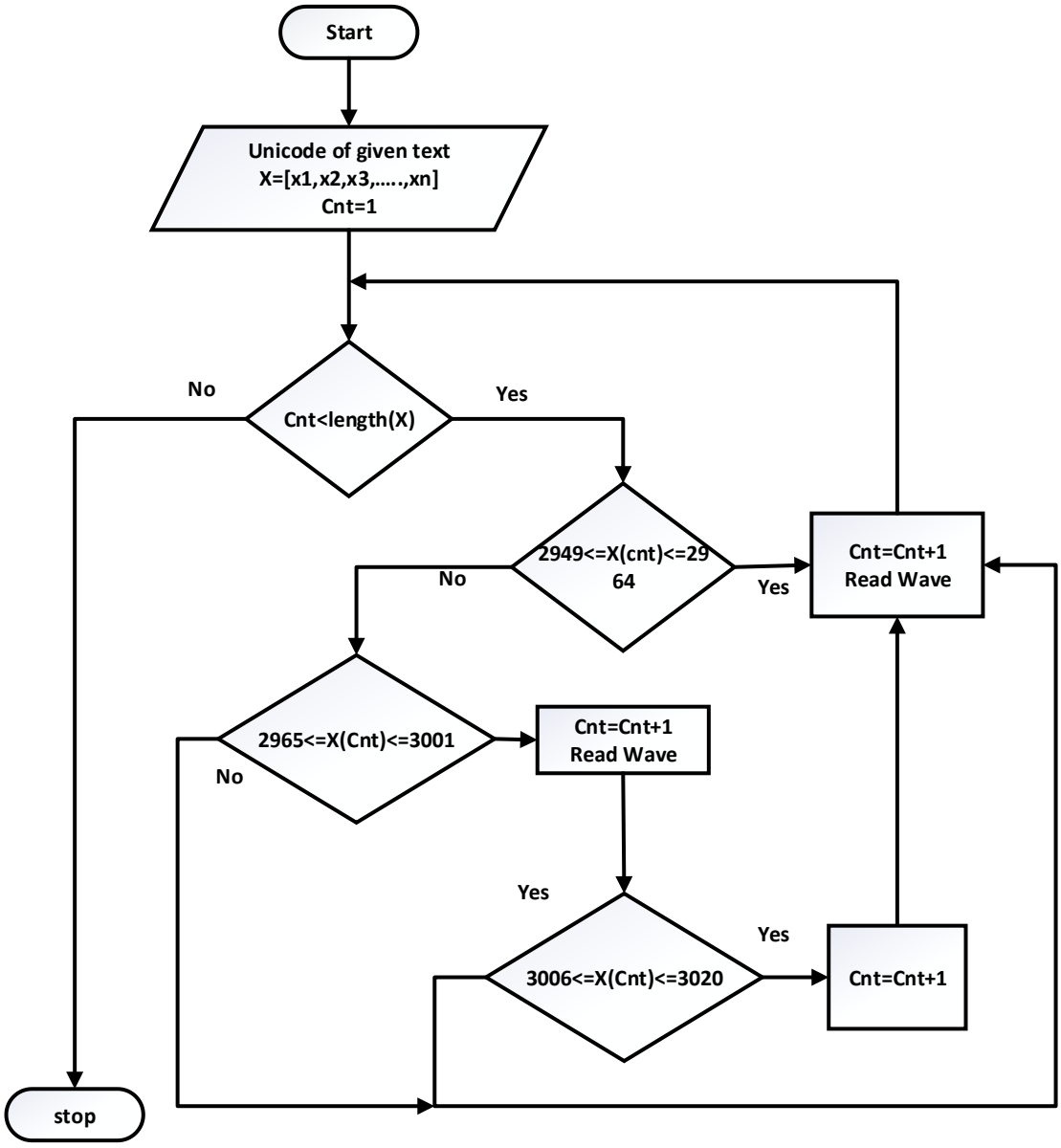


Fig 4. Flowchart for Extracting voice sample using Unicode

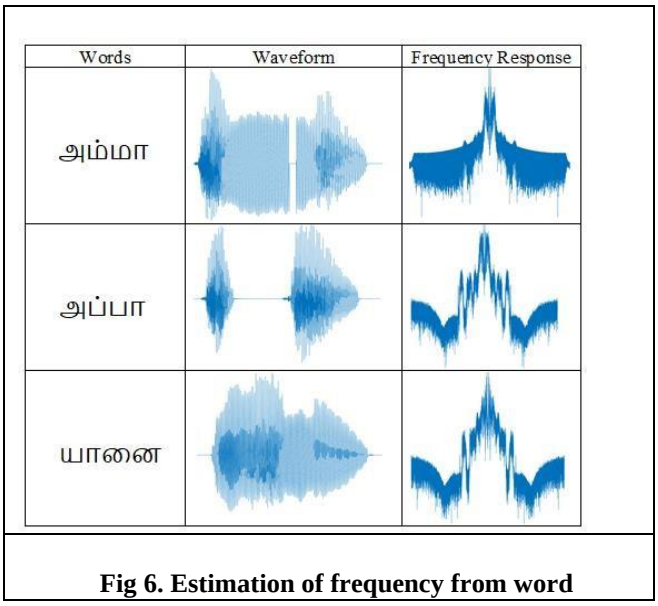


Fig 6. Estimation of frequency from word

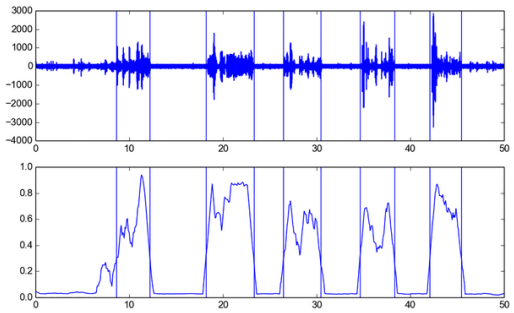


Fig 5: Segmenting characters from word which is stored in database

F. System Training

To train HMM, different combinations of randomly selected word's waveforms are collected and their corresponding speed and frequency are estimated.

The actual value of speed is set to perform training operation. The accuracy of system is directly proportional to the quantity of data samples which is used as training. More number of samples may increase the overall computational time for training. Similarly, the accuracy of the system is depending on the other factor that is how accurately setting the target while training.

Table I: Sample words used to perform testing operation

Sl. No	Words
1	அம்மா
2	அப்பா
3	யானை
4	பூனை
5	ஆனம
6	உளி

G. System Testing

Testing is performed by the same procedure which is used for training. By using feature extraction which is used for training only using for the testing operation also. But testing will be normally performed by single input single output manner.

IV RESULT AND DISCUSSION

Normally, training is performed for bulk number of sample waveform for accurate training. A MATLAB based graphical user interface (GUI) is used to perform testing operation. The performance of the system is measured by performing accuracy assessment which can measure the quality of the proposed system.Figure 7 shows the segmenting characters from word.

A. Data Set

Some of the sample words which is used in this work can be shown in Table I. for testing and training voice samples are recorded by using microphone with 44000Hz sampling. Voice samples are recorded in .wav format and stored in corresponding Unicode as filename. It is easy to retrieve the voice sample if saving in Unicode name.

B.Result Analysis

Table 2 shows the performance measures for different words with spectrum calculation. Column 1 shows the waveform which is Generated by using proposed synthesizer and column 2 is corresponding spectrum estimate. Column 3 of table II shows the actual word which is pronounced by a female. By this technique, voice samples can be directly retrieved from database.4th columns shows the spectrum estimate of the human pronounced word. Next two columns show the difference between the synthesized word and pronounced word and their corresponding spectrum differences. Figure 8: and figure 9 shows the graphical representation of MSE and MND.

C. Accuracy Assessment

The authorization of the proposed strategy can be assessed by

The proposed algorithm is simulated in the MATLAB2014a version in an I5 system with 4GB RAM. Real-time voice samples are recorded by using microphone and stored into the database. Similarly, to train HMM, a greater number of words samples are recorded and separately stored to perform training operation. To reduce the computational time, a lower sampled waveform is used to perform training operation.

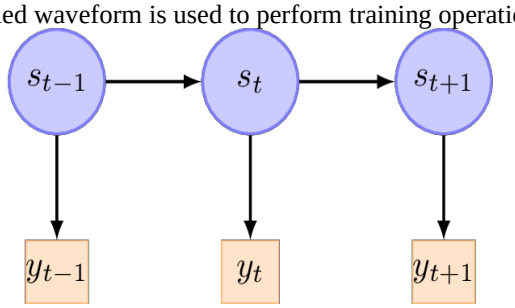


Fig 7. Segmenting characters from word which is stored in database

estimating the nature of the orchestrated Speech signal.. Mahala Nobis distance and Mean Square Error are two parameters used to measure the nature of the synthesized speech signal . The Mahala Nobis distance is greater than one and mean square error is less than one for good quality speech signal. This way we can evaluate the performance of the proposed system.

a) Mahala Nobis Distance

Mahala Nobis separation computes the separation between two examples dependent on their mean element vectors μA and μB, and the covariance matrix Σ of the features across all samples in the database.The distance is given as[D_M (μ〗 _A,μ_B)= 〖〖(μ〗 _A-μ_B)〗 ^T Σ^(-1) 〖(μ〗 _A-μ_B) (1)

On the off chance that the circulation of highlight vectors of all perceptions is ellipsoidal, at that point the Mahala Nobis remove between two mean vectors in include space is subject to the separation along each component measurement yet in addition on the change of that element measurement. This property makes the Mahala Nobis separate free of the size of the highlights. In administered order of music, Mandel and Ellis utilized an adaptation of Mahala Nobis remove, where the mean vector comprised of the considerable number of passages of the example astute mean vector and covariance framework.

This can be concluded from the value of MSE and MND. Both values should be minimum for better performance.Figure 10 shows graphical user interface for proposed method.

b) Mean Square Error

In bits of knowledge, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator gauges the normal of the squares of the errors that is, the typical squared

complexity between the assessed regards and what is surveyed. MSE is a risk work, identifying with

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 2 Performance measures for different words

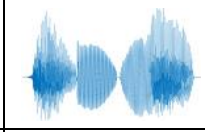
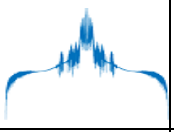
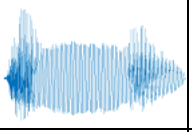
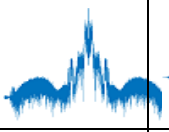
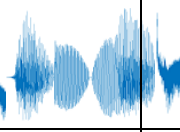
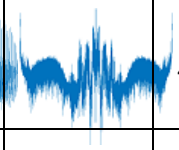
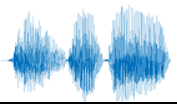
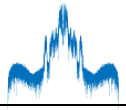
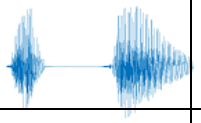
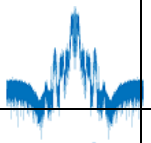
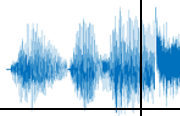
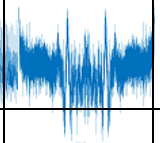
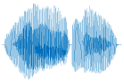
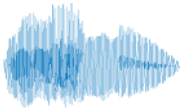
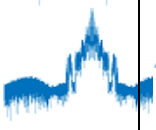
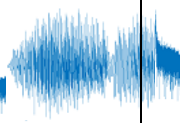
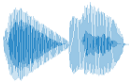
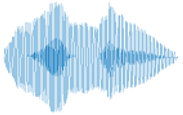
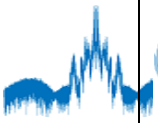
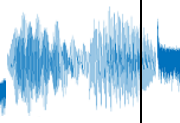
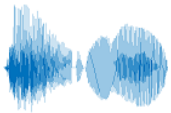
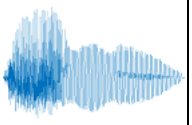
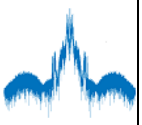
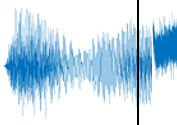
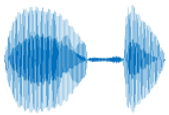
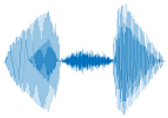
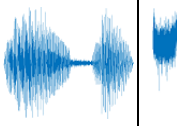
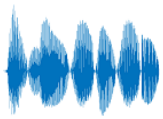
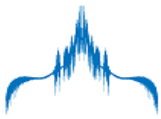
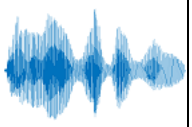
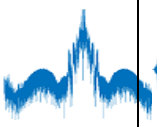
word	Synthesized		Pronounced		Difference		Performance	
	Waveform	Frequency Response	Waveform	Frequency Response	Waveform	Frequency Response	MND	MSE
அம்மா							4.2572	0.09 9378
அப்பா							9.6607	0.09 6659
யானை							7.0263	0.12 146
பூனை							5.6527	0.11 034
ஆமை							4.6882	0.10 411
ஊசி							5.3544	0.07 9278
அலை வடிவம்							4.5808	0.08 7698

Fig 8. Mahala Nobis distance for different words

Fig 9. Mean square error MSE for different words

Retrieval Number: I8589078919/19©BEIESP
DOI: 10.35940/ijitee.I8589.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

"Study on the consistency analysis between the prosody and the spectrum for Mandarin speech", Signal Processing IET, vol. 7, no. 2, pp.

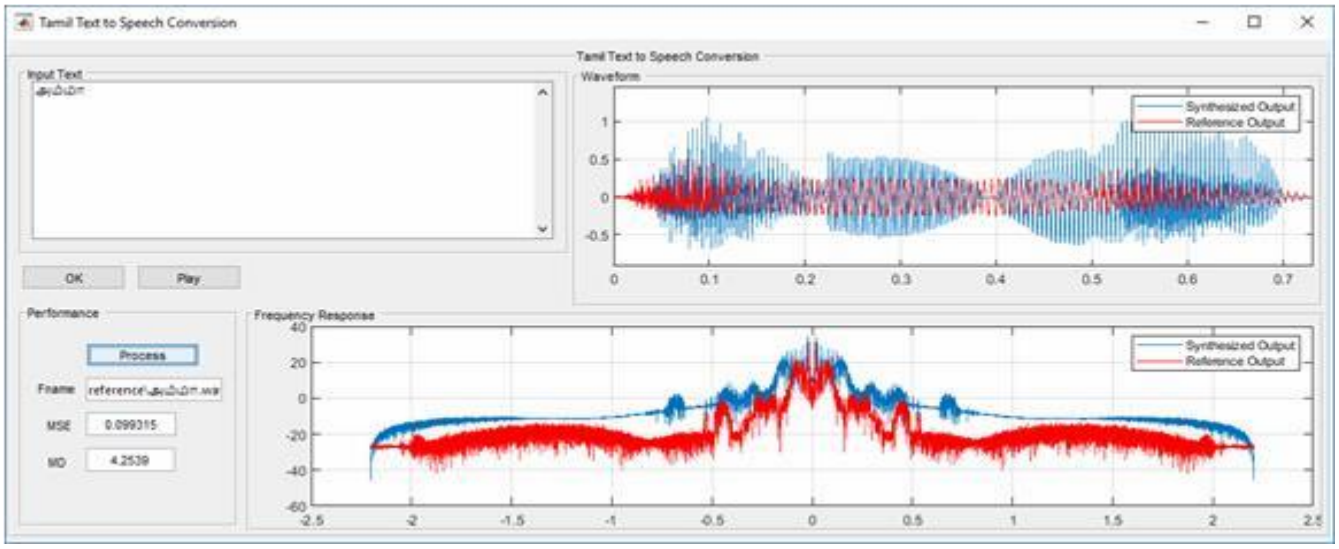


Fig 10: Graphical User interface for proposed method

Table III. Performance Comparison with GMM method

Words	GMM		HMM	
	MND	MSE	MND	MSE
அம்மா	4.75	0.10021	4.257	0.09937
அப்பா	9.9947	0.099988	9.6607	0.09665
யானை	7.14856	0.13121	7.0263	0.12146
பூனை	5.9378	0.12241	5.6527	0.11034
ஆமை	5.0013	0.12314	4.6882	0.10411

the normal estimation of the squared mistake misfortune. The way that MSE is quite often entirely positive (and not zero) is a result of arbitrariness or in light of the fact that the estimator does not represent data that could create a more precise measure.

V CONCLUSION

Highly accurate Tamil text to speech conversion using HMM is implemented and validated in this work. Also, some speech parameters are also discussed. By analyzing the result, it is clear that from the proposed system highly accurate speech waveforms are generated when compared to the conventional techniques. The proposed system can work all maximum available Tamil words without any degradation or miss pronouncement. Accuracy of the proposed algorithm is improved 7% more when compared to the conventional technique. Due to less computational complexity, the proposed system can be use easily with less computational complexity for real-time hardware implementation of the products

REFERENCES

1. Cheng-yu Yeh, Kuan-lin Chen, Shaw-hwa Hwang, Long-jhe Yan, 158-165, 2013.
2. Jalin, A. Femina, and J. Jayakumari. "Text to speech synthesis system for tamil using HMM." In Circuits and Systems (ICCS), *2017 IEEE International Conference on, pp*. 447-451. IEEE, 2017.
3. Arun Soman, S. Sachin Kumar, V. K. Hemanth, M. Sabarimalai Manikandan, and K. P. Soman, "Corpus driven Malayalam text to speech synthesis for interactive voice response system" International journal of computer application, vol-29, no.4, September 2011.
4. Dartmouth College: Music and Computers Archived 2011-06-08 at the Wayback Machine., 1993.
5. Boothalingam, R., Solomi, V. S., Gladston, A. R., Christina, S. L., Vijayalakshmi, P., Thangavelu, N., & Murthy, H. A. (2013, February). Development and evaluation of unit selection and HMM-based speech synthesis systems for Tamil. In Communications (NCC), National Conference on (pp. 1-5). IEEE, 2013
6. Priyanka Jose , Govindaru V " Malayalam Text-to-Speech" IJETR ISSN: 2321-0869, Volume-1, Issue-3, May 2013.
7. Rajeswari, K. C., & UmaMaheswari, P. (2014, December). A novel intonation model to improve the quality of tamil text-to-speech synthesis system. In Advanced Computing (ICoAC), Sixth International Conference on (pp. 335-340). IEEE, 2014
8. Patil, H. A., Patel, T. B., Shah, N. J., Sailor, H. B., Krishnan, R., Kasthuri, G. R., ... & Kishore, S. P. (2013, November). A syllable-based framework for unit selection synthesis in 13 Indian languages. In Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), International Conference (pp. 1-8). IEEE, 2013
9. Wan, V., Latorre, J., Yanagisawa, K., Braunschweiler, N., Chen, L., Gales, M. J., & Akamine, M. Building HMM-TTS voices on diverse data. IEEE Journal of Selected Topics in Signal Processing, 8(2), 296-306, 2014.
10. Chiang, C. Y, Cross-Dialect Adaptation Framework for Constructing Prosodic Models for Chinese Dialect Text-to-Speech Systems. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(1), 108-121, 2018
11. Liao, I. B., Chiang, C. Y., Wang, Y. R., & Chen, S. H. Speaker adaptation of SR-HPM for speaking rate-controlled Mandarin TTS. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(11), 2046-2058, 2016
12. Gopinath, Deepa P., et al. "A hybrid duration model using CART and HMM." TENCON 2008-2008 IEEE Region 10 Conference. IEEE, 2008

AUTHORS PROFILE

A. Femina Jalin is a fourth year PhD student in Electronics and Communication Engineering Department of Noorul Islam Centre for Higher Education, India. Her main research interest is how Tamil speech is produced from the Tamil input text. She received BE degree in ectronics and Communication Engineering and ME degree in Applied Electronics from Anna University. Her research interest is in Speech Processing, image processing and neural networks.

J. Jaya Kumari is Professor in Electronics and Communication Engineering Department of Mar Baselios College of Engineering and Technology, Kerala, India. She received BE degree in Electronics and Communication Engineering from MS University; Master degree in Applied Electronics and Instrumentation from Kerala University and PhD from Kerala University in 2009. She has teaching experience of 22 years, research experience of 14 years and administrative experience of 16 years. She had published several papers in international journals. Her research interest includes wireless communication and networking, signal and image processing, detection & estimation theory, spread spectrum systems and error correcting codes. She is a Fellow of Institution of Electronics and Telecommunication Engineers (IETE), Institution of Engineers (India) and council of Engineering and Technology also life member of Indian Society for Technical Education (ISTE), and Senior Member of Institution of Electrical and Electronics Engineers (IEEE).

Retrieval Number: I8589078919/19©BEIESP
DOI: 10.35940/ijitee.I8589.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)