International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Generation of Telugu Emotional Speech from Neutral Speech by using Voice Morphing Algorithm

D .Nagaraju. RJ Ramasree, N Hari, V. Pavan Kalyan, Sk. Shameer

Abstract: ***Voice morphing is the conversion of one voice signal into another voice signal i.e. source voice signal into target voice signal. Like image morphing, the aim of speech morphing is to preserve the pooled properties of the source signal and target signal during the generation of a smooth conversion between them. The major properties of speech signal are its pitch and envelope information. Pitch, duration and intensity of speech signal parameters are considered to get emotional voice from neutral or emotion. We are using TD-PSOLA technique to extract the prosodic parameters and during the training phase.***

Key words: voice morphing- image morphing-emotional voice-TD-PSOLA.,

I. INTRODUCTION

Conversion based voice morphing converts the speaker identity by modifying the parameters of an acoustic characteristics of the speech signal. It usually consists of two parts, first one is the training procedure and second one is transformation procedure [01]. The training phase operates on examples of speech from the source signal and the target signal. The main aim of this research is to convert a speech signal in one emotion or neutral [source] to another emotion [target] [02]. Voice morphing algorithm is used to execute the above task, it works based on characteristics of speech signal i.e. pitch, duration and intensity of speech signal[02]. Voice morphing have the several applications including text to speech (TTS) , mimicry voices between different film actors and celebrities. The heart process in a voice morphing is the conversion of the spectral characteristics of the source speaker to match the target speaker.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
D.Nagaraju, Associate professor in Audisankara college of engineering and technology, gudur, Nellore, Andhra Pradesh , India -524101
Dr.RJ Ramasree, Professor, dept of computer science, Rashtriya Sanskrit vidyapeetha, tirupati, Andhra Pradesh, India
N.Hari, Students in Audisankara college of engineering and technology, gudur, Nellore, Andhra Pradesh, India
V Pavan Kalyan, Students in Audisankara college of engineering and technology, gudur, Nellore, Andhra Pradesh, India
Sk Shameer, Students in Audisankara college of engineering and technology, gudur, Nellore, Andhra Pradesh, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

The input speech signals are first analyzed to extract the pitch, duration and intensity that represent the source speech signal identity [03]. After the extraction of prosodic parameters it is trained to get the relationship between the source parameters and the corresponding target parameters.

II. RELATED STUDY

Emotional speech plays a vital role in identifying the mental state of the speaker. The proposed model discusses the usage of TD-PSOLA as an important tool in converting one emotion to another for speech samples in Telugu language.

PSOLA algorithm is used to implement the pitch and time scaling of input speech samples. They implemented emotion conversions from angry to sad and sand to angry [04]. TD-PSOLA algorithm is used to get emotional speech from neutral for Punjabi language by modifying pitch of the speech signal [05]. Male voice converted into female and vice versa based on pitch scale modification [06]. Sentimental analysis for English text to speech synthesis systems are produced by TD-PSOLA. Happy, Fear, Neutral, and sad emotions are analyzed [07].

III. SOURCE AUDIO COLLECTION

Telugu male database wav files collected from IIT Madras [08], is considered as source signal of neutral and emotional signal. Prosodic parameters are collected from the above database, database have 2840 wav files.

IV. VOICE MORPHING ALGORTHM

A voice morphing algorithm is proposed to get emotional speech from neutral speech \by using TD-PSOLA technique. Voice morphing algorithm is designed by using MATLAB [09] and Praat tool [10]. Happy, sad and angry emotions are generated from neutral voice. Prosody means desired F0, duration and stress of phone etc. Here pitch and duration, two prosodic parameters are modified to get an emotional voice from neutral voice. Original speech signal and produced speech signals are compared to get better emotional speech. Voice morphing algorithm changes only duration, and pitch parameters. The general frame for generation of emotional speech from neutral speech is as shown in figure 1.

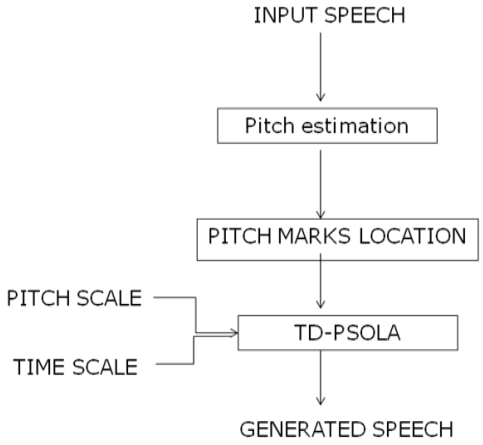

Fig 1 General frame work for voice morphing

V. PROPOSED VOICE MORPHING ALGORTHM

The proposed voice morphing algorithm is as shown in flow chart 1.

VI. EXPERMENTAL ANALYSIS

A speech sample is taken to analyze the prosodic parameters i.e. ullipaaya_ thokka _ti yaTaaniki_bledu _kaavaa laa_ katthi_kaavaalaa

ఉల్లిపాయ తొక్క తీయటానికి బ్లేడు కావాలా కత్తి కావాలా

Source and target speech synthesis prosodic parameters are given in table 1.

Table 1 Pitch and durations of source and target emotions

	Pitch	Min	Max	Intensity	Time
Neutral	196	98	352	86	3.27
Happy	245	77	501	85	3.08
Sad	216	74	479	85	6.58
Angry	259	126	497	86	1.89

Conversion of neutral speech into happy emotion

Figure 2 and 3 shows the source [neutral] and target [happy] speech signals. Neutral voice converted into happy emotional voice by changing the pitch and duration

Step1: Take two speech signals one source speech signal and other is target speech signal

Source ← neutral speech/emotional speech

Target ← emotional speech [happy, sad or angry]/ neutral speech

Step2: Found the prosodic parameters of both signals

Praat tool is used to extract pitch, duration and intensity

Step3: Change the source signal prosodic parameters

α- time scaling, β-pitch scaling factors

α_S- time scaling factor for source , α_T-time scaling factor for target

β_S- pitch scaling factor for source, β_T-pitch scaling for target

if ($\alpha_S > \alpha_T$)

then $\alpha_S = \alpha_S - 0.1$,

else $\alpha_S = \alpha_S + 0.1$

if ($\beta_S > \beta_T$)

then $\beta_S = \beta_S - 0.1$,

else $\beta_S = \beta_S + 0.1$

step4: Run the TD-PSOLA program

Mat lab is used to run TD-PSOLA algorithm

Step5: Get output wav file

Emotion.wav is the output wav file

Step6: Extract the prosodic parameters from output wav file

Extract pitch, duration and intensity of output wav file

Praat tool is used to get above parameters

Step 7: Compare the output parameters with target parameters

If ($\alpha_S = 0.95\,\alpha_T$ & $\beta_S = 0.95\beta_T$)

Consider output wav file is the target emotional wav file

step 8: Repeat step3 to step7 until wav file satisfy the target prosodic parameters.

Step 9: Stop

Flow chart 1 voice morphing

of the source. Figure 4-6 shows the iterations to convert source voice towards target voice. Table 2 gives prosodic variations from source signal into target signal.

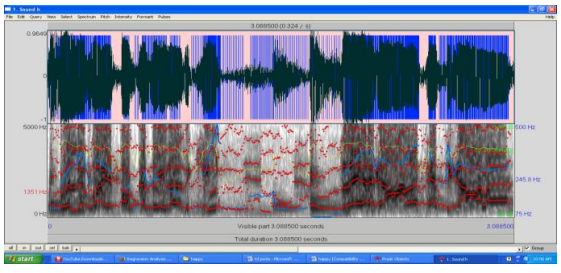

Figure 2. Target [happy] signal

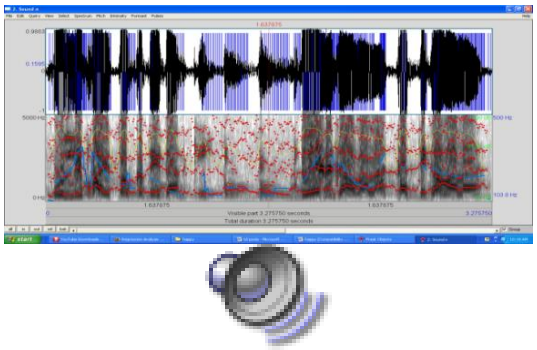

Figure 3. Source [neutral] signa
Iteration 1:

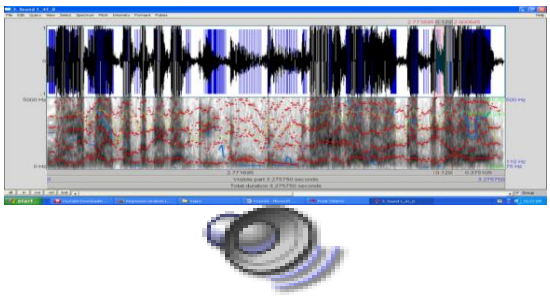

Fig 4 Iteration 1 beta=1.4 alpha=1.0
Iteration: 2

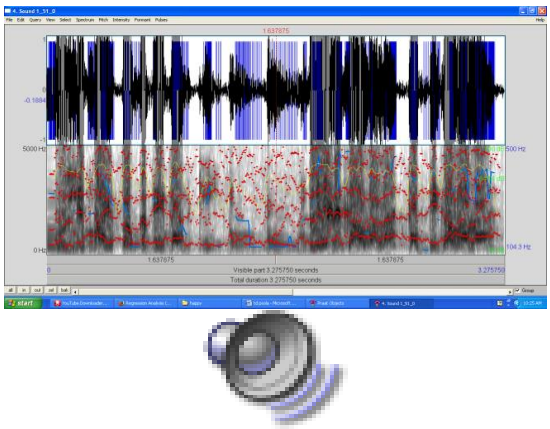

Fig 5 Iteration 2: beta=1.5 alpha=1.0
Iteration: 9

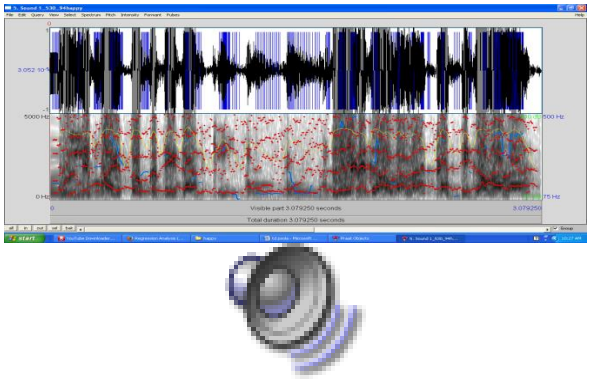

Fig 6 Iteration n: beta=1.53 alpha=0.94

Table 2. Prosodic parameters variations during the conversion of neutral to happy voice

	Beta	Pitch	Min	Max	Inten	Alpha	Time
Target speech							
Happy		245	77	501	85		3.08
Voice morphing speech							
Neutral	1.0	196	98	352	86	1.0	3.27
Iteration	1.4	233	85	482	86	1.0	3.27
Iteration	1.5	256	90	482	86	1.0	3.27
Iteration	1.5	226	79	487	86	0.9	2.94
Iteration	1.6	259	83	496	86	0.9	2.94
Iteration	1.6	270	83	497	86	0.95	3.11
Iteration	1.55	252	81	484	86	0.95	3.11
Iteration	1.55	254	81	482	86	0.93	3.04
Iteration	1.55	251	80	474	86	0.94	3.079
Iteration	1.53	243	76	488	86	0.94	3.079

Table 3 gives prosodic variations from source signal [neutral] into target signal [sad].

Similarly table 4 shows prosodic variations from source signal [neutral] into target signal [angry].

Table 3. Prosodic parameters variations during the conversion of neutral to sad

	Bet	Pitc	Mi	Ma	Int	Alp	Ti
Target speech							
Sad		216	74	479	85		6.5
Voice morphing speech							
Neutral	1.0	196	98	352	86	1.0	3.2
Iteratio	1.1	201	66	481	86	1.5	4.9
Iteratio	1.4	225	83	493	86	2.0	6.5
Iteratio	1.3	212	77	428	86	2.0	6.5

Iteratio	**1.3**	**219**	**80**	**480**	**86**	**2.0**	**6.5**
Iteratio	**1.3**	**217**	**79**	**438**	**86**	**2.0**	**6.5**

Table 4. Prosodic parameters variations during the conversion of neutral to angry

	Beta	Pitch	Min p	Max p	Intensity	Alpha	Time
Target speech							
Angry		259	126	497	86		1.89
Voice morphing speech							
Neutral	1	196	98	352	86	1	3.27
Iteration 1	1.5	244	86	479	86	0.7	2.29
Iteration 2	1.6	239	96	482	86	0.6	1.96
Iteration 3	1.7	293	117	469	86	0.55	1.8

VII. CONCLUSION

Present work focused on pitch, duration and intensity of speech signal are modified to get target speech i.e. neutral to emotional speech by using voice morphing algorithm.

REFERENCES

1. Vivek Vijay Nar "Verification of TD-PSOLA for implementing voice modification" IJERA ISSN:2248-9622 vol3 issue 3 pp 461-465.
2. D.Nagaraju, Dr.R.J.Ramasree Emo- tional speech by TD-PSOLA. GE-International Journal of Engineering Research Vol. 4, Issue 5, May 2016. 16-26.
3. Thomas Ewender and Beat Pfister "Accurate pitch marking for prosodic modifications of speech segments " in proceedings of ISCA INTERSPEECH 2010 pp 178-181
4. B. Akanksh Susmitha Vekkot Shikha Tripathi, Inter conversion of emotions in speech using TD-PSOLA , springer 2015.
5. Mamta Sharma, Payal Sharma, Implementation of TD-PSOLA to add emotions in Punjabi Speech, IJCBR 2012.
6. PV.Gujarathi and SS.Dinde, voice morphing, IJCT- 2015.
7. B Sudhakar, Different Sentiment Analysis of Text-to-Speech Synthesize Using Fuzzy Neural Network for English IJSRD 2016.
8. www.iitm.ac.in
9. http://www.tutorialspoint.com/matlab/
10. Pascal van Lieshout, Ph.D. "PRAAT", Oral Dynamics Lab V. 4.2.1, October 7, 2003.