International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Gender Classification for Emotional Speech using GMFCC and Deep LSTM

Sandeep Kumar, Jainath Yadav

Abstract—We have come to the point that one of the important aspects of the process speech emotion recognition is the gender classification. The correct classification of gender will improve the performance of Speech Emotion Recognition (SER) system towards its robustness. Here, we are specifically referring to Gammatone Mel Frequency Cepstral Coefficient (GMFCC) as a feature extraction method that extracts features from IITKGP-SESHC dataset, which is very crucial in deciding either male or female in gender classification. The well known classifier "Deep Long Short Term Memory (Deep LSTM)" is itself an important kind of Recurrent Neural Network (RNN) that handles the long-range dependencies more efficiently than the RNNs. The GMFCC feature has to pass through the Deep LSTM and get average percent gender identification accuracy of 98.3%.

Index Terms—GMFCC, Deep LSTM, Gammatone Filter, ERB

I. INTRODUCTION

THE speech signal has been considered as the best naturally understood mode of communication between people since the time being. Thus, the researches declared it as the speedy and methodical way of communication among people as well as machines. But, there is a condition over here that the machine should be capable enough to identify the human voices. As we have seen, during the last few years, there has been consistent research on speech recognition, which conveys the idea of converting human speech into a particular sequence of the word. Even though many research works have focused on speech recognition, we have been unable to establish the connection between man and machine as the reason being that the machine is not able to comprehend the emotional state of speaker [4]. This has given way to formation of the research name, speech emotion recognition. It is termed as a process of extracting some useful semantics from speech and thereby improving the performance of the speech recognition system [15].

With the technological advancement, the demands for emotion recognition in new speech dialogue system are very urgent as it has been advantageous in real life application. Citing the example of a health care field, the intelligent robot, which monitors the patient's emotion states, can help the doctor in diagnosing the mental and physical illness of a patient within a short period. Also, in an intelligent vehicle, the emotion recognition system monitors the drivers' emotion variation while during to escape vital accidents or mishaps. It can also be installed in human-computer interaction (HCI) based entertainment system [2]. The emotion recognition from the speech is flaming the field of research in speech recognition. So, to design a vigorous speech emotion recognition system, we need to discriminate between male and female speech as we are aware that the female speaker pitch is very high concerning male speaker pitch. The pitch is considered as a basic and most indispensable feature in determining the classification of gender in the field of speech emotion recognition. Gender identification is a process by which we try to identify the gender of a person with the help of emotional speech characteristics. Here, the gender-susceptible model is more correct than the gender-free model. All the previous research works have been carried out using a pitch threshold based on gender classification. But in this paper, we have proposed a new feature GMFCC that classifies gender with a Deep LSTM classifier.

II. PREVIOUS WORK

Rakesh et al. [17] have suggested two distinct models. Both have their own applications. The first model is useful in producing the format values of voice samples, whereas the other model aims at providing the pitch value of the voice sample. These two models help at identifying the genderbiased features and in extracting the pitch value of a speaker. The mean of formants and pitch of all the speakers have calculated by applying this model as it has got loops and counters. The model has generated the mean values of the formant and the pitch value of the speaker. With the help of euclidean distance from the mean value, the speakers have been classified between male and female. Meena et al. have demonstrated threshold-based gender classification from speech. Here, the authors use two methods, namely fuzzy Logic and neural network, to find out two thresholds from features extracted from speech, namely Short-Time Energy (STE), Zero Crossing Rate (ZCR) and Energy Entropy (EE). Then, the mean of two thresholds computed and consider the final threshold. This threshold used to identify male or female [14]. Ghosal.A et al. [6] has worked on the mechanical male-female voice discrimination model. The different types of time-domain features such as the zero-crossing rate, momentary energy, and maximum amplitude, these are frequency domain feature as well. The training and testing are carried out using the RANSAC method and neural-net classifier. G.S. Archana et al. [1] had done their researches using gender identification and performance analysis of speech signals. The pitch has applied for the gender discrimination of male and female voices.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author

Sandeep Kumar*, Lecturer, Department Of Computer Science, Government Polytechnic Banka, Banka, Bihar, India. E-Mail: Mailtoersandeep@Gmail.Com

Jainath Yadav,Asst. Professor, Department Of Computer Science, Central University Of South Bihar, Gaya, Bihar, India. E-Mail: Jainath@Cub.Ac.In

Here, we have found out that the outcomes given by SVM classification are more reliable than the artificial neural network. Kumari.M et al. [11] has put forward a method called a gender detection algorithm that deals in non-static of the speech signals. Here, we have found out the pitch, and with its help, we can find out the pitch gender. If the pitch range is loud, then it is a female voice, and if it is low, then it is a male voice. And the comparison of loud and quiet is made with the help threshold basis. Ramdinmawii.E et al. [18] gave the method of gender identification that, with the help of various speech signals, scrutinizes the speech signal producing features. In this entire procedure, what we have found out is signal features such as pitch, Mel-frequency coefficients, and signals energy. The lessons and trials of data were implemented using the SVM classifier. Mel-frequency coefficients precision had calculated as 69%. Shan. Li et al. [12] gave the multidimensional orator information identification system. The various speech qualities like pitch, formant, energy, linear prediction cepstrum coefficients (LPCC), and Mel-frequency cepstrum coefficients (MFCC) are progressively made to operate as a feature vector. Also, the SVM classifier had used for classification. Igor Bisio et al. [3] research had done with the help of audio signals recorded from speech samples. The features have extracted from the signal. A feature selection method is used to enhance classification accuracy. A support vector machine is used to build a classification ml. Gender has taken into consideration. Results show that involving gender and feature reduction steps increases the accuracy of classification by 81 percent.

III. PROPOSED METHODOLOGY

The proposed method uses a new feature extraction method, GMFCC and Deep LSTM as a classifier. It works well on an emotional speech dataset. All the previous research works had based on a threshold that fails to classify gender in an emotional speech dataset. Because the neutral emotional speech has different threshold value, compare to anger, fear, happy and sad speech emotion.

A. Emotional Speech Dataset: IITKGP-SESHC

We have gathered this data with the help of 5 male artists and 5 female artists from renowned Gyanavani FM radio station, Varanasi, India. All these professional artists were well-versed in their respective expertise to deliver the right emotion from the unbiased sentences. All the artists lie in the age group of 28 to 48 years with varied experience of 5-20 years. Similarly, female artists are of the age group of 2030 years with a 3-10 year of experience. For recording the emotions, 15 Hindi text prompts have used. The sentences used here are intensely unbiased in every sense. The artists are required to utter 15 sentences in 8 essential emotions in one session. The number of sessions recorded for preparing the database is 10. 12000 counts the total number of utterance and every emotion has got 1500 utterances. The word and syllables count in the sentences is 4-7 and 9-17, respectively. The time taken during the complete course is around 9 hours. Anger, Disgust, Fear, Happy, Neutral, Sadness, Sarcastic and Surprise are the emotions mentioned above used here for gathering the proposed speech emotion corpus [10]. Here, we have considered only five emotions from male and female speech voices, namely Anger, Fear, Happy, Neutral and Sadness.

B. Feature Extraction

1) GTCC: The Gammatone cepstral coefficients (GTCCs) are an anatomical inspired modification employing Gammatone filters with equivalent rectangular bandwidth bands.

Step I: Pre-emphasis the given speech signal to a high pass filter. Pre-emphasis aim to boost the high-frequency components.

$$z_1(n) = z_1(n) - \beta z_1(n-1) \tag{1}$$

Where β is called as pre-emphasis coefficient, ranging between 0.9 to 0.1

Step II: Frame Blocking is the process by which the input speech signal has fragmented into a frame size of 25 ms having 15 ms overlap. The sampling rate is supposed to be 1600 Hz ($25*10^{-3}*16000=4000$ samples). The frameshift has taken as 10 ms that allows some overlap of frames. Initially, the first sample begins at sample 0, and the corresponding 400 samples began at samples 160. We have sustained this entire process till the wind-up of speech file.

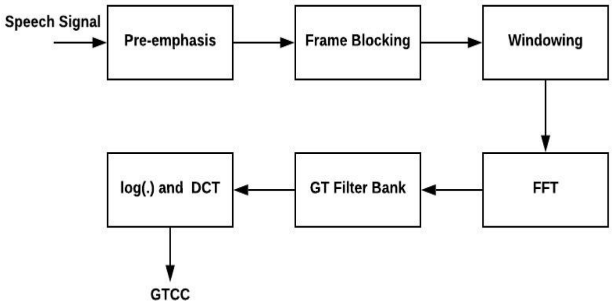


Fig. 1. GTCC Block Diagram

Step III: Hamming Window is a process to minimizes the delay and hindrance in getting the adequate signal both at the start and end of each frame. To achieve this, we have to multiply each frame with a hamming window ($W(n)$). $S(n)$ indicate signal in the frame, $n = 0,1,2...,N-1$.

$$W(n,a) = (1-a) - a\cos(\frac{2\pi n}{N-1})\ Where\ 0 < n < N-1 \tag{2}$$

Step IV: The Fast Fourier Transformation (FFT) converts a frame of N samples from the time domain scale to the frequency domain scale with the help of the following formula:

$$z_i(k) = \sum_{n=1}^{N} S_i(n)h(n)e^{\frac{-2\pi}{N}} \tag{3}$$

Where $S_i(n)$ signifies signal in the time domain scale and $S_i(k)$ signifies signal in the frequency domain scale for 1 to K, $h(n)$ is the window with N samples long and K is the length of the FFT.

Step V: Gammatone Filter Bank Analysis: This analysis is broadly used in the models of auditory system.

This function defined in time domain because of its impulse response

$$g(t) = at^{n-1}cos(2\pi ft + \theta)e^{-2\pi b_1 t} \qquad (4)$$

Here, n refers to Order of filter, b_1 refers to as Bandwidth of filter, a refers to as Amplitude, f refers to as Filter central frequency and θ refers to the phase.

Step VI: Compute log energy of square of summation of each filter bank energy. Then Compute DCT. In this step, DCT has been fitted to the log of energy acquired from the Gammatone rectangular filter bank. After DCT, we get the Gammatone cepstral coefficient(GTCC).

2) ***Skewness:*** It is a calculation of the unevenness of the data in the sample mean. If the outcome is negative, then we can see that data will turn towards the left of the mean rather than in the right direction. We can relate the skewness of the normal distribution is always 0. The skewness of a distribution is defined as

$$s = E(x - \mu)^3/\sigma^3 \qquad (5)$$

where μ stands for the mean of x, σ stands for the standard deviation of x, and $E(t)$ shows the desired value of the quantity t. The skewness calculate a sample version of the population value.

3) ***Pitch:*** Pitch of speech is that characteristic of speech by which an acute or shrill note has distinguished from a grave or a flat note. If a pitch is higher, the address is said to be sharp, and if the tone is low, the speech is dull. The various method can be used to compute pitch values such as zerocrossing, cepstrum, Average magnitude difference, etc. the autocorrelation describe as a time-domain method. In contrast, the cepstrum describe as a frequency-domain method.

4) ***GMFCC:*** The steps of the feature extraction algorithms are as follows:

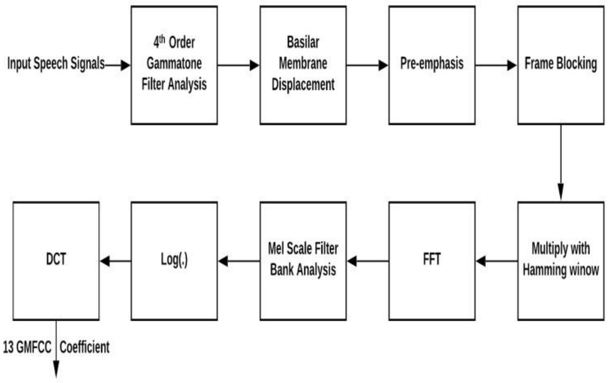

Input Speech Signals
4th Order Gammatone Filter Analysis
Basilar Membrane Displacement
Pre-emphasis
Frame Blocking
Multiply with Hamming winow
FFT
Mel Scale Filter Bank Analysis
Log(.)
DCT
13 GMFCC Coefficient

Fig. 2. GMFCC Block Diagram

Step I: Read the speech signal from dataset.

Step II: Gammatone Filter Bank Analysis: We know that Gammatone filter has extensively used in models of the auditory system. The Gammatone function defined in the time domain by its impulse response

$$g(t) = at^{n-1}cos(2\pi ft + \varphi)e^{-2\pi bt} \qquad (6)$$

Where, n = Order of filter, b = Bandwidth of filter, a = *Amplitude*, f = Filter central frequency and φ is the phase. Petterson et al. [16] has shown that surge behavior of the Gammatone function that has got the sequence number 4, gives the perfect figure to the human acoustic filter in diagrammatic shapes which we get by Peterson and Moore. It has been abridged by Glassberg and Moore [7]. The human data on the Equivalent Rectangular Bandwidth (ERB) of the acoustic filter having the following given function.

$$ERB = 24.7(4.37 * 10^{-3}f + 1) \quad (7)$$

Describing the fig 2, the order of the filter that recognize speech resemble to human ear is 4,$b_1$refers to the bandwidth of the Gammatone filter having value 1.019 ERB. The basic purpose of the Gammatone filter is that it simulates the motion of the Basilar membrane at a single place [13] .

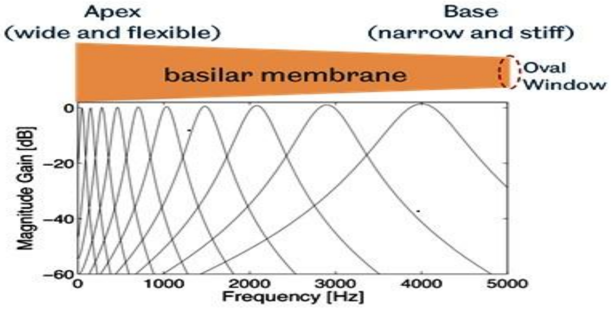


Fig. 3. The frequency response of the Gammatone filter bank consists of ten filters. 50 Hz and 4 kHz are the equally spaced centre frequencies respectively on the ERB scale [7].

Center Frequency (CF): It has expressed as the measurement of the Center frequency is falling between the uppermost and lowermost cutoff frequencies. In statistical term, it is either the arithmetic mean or the geometric mean of the lower cutoff frequency and the upper cutoff frequency. ERB scale is approximately logarithmic in nature, on which the center frequencies are equally spaced.

Step III: Pre-emphasis the Basilar displacement by using following equation:

$$b_m = filter([1 - preemph],1,b_m); \qquad (8)$$

Where, $preemph = 0.97$;

Step IV: Frame blocking: Take 25 ms frame duration with 10 ms overlap length.

Step V: Multiply the frame with Hamming window.

Step VI: Compute the Fast Fourier Transform of the b_m.

Step VII: Perform Mel-Scale Filter Bank Analysis using following equation.

$$F(Mel) = [2595 * log10(1 + F)700] \qquad (9)$$

Step VIII: Compute log of summation of Mel-Scale Filter Bank energy.

Step IX: Compute the Discrete Cosine Transform of log of the energy and obtain 13 GMFCC coefficient.

C. *Classifier*

LSTM : Since RNN fails to solve gradient vanishing problems (Increase in the duration of time gap between important event), then we use a special type of RNN that deals with the above problem is called LSTM units. The LSTM and the standard feedforward networks are distinguishable as the former one has got the lead in feedback connections, which marks it as a ”general purpose computer” (in other words, it has the capacity equivalent to a Turing machine). It has the potential to process single data points(e.g., images) as well as entire sequences of data (e.g., speech or video). These are remarkably composed of a cell that hold input till the next update, an input gate, an output gate and a forget gate. The cell function to recollect values above random time intervals and three gates are authorized to monitoring the flow of information inside and outside the cell. Here, the cell is superintended to keep an eye on the reliance of elements lying in the input sequence. It is designated that the input gate administers the level to which a new value flows into the cell, the duty of a forget gate is to adhere to the level to which the values reside in the cell and lastly the output gate monitor the level which is beneficial in calculating the output setting-off.

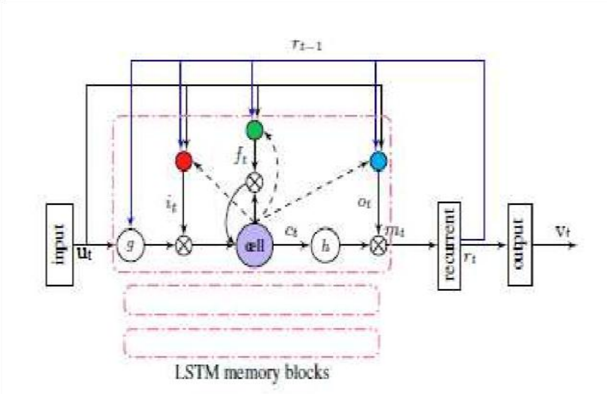


Fig. 4. LSTM [19].

From input sequence $u = (u_1,...,u_T)$ to get an output sequence $v = (v_1,...,v_T)$ by systematically calculating the network activation function for each network unit using the following series of equations from $t = 1$ to T:

Let $m_t = s_t$, $b = e$ and $c_t = d_t$.

$$i_t = \sigma(W_{iu}u_t + W_{is}s_{t-1} + W_{id}d_{t-1} + e_i) \quad (10)$$

$$f_t = \sigma(W_{fu}u_t + W_{fs}s_{t-1} + W_{fd}d_{t-1} + e_f) \quad (11)$$

$$d_t = f_t \odot d_{t-1} + i_t \odot g(W_{du}u_t + W_{ds}s_{t-1} + e_d) \quad (12)$$

$$o_t = \sigma(W_{ou}u_t + W_{os}s_{t-1} + W_{od}d_t + e_o) \quad (13)$$

$$m_t = o_t \odot h(d_t) \quad (14)$$

$$v_t = \varphi(W_{vs}s_t + e_v) \quad (15)$$

Deep LSTM: Since Deep Neural Network (DNN) with deeper architectures has been used for speech emotion recognition then deep LSTM have been successfully used for speech recognition [5], [8], [9]. Deep LSTM are built by stacking multiple LSTM layers as required by problem in hand. Minibatch gradient descent is a sequence that breaks the training dataset into compact batches which are further used to compute model error and is helpful in modernizing the model coefficient. The miniBatchSize is nor too large nor too small because of negative impact on LSTM network.

1) Proposed Model: The systematic sketch block of the proposed gender Deep LSTM is shown in fig 4. The proposed model contains eight layers. Input layer take sequential input from GMFCC feature vector and pass through first LSTM unit with 100 neuron in its hidden layer then pass through Dropout layer (0.2), which minimize the overfitting of training progress and again pass through second LSTM unit with 125 neuron in its hidden layer then Dropout layer (0.2) and so on as in block diagram.

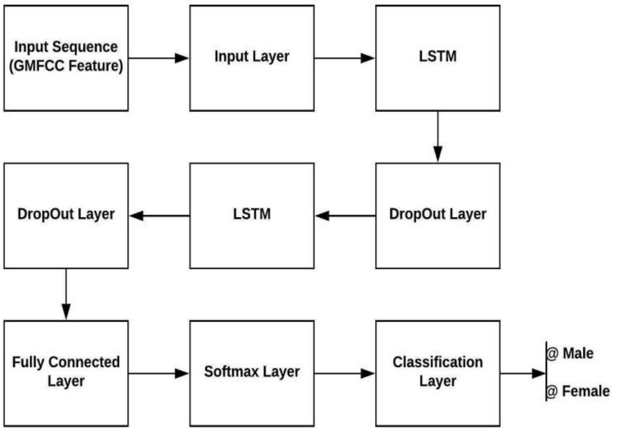
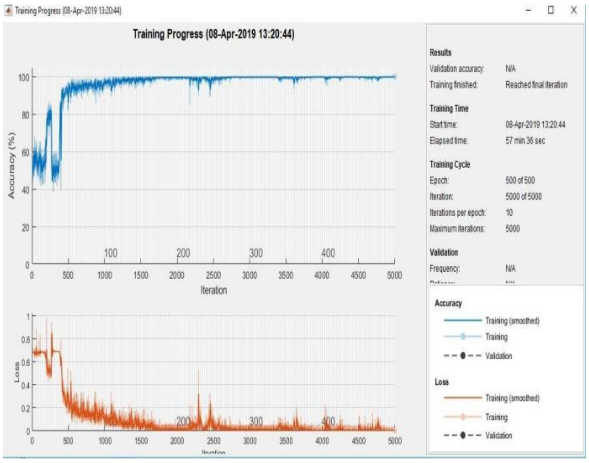


Fig. 5. Proposed Gender Deep LSTM Block Diagram

IV. EXPERIMENT AND RESULT

To train proposed model we are using 80% training data, which has been derived from IIT Kharagpur Simulated Emotion Hindi Speech Corpus (IITKGP-SEHSC) with specified target. And train under GPU (Graphical processing unit) environment up to 500 epoch for 57 min 36 sec with miniBatchSize=120 on Matlab as in fig 5. This model gives average training accuracy of 99.9%. Since miniBatchSize should not be too greater or too smaller because of the negative impact on Deep LSTM. To overcome training overfitting Dropout Layer with 20% dropout of training data is added. So that, it minimizes training overfitting and gives percent average testing accuracy with testing data of 98.3%. The proposed model also trained with features (i) feature finding by fusing traditional MFCC with pitch (ii) Feature by combining Gammatone cepstral coefficient (GTCC), pitch and skewness individually.



Results
Validation accuracy: N/A
Training finished: Reached final iteration

Training Time
Start time: 08-Apr-2019 13:20:44
Elapsed time: 57 min 36 sec

Training Cycle
Epoch: 500 of 500
Iteration: 5000 of 5000
Iterations per epoch: 10
Maximum iterations: 5000

Validation
Frequency: N/A

Fig. 6. Training Progress Of Gender Deep LSTM

Table I describes the percentage rate of male-female gender recognition rate for different speech emotion types. We can observe that the gender male recognition rate for anger and happy emotions is maximum and minimum in the case of fear emotion. The neutral and sad that is middle of max and min. In the case of female, the gender recognition rate of anger, happy, neutral and sadness emotion is the maximum and minimum for the fear emotion. In the fig 7 show the percentage gender identification rate for emotional speech with emotion anger, fear, happy, neutral and sad. Above figure shows that % gender recognition rate of female emotional speech is higher than the male emotional speech.

Table I
Percentage Emotion Gender Recognition Rate From Proposed Classifier And Gmfcc Feature

Gender	Speech Emotion Types				
	Anger	Fear	Happy	Neutral	Sadness
Male	99.3	94.0	99.5	96.68	96.69
Female	99.4	98.46	99.3	99.8	99.7

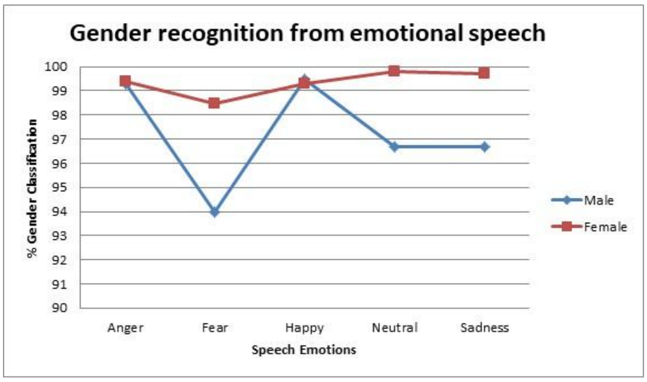


Fig. 7. Gender Recognition From Emotional Speech

In Table II describes the percent average gender recognition rate obtained from feature or combination of features. The feature used for comparison are (*i*) Mel Frequency Cepstral Coefficient (MFCC) with pitch (*ii*) Gammatone Cepstral Coefficient (GTCC) including Δ GTCC, ΔΔ GTCC, pitch, skewness, etc (*iii*) Proposed GMFCC feature that pass through Deep LSTM. We can observe that the percent average gender recognition rate from proposed GMFCC feature is 98.3%, which is maximum.

Table Ii
Average Percentage Emotion Gender Recognition Rate

Feature	MFCC + Pitch	GTCC + Pitch+ Skewness	GMFCC
% Avg. Gender Recognition Rate	92.7	93.6	98.3

V. CONCLUSION

Here in the given paper, we have proposed new features Gammatone Mel Frequency Ceptral Coefficient (GMFCC), which gives very efficient feature sequence that closely resembles to human ear basilar membrane displacement. This feature sequence pass through Deep LSTM, which has two LSTM units. And, the unit carries 100 and 125 hidden neuron in its hidden layer respectively. Each LSTM unit is followed by dropout layer with probability 0.2 that minimizes the training progress overfitting. The batching of feature sequences is done in such a manner that there is no negative impact on LSTM network. That means the size of batching of feature sequence is neither too large nor too small. Since LSTM based classifier, classifies the sequence data very well. Before this paper all the research were based on threshold based pitch value or combination of threshold and MFCC, combination Gammatone Cepstral Coefficient (GTCC) including Δ GTCC, ΔΔ GTCC, pitch, skewness, etc. The average maximum percent gender identification accuracy before this paper for emotional speech containing anger, fear, happy, neutral and sad emotion is 93.6 %. Reason being that the neutral emotion has high threshold pitch for male and low threshold pitch for female which is opposite in nature to anger, fear, happy and sad emotion. But in this paper we have used basilar membrane displacement based Mel cepstral coefficient, GMFCC, which is very susceptible, where pitch variation in speech took place and produced the best result with emotional speech dataset and Deep LSTM classifier. It gives average gender identification rate 98.3 %.

REFERENCES

1. GS Archana and M Malleswari. Gender identification and performance analysis of speech signals. In *2015 Global Conference on Communication Technologies (GCCT)*, pages 483–489. IEEE, 2015.
2. Suzanne Beeke, Ray Wilkinson, and Jane Maxim. Prosody as a compensatory strategy in the conversations of people with agrammatism.
Clinical linguistics & phonetics, 23(2):133–155, 2009.
3. Igor Bisio, Alessandro Delfino, Fabio Lavagetto, Mario Marchese, and Andrea Sciarrone. Gender-driven emotion recognition through speech signals for ambient intelligence applications. *IEEE Transactions on Emerging Topics in Computing*, 1(2):244–257, 2013.
4. Moataz El Ayadi, Mohamed S Kamel, and Fakhri Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3):572–587, 2011.
5. Florian Eyben, Martin Wollmer, Bj¨ orn Schuller, and Alex Graves. From¨ speech to letters-using a novel neural network architecture for grapheme based asr. In *2009 IEEE Workshop on Automatic Speech Recognition & Understanding*, pages 376–380. IEEE, 2009.

6. Arijit Ghosal and Suchibrota Dutta. Automatic male-female voice discrimination. In *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, pages 731–735. IEEE, 2014.
7. Brian R Glasberg and Brian CJ Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1-2):103–138, 1990.
8. Alex Graves, Navdeep Jaitly, and Abdel-rahman Mohamed. Hybrid speech recognition with deep bidirectional lstm. In *2013 IEEE workshop on automatic speech recognition and understanding*, pages 273–278. IEEE, 2013.
9. Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.
10. Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. Iitkgp-sehsc: Hindi speech corpus for emotion analysis. In *2011 International conference on devices and communications (ICDeCom)*, pages 1–5. IEEE, 2011.
11. Mamta Kumari, Nilakshi Talukdar, and Israj Ali. A new gender detection algorithm considering the non-stationarity of speech signal. In *2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS)*, pages 141–146. IEEE, 2016.
12. Shan Li, Longting Xu, and Zhen Yang. Multidimensional speaker information recognition based on proposed baseline system. In *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 1776–1780. IEEE, 2017.
13. Ning Ma, Phil Green, Jon Barker, and Andre Coy.´ Exploiting correlogram structure for robust speech recognition with multiple speech sources. *Speech Communication*, 49(12):874–891, 2007.
14. Kunjithapatham Meena, Kulumani R Subramaniam, and Muthusamy Gomathy. Gender classification in speech recognition using fuzzy logic and neural network. *Int. Arab J. Inf. Technol.*, 10(5):477–485, 2013.
15. Joy Nicholson, Kazuhiko Takahashi, and Ryohei Nakatsu. Emotion recognition in speech using neural networks. *Neural computing & applications*, 9(4):290–296, 2000.
16. Roy D Patterson, KEN Robinson, John Holdsworth, Denis McKeown, C Zhang, and Michael Allerhand. Complex sounds and auditory images. In *Auditory physiology and perception*, pages 429–446. Elsevier, 1992.
17. Kumar Rakesh, Subhangi Dutta, and Kumara Shama. Gender recognition using speech processing techniques in labview. *International Journal of Advances in Engineering & Technology*, 1(2):51, 2011.
18. Esther Ramdinmawii and VK Mittal. Gender identification from speech signal by examining the speech production characteristics. In *2016 International Conference on Signal Processing and Communication (ICSC)*, pages 244–249. IEEE, 2016.
19. Has¸im Sak, Andrew Senior, and Franc¸oise Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Fifteenth annual conference of the international speech communication association*, 2014.

AUTHORS PROFILE

Mr. Sandeep Kumar, Lecturer at Government Polytechnic Banka, Banka, Bihar-813105 has two year teaching experienced and appox. two year industry experienced. He has completed their M.TECH degree under guidance of Dr. Jainath yadav.

Dr. Jainath Yadav, Asst. professor, Department of Computer Science under School of Mathematics, Statistics and Computer Science Central University of South Bihar , Gaya , Bihar -824236 He has contributed several research papers in the referred journals like IEEE Transactions on Audio Speech and Language Processing, IEEE Signal Processing Letters, Speech Communication, etc.