Conversion of Neutral Speech to Storytelling Style Speech

Rashmi Verma
School of Information Technology
Indian Institute Technology
Kharagpur, India
Email:rashmi9425@gmail.com

Parakrant Sarkar
School of Information Technology
Indian Institute Technology
Kharagpur, India
Email:parakrantsarkar@gmail.com

K. Sreenivasa Rao
School of Information Technology
Indian Institute Technology
Kharagpur, India
Email:ksrao@iitkgp.ac.in

***Abstract*—Speech is the most basic and widely used method for communication. There is a growing need for an expressive speech synthesis especially, when human want to communicate with robots and computers. In this paper, prosody rule-sets are designed to convert neutral to storytelling style speech for Hindi language. In order to generate a storyteller speech from a neutral speech, modification in various prosodic parameters such as pitch, intensity, duration, tempo and pauses are considered. For each of the above mentioned prosodic parameters rules are developed separately for story-specific emotions (such as sad, anger, fear, surprise and neutral). Theses rules are designed by performing analysis on stories collected from a professional storyteller. In this work, modification are done both at phrase and sentence-level. These rules are derived for both male and female speakers. Subjective tests are conducted to evaluate the quality of the generated storytelling style speech. Also, influence of speaker characteristics on neutral to story speech conversion are analysed.**

***Keywords*—Neutral speech, Storyteller Style Speech, Prosodic Parameters, Expressive Speech, Story-specific Emotions, Prosody Rule-set, Pitch, Intensity, Tempo, Content Word.**

I. Introduction

Various real life applications such as automatic PDF reading, communication with astronauts, robots and human communication, digital storytelling systems, voices of different characters in video games, has drawn attention in the research field of expressive speech synthesis. The storytelling application is one of them, in which neutral speech is converted into storyteller speech. The definition of storytelling given by National Story-telling Association [1] is: *The art of using language, vocalization, and/or physical movement and gesture to reveal the elements of a story to a specific, live audience.*

Earlier works attempted to improve the quality of speech synthesized by text-to-speech (TTS) systems by incorporating basic emotions [2]. Murray *et al.*, [3] developed a rule based system for emotional speech synthesis. Silva *et al.*, [4], obtain an expressive agent that can resemble a human storyteller. The main drawback of this system was it does not provide flexibility to express emotions. In [5] proposes a concatenative speech synthesis approach for four emotional states such as anger, happiness, sadness and neutral. Major disadvantage of this diphone base technique is the quality of the synthesized speech was not upto the mark. The Virtual Storyteller, a multi agent framework for dynamic story creation was proposed by Szilas *et al.*, [6]. A digital storyteller system, ESPER was developed by Zhang [7], to achieve all emotion expressivity and render dialogues between two characters in a story. In [8] Fairy tale storytelling system was developed that can express four emotions. Theune *et al.* [9], [14] proposed a rule based expressive speech synthesis system using Dutch text to speech system for storytelling application. In [12] prosody rules were generated for Indian languages to convert neutral TTS speech to storytelling style speech.

In this work, we are generating storytelling style speech from the neutral speech. In order to achieve this, analysis was carried out for the stories narrated by a professional storyteller and same stories narrated in neutral style like a news reader speech. The emotions present in a story are classified into five story-specific emotions namely sad, fear, anger, neutral and surprise. For each of the story-specific emotions, rules are derived by perceptual analysis of the differences in the story speech narrated by storyteller in storytelling style and neutral style. The prosodic parameters used for analysis are pitch, duration, intensity, pause patterns and tempo (speaking rate). The quality of the modified speech is evaluated by conducting emotion recognition [12] test.

This paper is structured as follows. In Section II, an overview of storyteller style speech is discussed. In Section III, the process of deriving the prosody rule-set for converting neutral to storytelling style speech. Evaluation of the quality of modified neutral speech utterance using recognition and mean opinion tests are shown in Section IV. The conclusion and summary of the whole process in described in the Section V.

II. Storyteller Speech

In this section, we discuss about the prosodic parameters which are important in a human storyteller's speech. As a starting point of our research, we have done an informal analysis by comparing the stories collected from a professional storyteller and its corresponding neutral style stories. The objective of this analysis is to examine which aspects of the prosody are pertinent in storyteller speech. Based on these prosodic parameters, a set of rules are derived that can modify neutral speech into storyteller speech. In order to analyze and manipulate speech *Praat* [10] tool is used. The rules are implemented using praat scripts [11].

978-1-4799-7458-0/15/$31.00

A. Story Speech Corpus

In this work, stories are collected in Hindi language. The total number of stories recorded from the storyteller are approximately 25 and average duration of each story is about five minutes. Most of the stories are children's stories taken from story books like Panchatantra and Akbar-Birbal. The number of sentences in a story varies from 20 to 30. The stories are recorded by both female and male professional artists. Also, the same stories are re-recoded by five different male and female speakers in neutral style like news reader speech. Both storytellers and neutral wav files are manually transcribed at phrase and sentence-level. The wav files are manually annotated with one of story-specific emotions (sad, anger, fear, neutral and surprise).

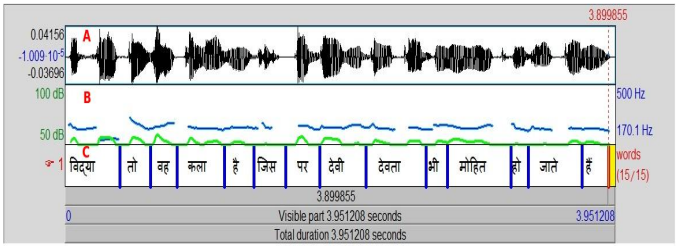

Fig. 1: Neutral speech (Utterance: "Vidhya To Wah Kala Hai Jis Par Devi Devta Bhi Mohit Ho Jate Hai.")

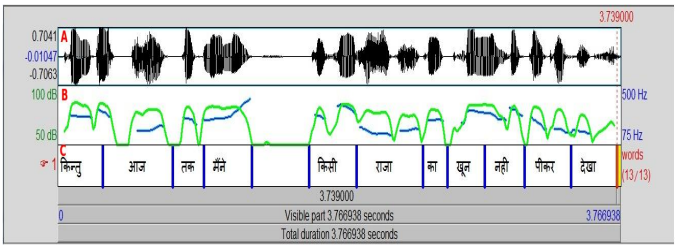

Fig. 2: Storyteller speech (Utterance: "Kintu Aaj Tak Maine Kisi Raaja Ka Khoon Nhi Pikar Dekha.")

B. Prosodic Parameters Analysis

For a female speaker, the Fig. 1 and 2, shows the waveforms of neutral and storytelling style speech respectively. In these figures the first tier (represented by the marker *"A"*) consist of the waveform of the neutral speech. The second tier (represented by the marker *"B"*) contains blue and green coloured lines which represents the pitch and intensity patterns respectively. The third tier (represented by the marker *"C"*) shows the word transcription of the utterance. Comparing both the neutral and storyteller speech waveforms, it is noted that most contributing prosodic features are the pitch, intensity, duration, tempo and pauses. These prosodic variations are more in the storyteller speech as compared to neutral speech. For instance, in the Fig. 1 and 2, it is observed that storyteller's average pitch (which is 249.79 Hz) value is larger than neutral's average pitch value (which is 187.69). Storyteller's average intensity (75.07 dB) value is larger than neutral's average intensity (56.35 dB). In case of tempo, neutral style is faster than storyteller speech for both the cases with pauses or without pauses.

C. Analysis of Story-specific Emotion

Analysis of the stories for both neutral and storyteller style are done to find out differences in terms of prosodic variations. In order to capture the attention of listeners, a storyteller exhibits various story-specific emotions while narrating a story. From our observation, most of the children stories begin with introducing the characters, followed by various events related to the character and finally concluding with a moral. In general, the beginning and ending parts of a story are mostly devoid of emotions, while the middle part of the story exhibits story specific emotions. In story corpus, we have observed five different story-specific emotions: anger, fear, sad , neutral and surprise. Apart from four emotions, surprise is an unexpected dramatic moment within a story. In case of anger emotion, the storyteller speech is louder (i.e. intensity is very high) and faster(i.e. tempo is fast). The neutral speech is less louder with slow speed. In case of fear emotion, the prosodic variations in storyteller speech are almost similar to anger emotion, but pitch is high. In case of sad emotion, the storyteller speech is having low intensity with high pitch and slow tempo. Also, in case of surprise emotion, events are announced with higher intensity and pitch on certain keywords like *"Achanak, Parantu, Kintu"* which introduces surprise in the storyteller speech.

III. Design of Prosodic Rules For Storyteller Speech

This section discusses the process involved in deriving rules for converting neutral to storytelling style (target) speech. The whole process of deriving prosody rule-set both at phrase and sentence-level is shown in Fig. 3. The rules are derived by analyzing 10 stories. Each story comprises of 20 to 30 sentences and 40 to 60 phrases. The speech utterances are analyzed perceptually for both neutral and storytelling style. A speech utterance may be considered as a phrase or sentence. For quantifying the prosodic variations, we determined the absolute difference between neutral and storytelling speech utterances. The difference is compensated by modifying the prosodic parameters of neutral speech according to the target speech. The conversion of neutral to storyteller speech is done both at sentence and phrase-level.

In order to derive a set of prosodic rules, we modified the prosodic parameters like pitch contour, duration and intensity patterns of neutral speech iteratively with a trial and error approach. For each trial, the quality of both modified and target speech utterances are evaluated perceptually. The factors used for modifications are manually adjusted based on the perception. The iteration process will be terminated whenever we are satisfied with the quality of modified speech (i.e. perceptual difference between target and modified speech utterances is minimum). These modification factors are used to generate the prosodic rule-set for five different story-specific emotions.

A. Pitch Rule

The storytelling style speech is having more variation in pitch dynamics than neutral speech. The pitch contour of storytelling style speech is either falling or raising pattern. The pitch contour of neutral style speech is not varying i.e. constant. To formulate this rise-fall patterns of the storytelling style speech

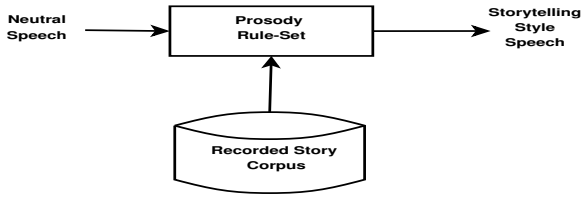

Fig. 3: Process of deriving prosody Rule-set

various non-linear functions were explored. Only sinusoidal function seems to be perfect after conducting an experimental study. Hence we formulated, the pitch contour patterns for the target speech using sine functions. The formulations based on sine function are derived for anger, fear, surprise and sad emotions. Variations among these emotions are formulated by using two variables namely *a* and *b*. The rule multiplies all pitch values within a relevant time domain [t_1,t_2] for a given utterance with a factor which is governed by the variables *a* and *b*. The sinusoidal function is shown in equation 1.

$$s'(t) = \begin{cases} s(t) * (1 + a \times sin((\frac{t-t_1}{t_2-t_1}) \times b \times pi)) & \text{if } t\epsilon[t_1, t_2] \\ y(t) & ,otherwise \end{cases} \quad (1)$$

where,
s Original pitch values at time *t*
s′ Manipulated pitch values at time *t*
a desired maximum shift in the pitch from the average pitch
b constant determining whether the sine function is constantly increasing or rising and then falling.

Table I shows the modification factors for converting neutral to storytelling speech at sentence-level for a female speaker. The values of the variables *a* and *b* are given in columns 2 and 3 of Table I for various story-specific emotions.

TABLE I: Prosodic Modification Factors for Converting Neutral to Storytelling Speech at Sentence-level (NC: No Change)

	Pitch (Hz)		Intensity (dB)	Tempo
	a	b		
Anger	-0.2	0.2	+10	0.84
Fear	0.4	0.5	+8	.86
Sad	0.4	0.2	-10	1.20
Surprise	0.8	0.75	+10	NC
Neutral	NC	NC	NC	0.95
Content Word	0.3	0.3	+5	NC

In a story, one sentence may exhibit multiple emotions. For example *"Gadadha bola, oye bhanje hum yaha chori karne aye hai."*. If this utterance is modified using sentence-level rules then only one emotion is achieved. But in original this sentence consist of having more than one emotion. Neutral emotion is present before comma where as, fear emotion is present at the later part of the sentence. To address these kind of issues and also to increase the feel of story, conversion from neutral to storyteller speech is done at phrase-level. At phrase-level, sentences are divided into phrases. Also, phrases are further divided into three parts initial, middle and final words. The values of the variable *a* and *b* derived at phrase-level for different story-specific emotions are shown in columns 2 and 3 of Table II.

B. Intensity Rule

The average intensity of a storyteller turns out to be higher as compared with neutral speech. Intensity in a speech signal may vary among different speakers. Also, intensity may vary due to the recording device used for recording the speech. In our work, we recorded the stories using a device without varying the intensity setting. The rules derived to convert intensity of neutral into storytelling style speech is given below.

$$y'(t) = \begin{cases} y(t) + c & \text{if } t\epsilon[t1, t2] \\ y(t) & ,otherwise \end{cases} \quad (2)$$

where:
y original intensity values at time *t*
y′ manipulated intensity values at time *t*
c constant
The intensity modifications at sentence-level is given in columns 4 of Table I. The intensity modifications at initial, middle and final parts of the phrase is given in column 4 of Table II.

C. Tempo Rule

Normally in a neutral style, speaker gives importance to each and every word in an utterance. It may look like the speaker is speaking all words having same duration. We analyzed a few neutral and storytelling style speech utterances and concluded that average speaking rate for neutral is slower than storytelling style. The adjustments made on the neutral speech for tempo are quite simple. Within an utterance, the duration of the signal is scaled by a modification factor. The factors at sentence-level and phrase-level (i.e. initial, middle & final words) are given in column 5 of Table I and Table II respectively.

D. Pause Rule

We analyzed two-three sentences of storytelling and neutral style speech. It is observed that stories which are recoded by storyteller have more duration of pauses as compared to neutral speech. The average pause length of neutral is 180 ms and storyteller 280 ms. At sentence-level, pauses are used for separating two sentences. At phrase-level pauses are used for separating initial, middle and final words. A rule is derived that in between sentences pause length is 350-450 ms and after a phrase pause length is 200-300 ms depending on the number of words in the phrase or sentence.

E. Neutral Emotion Rule

In a story it is not possible that all sentences or phrases consist of emotions. To get a feel of story, we need to modify these sentences which does not give any emotional evidence neither from the text nor speech. For example "eka taalaaba men: do machchha rahate the". This sentence is spoken by the storyteller in a neutral style i.e. storyteller is also not showing any emotion in it. Analyzing a few neutral emotion sentences for both neutral and storytelling style. Two main differences are observed. First, it is noticed that there are some words in a story which need more emphasis than other words. These words are termed as **content word** (for example *"taalaaba", "machchha"*). Based on this observation, we collected these set of words and used it for formulating the rules. The derived rule modifies the time interval of the content word with a factor which is given in row 8 of Table I for sentence-level modification. Second, the duration of storytelling style is less

as compared to neutral style. Hence, the rule for making this change is to decreased total duration of the utterance i.e. we are multiplying the time interval of the utterance with a factor. This increases the speed of utterance and give a feel that the narrator is speaking fast. The factor for modification at sentence-level is given in row 7 of Table I. At phrase-level, modification are carried out on the prosodic parameter tempo for initial and final parts of a phrase. The factors are given in Table II.

TABLE II: Prosodic Modification Factors for Converting Neutral Speech to Storytelling Speech at Phrase-level (NC: No Change)

	Pitch (Hz)		Intensity (dB)	Tempo
	a	b		
Anger				
Initial	NC	NC	+15	0.89
Middle	-0.2	0.2	+15	0.89
Final	0.3	0.3	+15	0.89
Fear				
Initial	NC	NC	+10	0.86
Middle	0.4	0.3	+10	0.86
Final	0.5	0.5	+10	0.86
Sad				
Initial	0.1	0.2	-10	1.20
Middle	0.4	0.3	-10	1.20
Final	0.4	0.5	-10	1.20
Surprise				
Initial	NC	NC	NC	NC
Middle	0.8	0.75	+10	NC
Final	NC	NC	NC	NC
Neutral				
Initial	NC	NC	NC	1.10
Middle	NC	NC	NC	NC
Final	NC	NC	NC	1.15

IV. Evaluation of the Prosodic Rules

Emotion recognition test is carried out to evaluate the quality of the modified speech utterances. In this test, subjects have to identify which emotion they are able to perceive in the modified speech fragments. 20 native Hindi speaking subjects in the age group of 20-35 participated in the listening test. A complete story is divided into small fragments. A fragment consists of one or two sentences, out of which one phrase is associated to one of the five story-specific emotions.

For listening test, fragments from the 10 stories which are not used for deriving the rules for both the neutral and storytelling style speech are considered. A total of 135 fragments are collected, out of which there are 50 neutral, 25 anger, 25 fear, 25 sad and 10 surprise for both the styles. The prosodic rules are applied on the neutral style speech to get modified neutral speech fragments. 135 modified neutral speech fragments are shuffled randomly and supplied to each subject for evaluation. In Table III, results are given for recognition test at sentence-level. It conveys that 70% of sad, 83.10% of surprise, 66% of fear emotion, 67.48% of anger emotion, are truly perceived as sad, surprise, fear and anger respectively. Also, results of the test at phrase-level are given in Table IV. It shows that 75% of sad, 92% of surprise, 50% of fear, 90% of anger emotions are truly perceived as sad, surprise, fear and anger respectively. Comparing the results for both phrase and sentence-level tests, it can be noted that phrase-level rules used for modification yields better result than sentence-level rules except for fear emotion as shown in the Fig. 4. The rules used for modification at phrase-level gives more story semantic information than at sentence-level. Also rules are derived at phrase-level for a male speaker. These rules are derived in a similar way as we derived the rules for a female speaker. For a male speaker, modification factors for different prosodic parameters are shown in Table V. In case of the male speaker, we also observed that the phrase-level rules outperforms the sentence-level rules.

TABLE III: Recognition Test Results at Sentence-level

	Sad (in %)	**Anger** (in %)	**Fear** (in %)	**Surprise** (in %)	**Neutral** (in %)
Sad	**70.00**	00.00	10.00	00.00	20.00
Anger	03.20	**67.48**	09.52	00.00	19.80
Fear	20.00	06.60	**66.00**	07.40	00.00
Surprise	00.00	04.80	04.50	**83.10**	07.60
Neutral	14.61	04.53	00.00	05.00	**75.86**

TABLE IV: Recognition Test Results at Phrase-level

	Sad (in %)	**Anger** (in %)	**Fear** (in %)	**Surprise** (in %)	**Neutral** (in %)
Sad	**75.00**	00.00	05.00	00.00	20.00
Anger	00.00	**90.00**	10.00	00.00	00.00
Fear	10.00	15.00	**50.00**	07.50	17.50
Surprise	00.00	03.00	05.00	**92.00**	00.00
Neutral	09.00	00.00	00.00	05.00	**86.00**

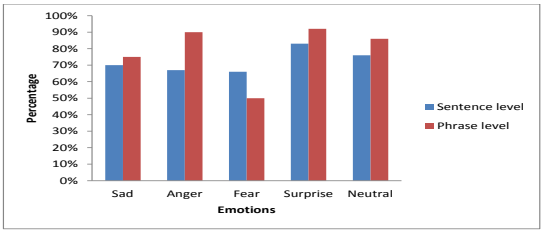

Fig. 4: Recognition test results at Phrase-level versus Sentence-level

To test the robustness and gender dependency various experiments are performed to see the influence of speaker characteristics on the neutral and storytelling style speech. To testify the robustness of the derived rules within a particular gender i.e. rules derived from one male/female speaker are tested on five different male/female speakers. Also, to test whether the phrase-level rules which uses modification factors are gender dependent or independent i.e. male speaker modification factors are tested on female speaker and vice versa. The test is performed on the 5 neutral fragments for each of the story-specific emotions taken form five different female/male speakers. For female gender, the factors derived in Table II from an individual female speaker are applied on five different female speakers. It is shown in Table VI where emotions sad and anger are perceived correctly by almost 50% listeners. Surprise is perceived correctly by 85% listeners. Fear seems to be speaker dependent at this level because it is perceived only by 20% of the listeners correctly.

Now by comparing the results obtained from the recognition test for one female speaker versus five female speakers is shown in Fig. 5. From this figure, it is noticed that less differences are present for neutral and surprise emotion as compared to other emotions when rules derived from individual female

TABLE V: Prosodic Modification Factors for Converting Neutral to Storytelling Style Speech for Male Speaker(NC: No Change)

	Pitch (Hz)		Intensity (dB)	Tempo
	a	b		
Anger				
Initial	NC	NC	+20	0.87
Middle	-0.2	0.3	+20	0.87
Final	0.3	0.4	+20	0.87
Fear				
Initial	NC	NC	+12	0.90
Middle	0.4	0.5	+12	0.90
Final	0.5	0.5	+12	0.90
Sad				
Initial	0.2	0.4	-15	1.20
Middle	0.4	0.5	-15	1.20
Final	0.5	0.5	-15	1.20
Surprise				
Initial	NC	NC	NC	NC
Middle	0.8	0.75	+10	NC
Final	NC	NC	NC	NC
Neutral				
Initial	NC	NC	NC	1.15
Middle	NC	NC	NC	NC
Final	NC	NC	NC	1.20

TABLE VI: Recognition Test Results for Individual Female Rules Tested on Multiple Females

	Sad (in %)	**Anger** (in %)	**Fear** (in %)	**Surprise** (in %)	**Neutral** (in %)
Sad	**55.00**	00.00	15.00	00.00	30.00
Anger	00.00	**45.00**	18.00	00.00	37.00
Fear	25.00	20.00	**20.00**	14.00	21.00
Surprise	00.00	05.00	00.00	**85.00**	10.00
Neutral	16.00	00.00	09.00	00.00	**75.00**

speaker is tested on five different female speakers. For male gender, the modification factors derived in Table V for an individual male are applied on five different male speakers. In the Table VII, results of the recognition test where sad and anger emotions are perceived correctly by almost 70% of the listeners. In Fig. 6, results obtained from the recognition test for one male speaker versus five male speakers is shown. From this figure, it is observed that less differences are present for fear, neutral and surprise emotions as compared to other emotions. Now from the figures 5 and 6, we observed that, variation of results evaluated on male are less compared to female speaker. Hence we can conclude that the modification factors derived for male speaker is more robust than female speaker. To perform the gender dependency test, the female

TABLE VII: Recognition Test Results for Individual Male Versus Multiple Males

	Sad (in %)	**Anger** (in %)	**Fear** (in %)	**Surprise** (in %)	**Neutral** (in %)
Sad	**60.00**	00.00	10.00	00.00	30.00
Anger	00.00	**70.00**	05.00	05.00	20.00
Fear	15.00	20.00	**30.00**	10.00	25.00
Surprise	00.00	10.00	00.00	**80.00**	10.00
Neutral	21.00	00.00	09.00	00.00	**70.00**

modification factors derived in Table II are tested on male neutral stories. The recognition test results are shown in Table VIII. From this table, it is noted that surprise is perceived by 80

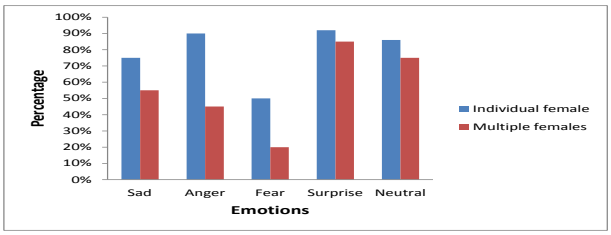

Fig. 5: Recognition test results for Individual Female versus Multiple Females

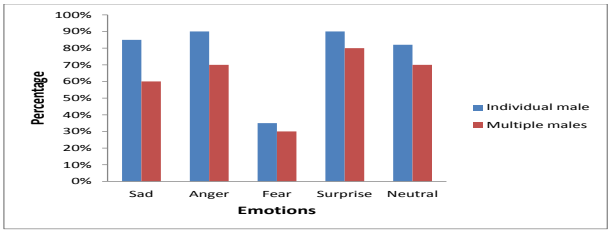

Fig. 6: Recognition test results for Individual Male versus Multiple Males

% of the listeners as compared to other emotions. Similarly, the male modification factors derived in Table V are tested on female neutral stories. The results of the recognition test carried on the modified stories are shown in Table IX. From this table it is noticed that, Fear is not perceived properly by the listeners compared to other emotions. In the Fig. 7,

TABLE VIII: Result for Female Speaker Rules Applied on Male Speaker

	Sad (in %)	**Anger** (in %)	**Fear** (in %)	**Surprise** (in %)	**Neutral** (in %)
Sad	**35.00**	00.00	00.00	00.00	65.00
Anger	00.00	**40.00**	05.00	10.00	45.00
Fear	15.00	30.00	**10.00**	00.00	45.00
Surprise	00.00	10.00	00.00	**80.00**	10.00
Neutral	16.00	00.00	09.00	00.00	**75.00**

TABLE IX: Result for Male Speaker Rules Applied on Female Speaker

	Sad (in %)	**Anger** (in %)	**Fear** (in %)	**Surprise** (in %)	**Neutral** (in %)
Sad	**45.00**	00.00	00.00	00.00	55.00
Anger	00.00	**50.00**	00.00	15.00	35.00
Fear	20.00	10.00	**05.00**	00.00	65.00
Surprise	00.00	20.00	00.00	**80.00**	00.00
Neutral	20.00	00.00	10.00	00.00	**70.00**

comparison between, test results of a female speaker and a male speaker where female modification factors are used is shown. In the Fig. 8, test results of a male speaker and a female speaker where male factors are used is shown. From the Fig. 7 and 8, it is observed that the quality and perception of story-specific emotions are improved when the male speaker modification factors are tested on female neutral stories.

We considered two neutral stories and modified them by using the phrase-level rules for both male/female speakers.

These modified stories are evaluated by the subjects in terms of naturalness on a five-point scale (1: very poor, 2: poor, 3: fair, 4: good and 5: excellent) i.e mean opinion score (MOS). The MOS for neutral and modified stories are shown in the column 2 and 3 of Table X. The significance of the differences in the pairs of the MOS for neutral and modified speech is tested using hypothesis testing [13]. The level of confidence for the observed differences in the sample means was obtained in each case using the sample variances and values of Student-t distribution. The level of confidence is high (99.5%) for both gender as shown in the Table X. This indicates that the differences in the pairs of the MOS in each case is significant.

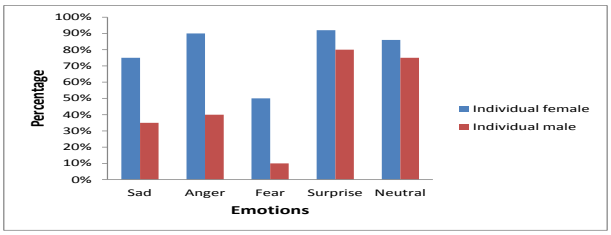

Fig. 7: Recognition test results for Individual Female rules applied on Individual Male

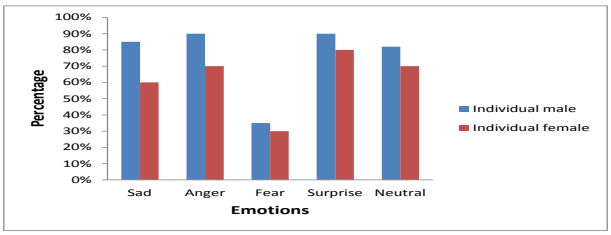

Fig. 8: Recognition test results for Individual Male rules applied on Individual Female

TABLE X: Mean Opinion Score and Confidence Values

Category	MOS		Level of confidence in % for the significance of difference in MOS
	Neutral Speech	Modified Speech	
Male	2.95	3.25	99.5
Female	2.85	3.45	99.5

V. Conclusion

This work is carried out for converting neutral to storyteller style speech. For the conversion, stories are recorded in Hindi language for both male and female voices. The importance of the prosodic parameters relevant for storytelling style speech are discussed. These parameters are considered for conversion form neutral to storytelling style speech. The conversion is performed both at sentence-level and phrase-level. At phrase-level modification yields better results than sentence-level except in the case of fear emotion. Emotion recognition tests are carried out on rules used for modification. Robustness of the derived rules for both male and female speakers are verified and concluded that male speaker rules are more robust as compared to female speaker rules. To check the gender dependency between the rules derived for male and female speaker. It is observed that quality of story-specific emotions are improved when male speaker modification factors are used on female speaker neutral stories. Further studies may be carried out to derive prosody rule-sets for various Indian languages. Prosody rules can be refined for improving the quality of storyteller speech. The existing prosody modification methods could be improved to enhance the quality of synthesized story speech.

Acknowledgements

The authors would like to thank the Department of Information Technology, Government of India, for funding the project, *Development of Text-to-Speech synthesis for Indian Languages Phase II*, Ref. no. 11(7)/2011HCC(TDIL). The authors would also like to thank all the participants for the listening tests and storytellers for recording the stories.

References

[1] National Storytelling Association, Available at http://www.eldrbarry.net/roos/art.htm,1997

[2] M. Schröder. Emotional Speech Synthesis: a review. In *Proceedings of the seventh European conference on Speech Communication and Technology*, pages 561–564, 2001.

[3] I. Murray and J. Arnott. Synthesizing Emotions in Speech: Is it time to get excited?. In *Proceedings of the fourth international conference on Spoken Language*, pages 1816–1819, IEEE, 1996.

[4] A. Silva, M. Vala and A. Paiva. The Storyteller: Building a Synthetic Character that tells Stories. In *Proceedings Workshop Multimodal Communication and Context in Embodied Agents*, pages 53–58, 2001

[5] M. Bulut, S. S. Narayanan and A. K. Syrdal. Expressive Speech Synthesis using a Concatenative Synthesizer. In *Proceedings of seventh international conference on Spoken Language Processing*, pages 1265–1268, INTERSPEECH, 2002.

[6] N. Szilas. A new Approach to Interactive Drama: From Intelligent Characters to an Intelligent Virtual Narrator. In *Proceedings of the AAAI Spring Symposium on AI and Interactive Entertainment*, pages 72–76, 2001.

[7] J. Y. Zhang, A. W. Black and R. Sproat. Identifying Speakers in Children's Stories for Speech Synthesis. In *Proceedings of the eighth European conference on Speech Communication and Technology*, pages 2041-2044, ISCA, 2003.

[8] H. Lee. Fairy Tale Storytelling System: Using Both Prosody and Text for Emotional Speech Synthesis. In *Convergence and Hybrid Information Technology*, pages 317–324, Springer, 2012.

[9] M. Theune, K. Meijs, D. Heylen and R. Ordelman. Generating Expressive Speech for Storytelling Applications. In *IEEE Transactions on Audio, Speech, and Language Processing*, pages 1137–1144, 2006.

[10] P. Boersma and D. Weenink. Praat Script and Software. Available at http://www.fon.hum.uva.nl/praat/.

[11] F. Jaeger. Praat Scripting Tutorial Basics. Available at http://www.stanford.edu/dept/linguistics/corpora/material/, 2004.

[12] P. Sarkar, A. Haque, A. Kr. Dutta, G. Reddy, Harikrishna D. M., P. Dhara, R. Verma, Narendra N.P., Sunil Kr.S.B., J. Yadav, K. S Rao Designing Prosody Rule-set for Converting Neutral TTS Speech to storytelling style speech for Indian Languages: Bengali, Hindi and Telugu, In *Proceedings of IC3*, JIIT Noida, India, page no. 473–477, 2014.

[13] Hogg, Robert V and Ledolter, Johannes *Engineering statistics*, MacMillan New York,volume 358,1987

[14] K. Meijs. Generating Natural Narrative Speech for the Virtual Storyteller. Master's thesis *University of Twente, Enschede The Netherlands*, 2004.