International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Modification of Prosody for Emotion Conversion using Gaussian Regression Model

Geethashree A, D J Ravi

Abstract*: Emotion conversion is one of the most inspiring forefronts of research in the arena of emotional speech synthesis. The main focus of the work is to convert a neutral speech sentence to the target emotional speech sentence using signal processing techniques. The parameters used for emotion conversion are pitch contour and intensity along with the duration of the sentence. Kannada Emotional Speech (KES) Database is created and used for analysis. The database consists of 4 (sadness, happy, anger, and fear) emotions with neutral. The pitch contour of different emotional sentences are analyzed and Gaussian Regression Model (GRM) is proposed for predicting the target pitch contour. The evaluation of the proposed method is done using Objective test & Subjective test. For objective test, mean pitch, the standard deviation of pitch, mean intensity and duration of the sentences are used. Evaluation using a subjective test is performed by calculating Emotion Recognition Rate (ERR) with the help of confusion matrix and also by taking the Mean Opinion Score (MOS) rating of the conversion system on the scale of 1-5. The result of Subjective test indicates that the effectiveness and discernment of emotion are improved when GRM is used for pitch contour modification with intensity and duration. The most recognized emotion was sadness with MOS of 3.52 and ERR of 83% and the least recognized emotion was anger with MOS of 1.74 and ERR of 66%. The results of the subjective and objective test show that the converted sadness, happy and fear speech is seeming very close to usual sadness, anger and fear emotion.

***Index Terms*: Emotion Conversion, Gaussian Regression Model, Kannada Emotional Speech Database.**

I. INTRODUCTION

Human-Machine Interaction (HMI) is one of the evolving field of research. Human communicates with machines in a large number of modes, like text, speech, and some input/output devices. Amoung which speech is one of the most spontaneous and regular modes of communication. Speech synthesis and speech recognition are the two noticeable methods used in HMI. In speech recognition, the speech signal is the input, the machine tries to recognize what is articulated and displays the text on the output device, i.e., machine converts speech to text. Speech synthesis or Text-To-Speech (TTS) systems are just the opposite of speech recognition, given a text, the machine needs to convert it to speech. Many types of speech synthesizers are available in different languages, but still, the output of these text-to-speech systems do not sound natural. There are two standard approaches to add expressiveness to the output of the TTS system. The first approaches are corpus-driven unit selection [1] and the second one is modification by post-processing [2].

Cen [3] stated that a highly natural sounding speech can be synthesized using corpus-driven approach, where a very large amount of unit selection database which contains sufficient unit to match the target speech for synthesis is used. The database should contain the recording of each unit in all the required emotions. During synthesis, the required units in the required emotion from the database are selected. This requires a huge amount of memory space, time and very high processing speed.

Post processing modification is a cost-effective speech synthesis approach with the reduction in the expressiveness. An ordinary TTS system will convert the given text to a neutral speech, by using different signal processing technique on neutral speech the expressiveness is added to it. However, the application of emotion conversion is well behind the scope of HMI.

The proposed method converts the neutral speech sentence to target emotional speech sentence so that the converted sentence is perceptually similar to natural emotional speech. Emotion conversion in speech requires manipulation of a wide range of prosodic features like pitch, intonation, intensity, speaking rate, vocal quality, and articulation [4, 5]. Among which Pitch and intonation modification is one of the integral parts of emotion conversion systems. For pitch contour modification Gaussian Regression Model is used. Intensity and duration modification is done at sentence level with a different scaling factor. These scaling factors are obtained by analysis of emotional sentences with respect to the neutral sentences.

The rest of the paper is structured as follows. Section II presents a literature review of related work. Section III describes the Kannada Emotional Speech (KES) database. Section IV analyzes the prosody parameters of neutral and emotional speech. The parameters used in analysis are pitch, intensity and duration. Section V proposes Gaussian Regression Model (GRM) for modification of pitch contour of neutral speech to target emotional speech. Section VI discusses the results and VII concludes with a summary.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Geethashree A*, Department of ECE, Visvesvaraya Technological University, Vidyavardhaka college of Engineering, Mysuru, Karnataka, India.
D J Ravi, Department of ECE, Visvesvaraya Technological University, Vidyavardhaka college of Engineering, Mysuru, Karnataka, India.

II. LITERATURE REVIEW

The technique in which the parameters of input speech is manipulated to the target emotion and then resynthesized using new parameters is called emotion conversion. Transformation of the emotion in speech involves manipulation of a wide range of prosodic parameters like pitch, intonation, intensity, speaking rate, vocal quality, and articulation [4,5]. In literature, many works are available on emotional speech synthesis. In Formant synthesis [6], a large number of emotions in speech can be modeled, but the output speech is distinct and has a robot-like quality due to which it remains unacceptable. Unit selection proposed in [7] can synthesize extremely natural-sounding speech, with the huge unit-selection database. Prosodic rule-based (RB) approach is employed to model prosodic features [8, 9]. The rules are formed by analyzing the difference between emotional and neutral speech. RB approach is relatively simple and direct when compared to other methods, the rules employed decides the naturalness and quality of the emotional speech. The rule-based approaches have been used in English [10], Dutch [11], Spanish [12], Catalan [13], German [14], Korean [15] and some Indian languages [16, 17, 18].

In statistical methods, a mapping between emotional speech and neutral speech is constructed. Some of the statistical methods are Vector Quantization (VQ) Codebook [19], Classification and Regression Tree (CART) [20] and Gaussian Mixture Model (GMM). In GMM method the spectrum of the speech signal is used in the conversion and the prosodic parameters like pitch and duration are ignored. In RB methods prosodic parameters are modified, while the spectrum of the speech signal is ignored.

Emotional conversion is a post-processing technique, wherein the neutral speech from the synthesizer is obtained and the transformation function is applied either in sub-segmental level or sentence level [21, 22]. Tao [20] used vocal track features and excitation source parameters in conversion. [23, 24] adapted copy synthesis approach and stated that both prosodic and spectral parameters need to be modified to obtain good results. Turk [25] showed that the identification rate of each emotion can be increased by combining spectral and prosodic transformation, but this decreases the output quality. Several machines learning techniques have been used to map the neutral speech parameters to different emotions. Numerous techniques like Linear Modification Model (LMM), Gaussian Mixture Model (GMM), and Classification and Regression Tree (CART) were explored for pitch conversion. A comparison between these methods was presented in [20].

Govind et al [32, 33], has analyzed excitation source information across different emotions and proposed an emotion conversion approach by using source and supra-segmental information. The parameters used were instantaneous pitch, the strength of excitation and duration of the syllable. The epoch strength and prosody modification was done by scaling the Hilbert envelope (HE) of the LP residual. Vuppala and Rao [35] have proposed a method of non-uniform duration modification for emotion conversion in speech. Yadav and Rao [36] recently explored feed-forward neural network models for mapping the prosodic parameters between neutral and target emotions. Pathak [34] has used Discrete Wavelet Transform (DWT) for modeling emotional speech of Source and Target speakers. Haque [17] has used spectral energy, epoch strength and epoch sharpness with Pitch and intensity. Filter bank approach was used to modify energy spectra and the pitch contour of target emotion was predicted using Gaussian Normalization and polynomial regression method. Recently, Singh [37] has used Quadratic Multivariate Polynomial (QMP) for converting neutral speech to emotional speech.

In this work, an attempt is made to convert a neutral pitch contour of a sentence into a target emotional pitch contour by fitting a Gaussian function. The peaks of the pitch contour are fitted using a Gaussian function with peak values of 1 to 6. The linear least square fitting method is used for fitting the speech signal and trust region algorithm is used for the optimization of the fit. Guo has given a simple and improved algorithm for estimating the parameters of a Gaussian function fitted to observed data points [26]. Pastuchová [27] has compared Caruana and Guos fitting algorithms for Gaussian fitting of sensor data. Conder [28] has stated that bell curves are applicable for understating many observations and measurements across the sciences. The review on Trust region algorithm for nonlinear optimization is given in [29] and states the method is robust and can be applied to an ill-conditioned problem [30].

III. KANNADA EMOTIONAL SPEECH (KES) DATABASE

An emotionally rich simulated parallel KES database is used for prosody conversion. The database contains 4 primary emotions (sadness, anger, fear and happy) with neutral by four Kannada native actors two male and two female. 30 sentences in Kannada, which can be reproduced in all the 4 emotions were selected from Kannada textbook and given to the speaker for preparation. The speakers were asked to record the sentences with bursting emotion in an acoustic room to reduce HNR and SNR. The sampling rate of 44.1 KHz and 16-bit precision with a mono channel was used in the recording. Recording of all the speakers took place in different sessions to prevent influencing each other's speaking style. 4 successive recording was done and the best one was used for conversion. The database was also evaluated using 20 listeners and LVQ classifier. The recognition rate by listeners was around 95 percent and LVQ classifier with LFCC coefficients was 71 percent [31]. Simulated Parallel KES database consisting of a total of 600 sentences with minimum 3 to maximum of 10 words.

IV. PROSODY ANALYSIS OF NEUTRAL AND EMOTIONAL SPEECH

The proposed method uses 800 parallel sentences which includes neutral, happy, sadness, anger and fear speech from the database for analysis.

Using the PRAAT tool all the sound files are manipulated with a time step of 0.01 sec, the maximum pitch of 600 Hz and minimum pitch of 75 Hz, then the pitch tier, intensity tier, and duration tiers are extracted and stored in a one-dimensional array. From pitch tier mean pitch ($F0_{mean}$) and standard deviation of pitch ($F0_{SD}$), mean intensity from intensity tier and duration from the duration tier is extracted and tabulated. The conversion rule for intensity and duration modification is expressed as

$$y_T = \alpha_{NT} X_N \quad (1)$$

Where X_N represents prosody parameters of neutral speech y_T is target emotional prosody output (where T indicates target emotion N-Neutral, S-Sadness, F-Fear, A-Anger, and H-happiness) and α_{NT} is the Transformation scaling factor of the prosodic parameters from neutral the target emotions. The transformation scaling factor α_{NT} is estimated based on the analysis of parallel sentences in the KES database. The scaling factor α_{NT} is expressed as

$$\alpha_{NT} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_{Ti}}{X_N} \quad (2)$$

Where y_{Ti} represents prosody parameters of the recorded target emotion and n represents the number of parallel sentences used in the analysis process to estimate the conversion scaling factor. Intensity and duration scaling factors are obtained by taking the ratio of emotional speech parameters to the neutral speech parameter using Eq. (2) and are tabulated in Table I.

From Table I it can be analyzed that, anger emotions have smaller durations, higher mean pitch, and intensity when compared to other emotions. Whereas fear and sad has a larger duration, almost the same mean pitch, and intensity with respect to neutral. Happy has the higher $F0_{Mean}$, $F0_{SD}$, and Duration when compared to neutral

A. Pitch and Intensity Contours

As in linguistics, the pitch contour of a speech is a curve that tracks the observed pitch of the sound over time. Many studies in different languages have shown that pitch contour or intonation reflect specific emotions [9]. The present work analyzes the intonation pattern of different emotions in the Kannada language. Fig 1 shows the variation of Pitch contour and intensity contour of female speaker for different emotions, where (a - Neutral, b – Happy, c – Sadness, d – Anger, e - Fear). The pitch contour of neutral, sadness and fear are almost flat when compared to happy, and anger.

Table I: Scaling factor of features for emotion conversion from neutral to different target emotions

Scaling Factor		$F0_{Mean}$	$F0_{SD}$	Intensity	Duration
Female	α_{NN}	1	1	1	1
	α_{NH}	1.1	1.5	1	1.1
	α_{NS}	1.3	0.9	1.0	1.2
	α_{NA}	1.8	2.0	1.2	0.9
	α_{NF}	1.3	0.9	0.9	1.3
Male	α_{NN}	1	1	1	1
	α_{NH}	1.4	1.8	1.1	1.0
	α_{NS}	0.9	0.4	1.0	1.1
	α_{NA}	1.7	1.7	1.2	0.9
	α_{NF}	1.0	0.5	1.0	1.2

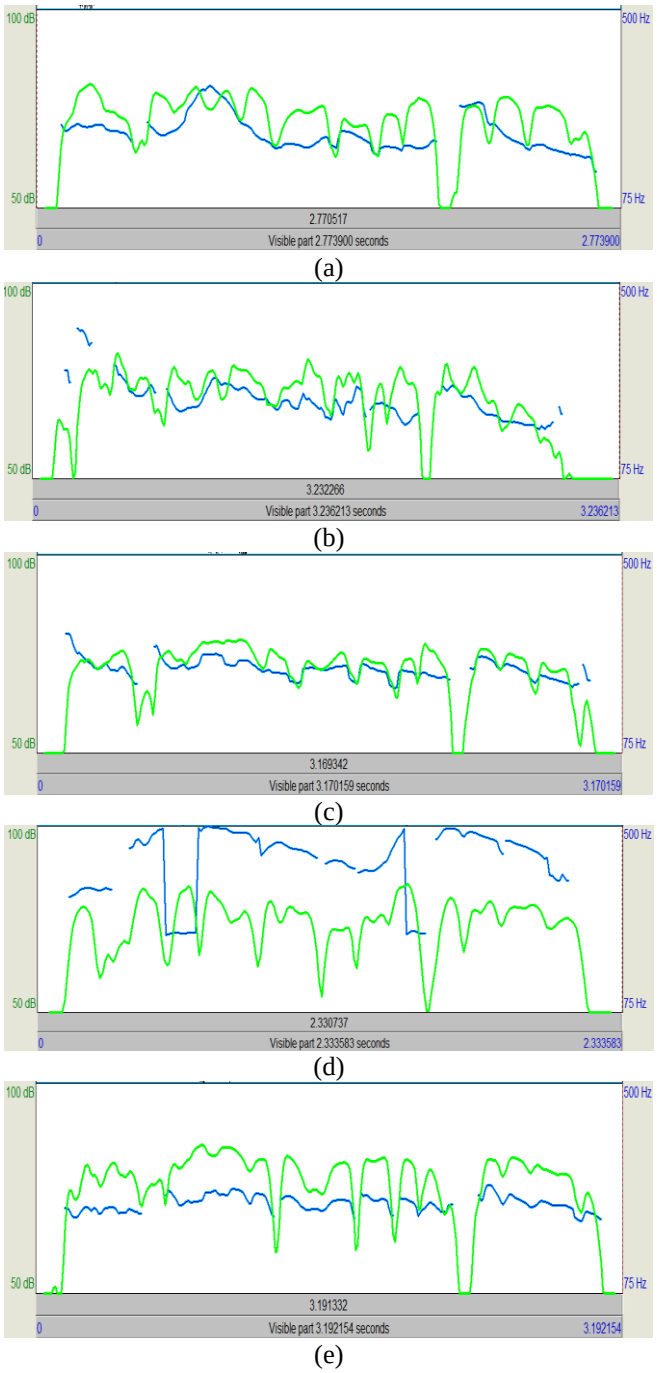


Figure.1 Pitch (blue) and Intensity (green) contour of Kannada sentence "ನಾನು ಸಾಮಾನ್ಯವಾಗಿ ಓದುವುದು ರಾತ್ರಿವೇಳೆಯಲ್ಲಿ" (I usually study at night) *where (a - Neutral, b – Happy, c– Sadness, d – Anger, and e - Fear)

In happy the pitch contour rises at the end of the sentence, there is a deep rise-fall, rise-fall contour in happy and anger. The emotions like sadness, anger and fear requires the change in intensity level. From the analysis, it is observed that the intensity level of fear and sadness is lower than neutral, whereas the intensity of anger is higher than that of neutral. Intensity modification is done using a function scale intensity in PRAAT tool.

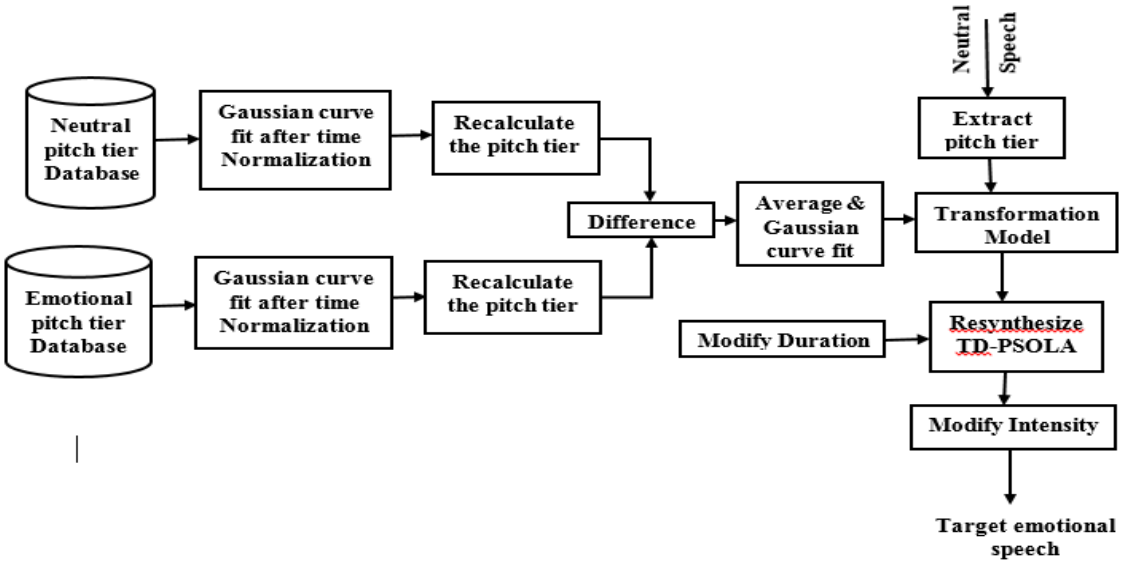


Figure. 2 Emotion conversion system using GRM

Scale intensity function multiplies the pitch-converted speech by scaling factors shown in Table I.

Duration conversion is performed in sentence level. From the analysis, it is observed that the duration of sadness and fear has to be increased, while anger has to be decreased by the factor shown in Table I. Lengthen (overlap-add) function in PRAAT tool has been used to modify duration.

V. GAUSSIAN REGRESSION MODEL (GRM)

The dynamic feature of speech such as pitch contour has a significant effect on the emotion in speech. The static features such as Z-score, standard deviation (σ) and mean (μ) pitch mainly characterize the voice. The paper proposes a Gaussian curve fitting method to model the transformation function for modification of neutral pitch contour to a fresh pitch contour that would nearly imitate the pitch contour of target emotional speech. As the Gaussian curve fitting method fits the number of peaks in the given signal, the method is called Gaussian Regression.

By fitting the data to a curve of known form, the information related to the time series of discrete data points can be extracted, thereby reducing the data to a few describable parameters [28]. By doing this the system will be easier to describe and simple to understand.

Guo H. [26] stated that Gaussian function is of the form

$$Y = Ae^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]} \quad (3)$$

The Gaussian function is a symmetric bell curve that is centered at the position x = μ. The tails on both sides of the bell peak fall quickly and approach the x-axis. Where peak height and peak width is denoted by A and σ respectively. Determination of parameters A, μ, and σ is not an easy task, as it is associated with an over-determined system of nonlinear equations and its solutions, which is obtained by substituting the data into Eq. (3). The standard solution is obtained by using the iterative algorithm proposed by Newton-Raphson.

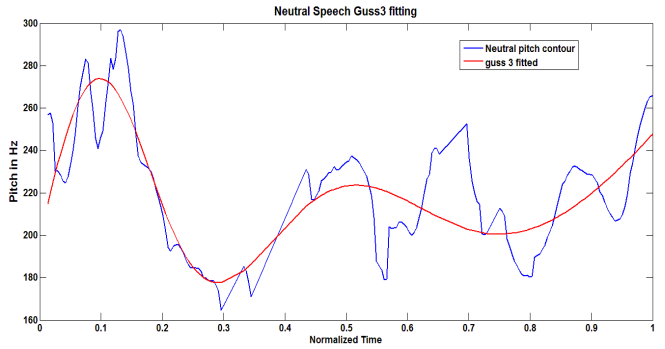
The focus of the work is to fit Gaussian Model to the pitch contour of neutral and target emotional sentences for the purpose of emotion conversion. Fig.2. shows the steps followed in the proposed method. The Gaussian model fits peaks and is given by

$$Y = \sum_{i=1}^{n} Ae^{\left[-\left(\frac{x-\mu_i}{\sigma_i}\right)\right]} \quad (4)$$

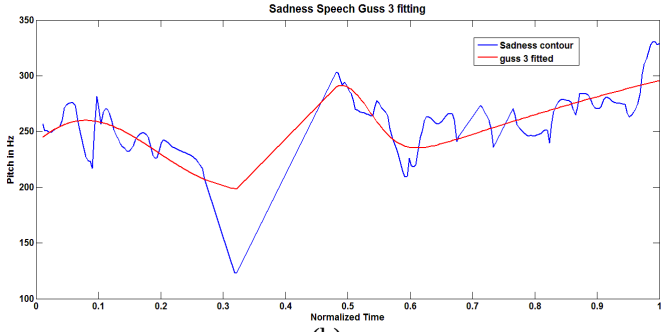
Where n denotes the number of peaks to fit ($1 \leq n \leq 8$). Nonlinear Least Squares method is used to fit the Gaussian curves to the pitch contour data. Trust – Region algorithm is used to further reduce the least squares error. For modifying the pitch contour of neutral utterance to target emotional utterance, 70% of parallel utterance from neutral, happy, sad, Anger and fear was used for generation of transformation model and 30% of the database was used for conversion.

Neutral and emotional pitch tier database was prepared by extracting the pitch tier of both neutral and the corresponding emotional utterances using PRAAT. For conversion, the neutral pitch tiers are modified by the transformation model obtained in the training procedure, then the generated new pitch tiers are resynthesized. An example of the steps followed for prosody conversion of the sentence "□□□□□□□□□□□□□□□, □□□□□ □□□□□□ □□□□□□□□□?" (I realized, what is the reason for your fear?) is shown in Fig 3.

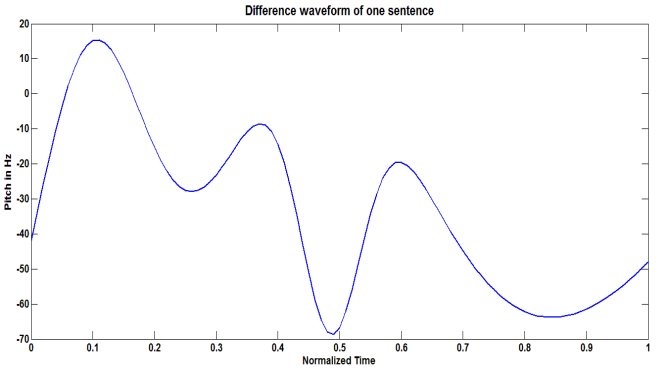
The Guss3 fitting of a neutral sentence is shown in Fig. 3(a). Guss3 fitting of the emotional sadness sentence is shown in Fig. 3(b). The difference of recalculated pitch tiers of Guss3 fitted neutral and sadness sentences are shown in Fig. 3(c). The transformation curve generated for neutral to sadness conversion when 10 sentences were used for training is shown in Fig. 3(d).



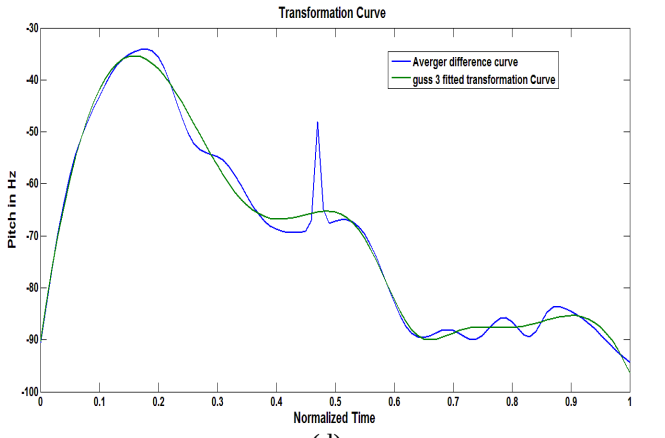
(a)



(b)



(c)



(d)

Figure.3. Example of training phase of GRM. (a) Neutral pitch tier Guss3 fitted, (b) Sadness pitch tier Guss3 fitted, (c) Difference of recalculated pitch tiers after Guss3 fitting, and (d) Transformation curve generated by Guss3 fitting of the averaged difference curves

VI. RESULTS AND DISCUSSION

The proposed method is evaluated by conducting subjective and objective tests. The evaluation is done at 4 different stages of the conversion process: (i) Pitch contour (P), (ii) Pitch contour + Intensity (P+I), (iii) Pitch contour + Duration (P+D), and (iv) Pitch contour + Intensity + Duration (P+I+D).

A. Objective Evaluation

The parameters used for the objective test are $F0_{Mean}$, $F0_{SD}$, I_{Mean} and D. 20 sentences (10 Male & 10 Female) in each of the recorded and converted emotions are used in the evaluation. The results of the objective test are tabulated in Table II. The objective test results show that the $F0_{Mean}$ and $F0_{SD}$ values of pitch, intensity, and duration of the recorded emotional speech of KES database and that of converted using GRM are very close to one another.

B. Subjective Evaluation

The effectiveness of Gaussian Regression model is demonstrated using subjective listening tests. The Google forms were created for the subjective listening test. The converted 10 sentences (5 male and 5 female) in all 4 emotions were divided into 4 groups. A reasonable number of subjects were asked to evaluate the converted sentences. Subjects with basic knowledge of speech and native speakers of Kannada were used for evaluation.

The subjects were made to listen to the neutral sentence first and then the converted sentence. They were asked to recognize the converted emotion first and then give the opinion score of the same on the scale of 1-5. Table III gives the details of the rating scale used for subjective evaluation.

A subjective test was performed to assess the extent to which the converted speech is perceived as having the intended expressivity. The Mean Opinion Score (MOS) of the subjective test is calculated for all combination of parameters i.e. P, P + I, P + D and P + I + D.

Table II: Objective test of GRM for converter sad, happy, anger & fear speech

	Female				Male			
	$F0_{Mean}$ (Hz)	$F0_{SD}$	I(dB)	D (sec)	$F0_{Mean}$ (Hz)	$F0_{SD}$	I(dB)	D (sec)
Recorded Sadness	256	30	69	2.9	105	8	70	2.6
Converted sadness	238	28	70	3.1	104	15	69	2.9
Recorded Fear	276	27	67	2.9	124	15	71	3.1

Converted Fear	267	29	68	3.5	179	25	74	3.5
Recorded Anger	413	90	70	2.6	285	41	76	2.1
Converted Anger	404	77	77	2.3	190	45	80	2.1
Recorded Happy	285	35	72	2.8	174	60	77	2.4
Converted Happy	285	35	79	2.7	164	26	80	2.3

Table III. Rating scale used for subjective listening test

Quality of speech	Expressiveness	Rating
Very poor	Sounds exactly like neutral	1
Poor	Sounds slightly different from neutral	2
Fair	Sounds slightly like target	3
Good	Sounds more like target	4
Excellent	Sounds exactly like target	5

Table IV. Evaluation of GRM using MOS and Emotion Recognition Rate (ERR)

	Parameters	P	P+I	P+D	P+I+D
Converted Sadness	**MOS**	2.8	3.2	3.2	3.52
	ERR (%)	70	73	79	83
Converted Fear	**MOS**	2.5	2.6	2.6	2.71
	ERR (%)	63	65	65	70
Converted Anger	**MOS**	1.3	1.5	1.5	1.74
	ERR (%)	60	65	61	66
Converted Happy	**MOS**	2.9	3.2	3.1	3.41
	ERR (%)	62	63	64	68

It is evident from the literature review that pitch contour plays a vital role in any emotion conversion system. Therefore modification of the pitch using GRM was evaluated separately. As the MOS and the percentage of recognition obtained were satisfactory for sadness, happy, anger and fear converted sentences, modification of other features like intensity and duration were combined with pitch contours.

Then the evaluation of each combination was done. Table IV summarizes GRM performance. The evaluation of GRM witnessed the important role of pitch contour in the perception of emotions. Fig 5 shows the neutral and GRM converted pitch contour of sadness, fear, anger, and happy for the sentence "ಇವನ□□□□ □□□□ □□□□□□□□." (I will come along with him). The MOS score and ERR show that converted speech is satisfactory in terms of emotion perception in sadness, fear, and anger converted sentences when all the three parameters pitch, intensity, and duration are modified.

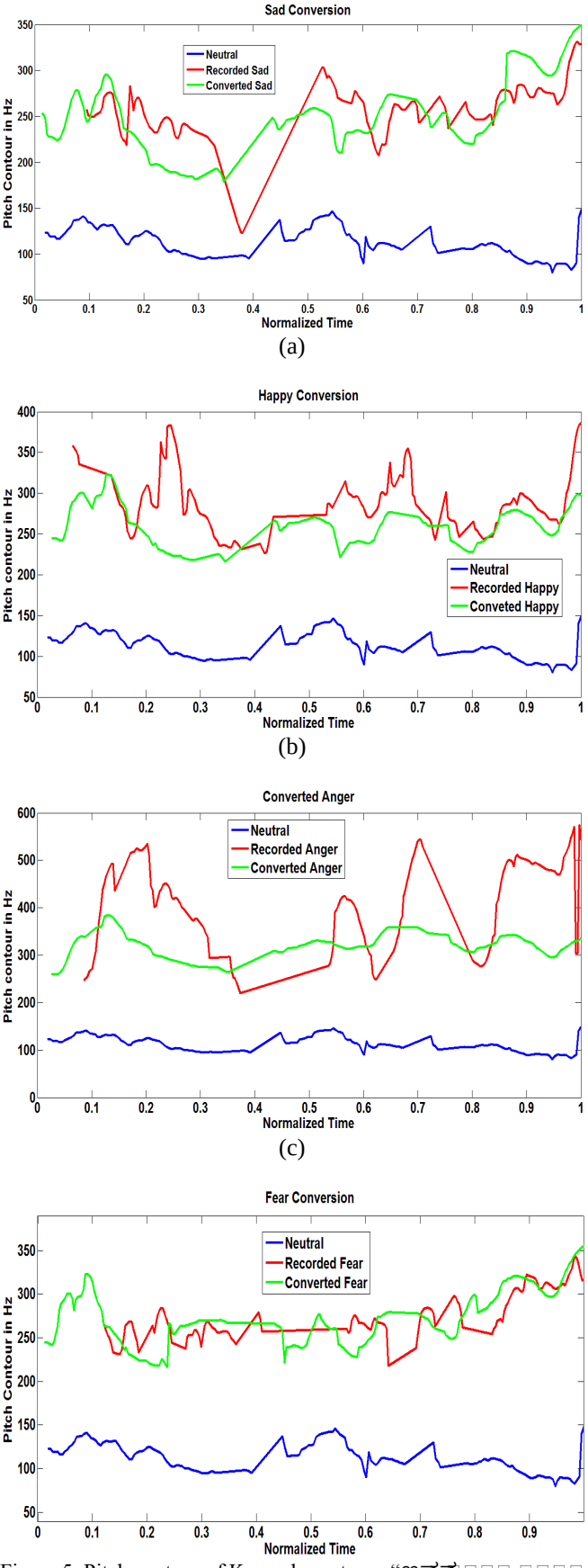


Figure 5. Pitch contour of Kannada sentence "ಇವನ□□□□ □□□□ □□□□□□□□." (I will come along with him) of female speaker in neutral, recorded target emotion and converted target emotion. (a) Sadness conversion, (b) Happy conversion, (c) Anger conversion, and (d) Fear conversion.

The MOS of 2.8, 2.5, 1.3, and 2.9, ERR of 70%, 63%, 60%, and 62% is obtained for sadness, fear, anger and happy conversions respectively, when only pitch contour alone is modified. Modification of pitch contour gave poor MOS for Anger and low ERR for anger and happiness. The modification of P + I gave an emotion recognition rate of 65% for anger, which shows intensity modification is important for the conversion of neutral speech to anger speech. The interesting result to note is that when pitch contour is combined with intensity and duration the emotion identification rate was best. The most recognized emotion was sadness with the MOS of 3.52 and an ERR of 83%. The least recognized emotion was anger with the MOS of 1.74 and ERR of only 66%. When pitch contour is combined with only the duration, sadness and fear gave the MOS of 3.2 and 2.6 with the ERR of 79% and 65% respectively. The results prove that duration plays a very important role in the perception of sadness and fear.

The MOS and ERR of the proposed method are compared with the recent techniques [17, 36, 37]. The comparison of the estimated values of MOS and ERR of the recent techniques and the proposed techniques are shown in Table V. Singh [37] has used Quadratic Multivariate Polynomial (QMP) method with Gaussian Normalization method for pitch contour modification for conversion of neutral speech sentences into target emotional (sadness, happy, anger and fear) speech sentences.

Table V. Comparison of the proposed method with other works

Author	Sad/Boredom	Happy	Anger	Fear
MOS				
Singh,	3.25	3.29	3.21	3.27
Haque	3.31	-	2.21	-
Yadav	-	2.37	2.53	2.78
Proposed technique	3.52	3.41	1.74	2.71
ERR (%)				
Singh,	18	56	51	93
Haque	84	-	53	-
Proposed technique	83	68	66	70

Haque [17], has also used Gaussian Normalization method to modify the pitch contour of a neutral speech signal to target (Sadness and anger) emotional speech signal. With the pitch contour, the epoch strength, epoch sharpness and spectral energy are modified. Yadav [36] has used the prosody imposition method at three (sentence, word, and syllable) levels. In this method, the neutral speech segments were first properly aligned with the target emotional speech segments. The process of stretching, shrinking or inserting silence between the syllable and words were used for alignment. After proper alignment, the prosody parameters of the target emotional speech segments are superimposed on the neutral speech segments.

The results of the proposed technique are compared with the results of sentence-level imposition for happy, anger and fear conversion.

From Table V it can be observed that the proposed technique is giving a better MOS and ERR for converted sadness, converted happy has almost same MOS and better ERR than that of Singh [37], and the ERR of anger is better when compared to other methods.

VII. CONCLUSION

The paper discusses the GRM method for deriving the target emotional pitch contour from the neutral pitch contour. Both static and dynamic parameters were used for analysis. The parameters used are pitch, intensity, and duration. Pitch was modified using GRM, intensity and duration were modified by the scaling factor obtained from the analysis. To test the effectiveness of GRM objective test (F0mean, $F0_{SD}$, Intensity, and duration) and subjective test (MOS and recognition of emotion) are performed. It was again proved that pitch contour plays a very important role in the conversion of neutral to sadness and anger, intensity plays an important role in the conversion of neutral to anger and duration play an important role in the conversion of neutral to fear and sadness.

The performance of neutral to sadness conversion (MOS of 3.52 & recognition of 83%) using the proposed method is found to be far superior, followed by neutral to happy (MOS of 3.41 & recognition of 68%), neutral to Fear (MOS of 2.71 & recognition of 70%) and neutral to Anger (MOS of 1.74 & recognition of 66%) with least MOS and recognition. To improve the effectiveness of neutral to happy conversion, laughter signal can be synthesized and incorporate in happy converted sentences, and this can be addressed as future work.

REFERENCES

1. J.F. Pitrelli, R. Bakis, E.M. Eide, R. Fernandez, W. Hamza, M.A. Picheny. *"The IBM expressive text-to-speech synthesis system for American English."* IEEE Transactions on Audio, Speech, and Language Processing, vol.14, No.4, pp.1099-1108, 2006.
2. O.Türk and M. Schröder. *"A comparison of voice conversion methods for transforming voice quality in emotional speech synthesis."* Ninth Annual Conference of the International Speech Communication Association. 2008.
3. L.Cen, P.Chan, M.Dong, *"Generating emotional speech from neutral speech."* 2010 7th International Symposium on Chinese Spoken Language Processing. IEEE, 2010.
4. Y.Stylianou, O.Cappé, and E.Moulines. *"Continuous probabilistic transform for voice conversion."* IEEE Transactions on speech and audio processing, Vol.6, No.2, pp.131-142, 1998.
5. M.Schröder,*"Emotional speech synthesis: A review."* Seventh European Conference on Speech Communication and Technology, in In EUROSPEECH, pp. 561-564, 2001.
6. F.Burkhardt, WF. Sendlmeier. *"Verification of acoustical correlates of emotional speech using formant-synthesis."* ISCA Tutorial and Research Workshop (ITRW) on speech and emotion. 2000.
7. A.Iida, N.Campbell, S.Iga, F.Higuchi, & M.Yasumura,*"A speech synthesis system with emotion for assisting communication."* ISCA tutorial and research workshop (ITRW) on speech and emotion. 2000.
8. J E.Cahn, *"The generation of affect in synthesized speech."* Journal of the American Voice I/O Society, Vol.8, No.1, pp. 1-1, 1990.
9. I R.Murray, & JL Arnott,*"Implementation and testing of a system for producing emotion-by-rule in synthetic speech."* Speech Communication, Vol.16, No.4. pp.369-390, 1995

10. J Y.Zhang, AW.Black, & R.Sproat. *"Identifying speakers in children's stories for speech synthesis."* Eighth European Conference on Speech Communication and Technology, In EUROSPEECH, pp.2041-2044. 2003.
11. S J L.Mozziconacci,*"Speech variability and emotion: Production and perception."* Doctoral thesis, Technische Universiteit Eindhoven, 1998.
12. J.M.Montero, JM.Gutierrez-Arriola, S.Palazuelos, E.Enriquez, S.Aguilera, & JM.Pardo,*"Emotional speech synthesis: From speech database to TTS."* Fifth International Conference on Spoken Language Processing, paper 1037, 1998.
13. I Iriondo, F Alías, J Melenchón, MA Llorca,*"Modeling and synthesizing emotional speech for Catalan text-to-speech synthesis."* Tutorial and research workshop on affective dialogue systems. Springer, Berlin, Heidelberg, pp.197-208, 2004.
14. M.Schröder, *"Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions."* Tutorial and research workshop on affective dialogue systems. Springer, Berlin, Heidelberg, pp.209-220, 2004.
15. H J Lee,*"Fairy tale storytelling system: Using both prosody and text for emotional speech synthesis."* International Conference on Hybrid Information Technology. Springer, Berlin, Heidelberg, pp.317-324, 2012.
16. NP.Nataraja,*"Intonation in four Indian languages under five emotional conditions."* Journal of AIISH, Vol.12, pp.22-27, 1981.
17. A Haque, KS Rao, *"Modification of energy spectra, epoch parameters and prosody for emotion conversion in speech."* International Journal of Speech Technology, Vol.20, No.1, pp.15-25, 2017.
18. A Jain, SS Agrawal, N Prakash, *"Transformation of Emotion Based on Acoustic Features of Intonation Patterns for Hindi Speech and their Perception."* IETE Journal of Research, Vol.57, No.4, pp.318-324, 2011.
19. M. Abe, S.Nakamura, K.Shikano & H.Kuwabara, *"Voice conversion through vector quantization."* Journal of the Acoustical Society of Japan (E), Vol.11, No.2, pp.71-76, 1990.
20. J.Tao, Y.Kang, A.Li, *"Prosody conversion from neutral speech to emotional speech."* IEEE Transactions on Audio, Speech, and Language Processing, Vol.14, No.4, pp.1145-1154, 2006.
21. Z.Inanoglu, S.Young, *"Intonation modelling and adaptation for emotional prosody generation."* International Conference on Affective Computing and Intelligent Interaction, pp.286-293, Springer, Berlin, Heidelberg, 2005.
22. E Helander, J.Nurminen, *"A novel method for prosody prediction in voice conversion."* 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Vol. 4, pp.IV-509, IEEE, 2007.
23. M.Bulut, C. Busso, S.Yildirim, A.Kazemzadeh, CM. Lee, S.Lee, S. Narayanan, *"Investigating the role of phoneme-level modifications in emotional speech resynthesis"* , Ninth European Conference on Speech Communication and Technology. 2005.
24. Y.Shao, Z.Wang, J.Han, T.Liu, *"Modifying spectral envelope to synthetically adjust voice quality and articulation parameters for emotional speech synthesis."* International Conference on Affective Computing and Intelligent Interaction, pp.334-341, Springer, Berlin, Heidelberg, 2005.
25. O.Turk, M.Schroder, *"Evaluation of expressive speech synthesis with voice conversion and copy resynthesis techniques."* IEEE Transactions on Audio, Speech, and Language Processing, Vol.18, No.5, pp.965-973, 2010.
26. H.Guo, *"A simple algorithm for fitting a Gaussian function"*, Streamkining Digital Signal Processing A Tricks of the Trade Guidebook, pp.297-305, 2012.
27. E.Pastuchová, M.Zákopčan,*"Comparison of Algorithms for Fitting a Gaussian Function used in Testing Smart Sensors."* Journal of Electrical Engineering, Vol.66, No.3, pp.178-181, 2015.
28. J A.Conder, *"Fitting multiple bell curves stably and accurately to a time series as applied to Hubbert cycles or other phenomena."* Mathematical Geosciences, Vol.47, No.6, pp.663-678, 2015.
29. Y.Yuan, *"A review of trust region algorithms for optimization."* Iciam. Vol. 99. No. 1. 2000.
30. W Chengjing, *"A trust region method with a conic model for nonlinearly constrained optimization."* Applied Mathematics-A Journal of Chinese Universities, Vol.21, No.3, pp-263-275, 2006.
31. A.Geethashree, D J. Ravi. *"Kannada Emotional Speech Database: Design, Development and Evaluation."* Proceedings of International Conference on Cognition and Recognition, pp.135-143, Springer, Singapore, 2018.
32. Govind, D., SR Mahadeva Prasanna, and Bayya Yegnanarayana. *"Neutral to target emotion conversion using source and suprasegmental information."* In Twelfth annual conference of the international speech communication association. 2011.
33. S. R. M. Prasanna and D. Govind, *"Analysis of excitation source information in emotional speech,"* in Proc. INTERSPEECH, Sep. 2010, pp. 781–784.
34. Pathak, Bageshree Sathe, Manali Sayankar, and Ashish Panat. "*Emotion transformation from neutral to 3 emotions of speech signal using DWT and adaptive filtering techniques."* In India Conference (INDICON), 2014 Annual IEEE, pp. 1-5. IEEE, 2014.
35. Rao, K. S., & Vuppala, A. K. "*Non-uniform time scale modification using instants of significant excitation and vowel onset points.*", Speech Communication, 55(6), 745–756, 2013
36. Yadav, J., & Rao, K. S. *"Prosodic mapping using neural networks for emotion conversion in Hindi language."* Circuits, Systems, and Signal Processing, 35(1), 139–162, 2016.
37. Singh, J. B., & Lehana, P, *"STRAIGHT-Based Emotion Conversion Using Quadratic Multivariate Polynomial."* Circuits, Systems, and Signal Processing, 37(5), 2179-2193, 2018

AUTHORS PROFILE

Geethashree A received the BE degree in Telecommunication Engineering in 2002 and M.Tech in VLSI Design and Embedded Systems in 2010 from Visvesvaraya Technological University, Karnataka, India, in 2002. She is currently pursuing the Ph.D. degree in ECE research center, Vidyavardhaka College of Engineering, Mysuru, India. She is working as an associate professor in the department of ECE, Vidyavardhaka College of Engineering, Mysuru, India. Her main research interest include Expressive Speech Synthesis and Speech Signal Processing.

.

D J Ravi received his BE degree in Electrical and Electronic Engineering in 1989 from the University of Mysore and M.Tech in Industrial Electronic from Sri Jaya Chamaraja Engineering college in 1992, Mysuru and Ph.D. in Speech Processing from University of Mysore, India, in the year 2013.

He is working as Dean Academic and professor in the department of electronics and communication engineering, Vidyavardhaka College of Engineering, Mysuru, India. His major research interest includes Speech synthesis and speech signal processing.