Transformation of feelings using pitch parameter for Marathi speech

Sangramsing N. Kayte[1], Monica Mundada[1], Dr. Charansing N. Kayte[2], Dr. Bharti Gawali*

[1,3]Department of Computer Science and Information Technology Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.
[2]Department of Digital and Cyber Forensic, Aurangabad, Maharashtra

ABSTRACT
The Many researches have been done in the transformation of emotion. However, for Marathi not many studies have been done. In this paper we construct a Marathi speech database to study the effects of change of emotion. Emotion is an important element in expressive speech synthesis and is investigated by many researchers. In this research paper we build a Marathi speech database to study the effects of change of emotion. We describe methods to optimize the database for analysis and study. The pitch information is extracted from the database for different emotions like Joy, Angry and Sad. Pitch analysis is done on the database using the extracted pitch points, and a general algorithm is devised for the change of neutral state to emotional state. To perform the experiments, three expressive styles- Joy, Anger and Sad are done with Neutral.
KEYWORDS: Emotion conversion, Emotional Speech database, Joy, Anger, Sad,Pitch.

I. INTRODUCTION

Emotion plays an important role in day-to-day interpersonal human interactions. Recent findings have suggested that emotion is integral to our rational and intelligent decisions. A successful solution to this challenging problem would enable a wide range of important applications. Correct assessment of the emotional state of an individual could significantly improve quality of emerging, natural language based human-computer interfaces [2, 3, 4, 5]. It helps us to relate with each other by expressing our feelings and providing feedback. This important aspect of human interaction needs to be considered in the design of human– machine interfaces [3, 5].

To recognize emotions, we need to know not only what information a user conveys but also how it is being conveyed. Speech signals convey not only words and meanings but also emotions. In terms of acoustics, involve variation in syllable length, loudness, pitch, and the formant frequencies of speech sounds [6,7]. To build interfaces that are more in tune with the users' needs and preferences, it is essential to study how emotion modulates and enhances the verbal and nonverbal channels in human communication. Emotional speech can provide invaluable clues in forensics.

Pitch is strongly correlated with the fundamental frequency of the sound. It occupies a central place in the study of prosodic attributes as it is the perceived fundamental frequency of the sound [2,4]. It differs from the actual fundamental frequency due to overtones inherent in the sound. Pitch Contours are essentially curves that trace the pitch of the sound with time. In this paper, we extract and analyze patterns in the pitch points of utterances when spoken in the Normal, Happy, Anger and Surprise emotion states, as compared to when spoken in the Normal emotion state. The observed patterns and norms are then used to devise an algorithm for the conversion of utterances from the normal emotion state to the happy emotion state [8,9,15].

The paper is organized in five sections, the second and third section details the creation and optimization of the emotional speech database used in the analysis; the fourth and fifth section gives patterns observed in the pitch, and finally the description and implementation of the algorithm which was devised for the conversion from neutral state to emotional state.

II. EMOTIONAL SPEECH DATABASE

A database consisting of a total of 1000 Marathi words and sentences was prepared. The database used for the purpose of analysis. The first step was to record the voice of each words and sentences [14]. The recordings of all the words and sentences were done using the software Praat [9,10]. These words and sentences were recorded at a sample rate of 44100 Hz with a single (Mono) channel. The recordings were done in a silent room with minimal noise. For pitch analysis minimum and maximum pitch is taken as 75 Hz and 500 Hz.

Speakers recorded the words and sentences, with each word and sentence being spoken in four

different emotions like Neutral, Joy, Anger and Sad and then optimize it. These words and sentences are emotionally rich and can be spoken in all four other emotions. Once the speaker is in emotionally charged mood, each speaker was asked to record the words and sentences with full emotion and stored in the computer. As a part of the optimization process, the files were subject to three steps:

1) Normalization
2) Noise Reduction
3) Auto Trim

To normalize is to adjust the volume so that the loudest peak is equal to the maximum signal that can be used in digital audio. Two methods were used for mitigating the background noise from the sound one is the manual spectral subtraction was done on each sound file by hand picking the noise regions and the second is the sound was subject to a noise gate to clean up the sound. As a last part of the optimization process, the files were trimmed so that silent regions at the beginning and end of each utterance were eliminated. The resulting database had an average duration for each sentence utterance [12].

III. ANALYSIS

In our experiment we extracted all the pitch values of each word and sentences recorded in normal and emotional speech using PRATT software tool [13]. Each recorded word and sentence has three parts i.e. starting, middle and end using the fact that the pitch values should be continuous. Then we extracted the pitch values of each word and sentence from starting and ending point as well as maximum and minimum pitch value between starting and end points of particular word and sentence. But, it may joy pitch value at starting point, may be same as either maximum or minimum pitch value or same is true for end value, whether it is starting, middle or end. Each recorded word and sentence has maximum and minimum pitch. On comparing the emotional speech with the normal speech, we will see both are different. When we take Joy it has high pitch as compared to normal. Then we manually noted each pitch value of emotional speech with respect to normal speech pitch value. Then we extracted the pitch value starting point to ending point as well as maximum and minimum pitch value.

The database constructed was used to study and analyze patterns and similarities in the pitch contour when one sentence was spoken in Joy emotion, as compared to when the same sentence was spoken in the neutral emotion. For the purpose of the analysis, pitch points were computed for the utterances and a pattern was found. An algorithm for the conversion from neutral state to emotional state was implemented. The database files pertaining to neutral and emotional state were stored as sequentially numbered files so that scripting could be used to automate the process of pitch point's extraction. For computing the pitch points, the sound files were loaded in Praat, and then sent to its "Manipulation" editor. The pitch information of the utterance was extracted from that manipulation as a "Pitch-Tier" object which was used to write the pitch information to a text file. The text file was named sequentially with the same name as the sound file, with a .txt extension. The whole process was repeated for all the sound files using a script specifically written for this purpose. So after running the script, individual text files were generated which the pitch information had contained in them. Each text file had all the pitch points of its corresponding sound file. The text file had a set of tab separated values, corresponding to time and pitch. For the purpose of analysis, each utterance was partitioned into three sections Start, Mid and End part of the sentence. Also, in each part, the pitch points pertaining to the maximum pitch and the minimum pitch were computed [10,11,12].

Fig. 1 represents Graphical Pitch Tier comparison of different emotions. The table 1 given below also shows pitch values of a particular sentence in different emotions with the neutral emotion. The table 2 gives maximum, minimum, start and end value of a pitch of each sentence.
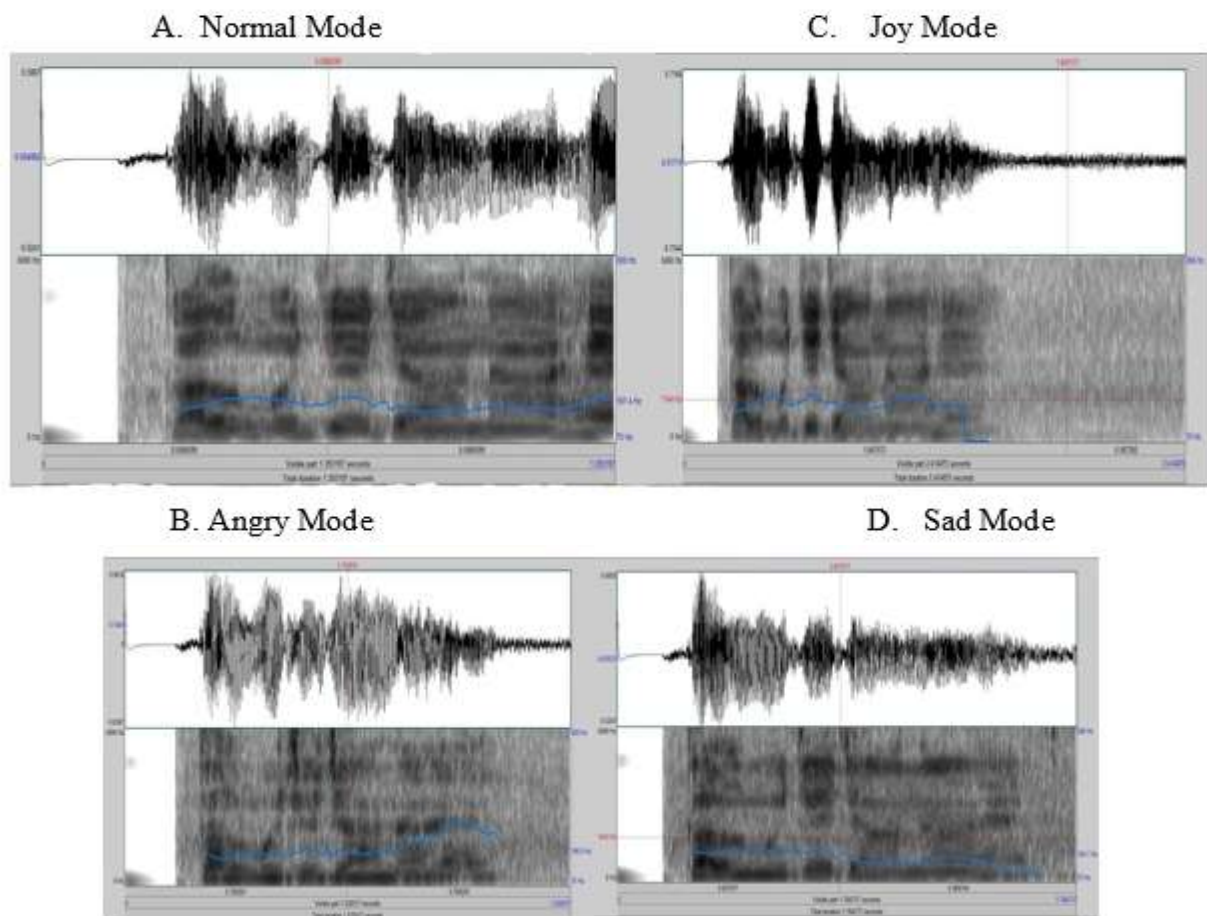
FIG 1 REPRESENTS THE GRAPHICALLY DIFFERENT EMOTIONS OF A PARTICULAR SENTENCE
"मीमाझेडोळे, घट्टबंदकेले"

Table 1 pitch values of a single sentence in different emotions [9][10]

Pitch Values	Neutral	Joy	Angry	Sad
1	165.68	190.51	163.42	144.73
2	155.55	187.37	164.45	155.67
3	152.23	182.41	167.78	154.61
4	144.96	178.44	170.52	155.89
5	142.09	179.59	173.67	147.64
6	145.78	182.85	171.34	145.86
7	149.45	185.4	174.67	151.62
8	151.04	187.82	176.23	155.78
9	156.17	190.47	178.9	159.45
10	159.04	192.75	169.67	161.59
11	162.97	195.91	177.56	165.29
12	164.68	192.75	180.49	161.17
13	160.56	195.92	183.19	169.9
14	165.87	201.06	189.62	172.21
15	168.56	202.38	186.18	165.11

Table 2 pitch values of each word on the bases of word level[9][10]

Pitch Value	मीमाझे	डोळे	घट्टबंद	केले
1	161.39	143.95	151.07	157.21
2	159.87	147.54	141.66	161.39
3	158.58	151.08	147.09	163.81
4	158.01	145.24	144.34	162.91
5	159.71	143.99	144.61	164.87
6	160.97	143.57	143.33	163.37
7	162.8	141.07	141.48	160.53
8	163.98	142.35	140.49	160.78
9	164.41	143.06	138.07	160.92
10	164.91	142.94	136.92	162.67

Table 3 pitch values of each sentence [9][10]

Sentences	Min	Max	Start	End
मी माझे डोळे, घट्ट बंद केले	126.67	174.01	124.27	170.45
आजही, ते दोघे, संगीताचे क्लास घेतात	90.37	159.23	105.23	155.89
म्हणाली, माका सत्तर मैलार, जांवचा आसा.	116.35	173.95	116.35	178.82
पुल,उत्कृष्ट संगीतकार होते.	98.94	160.33	110.02	170.78
डॉक्टरांनी मेंदूचे स्कॅन करून घ्यायला सांगितले.	120.67	170.21	118.27	169.67
बिस्किटे नाष्टा, अशी येथे पद्धत नाही.	109.88	156.12	125.11	148.99
शंकरराव भोईं, स्मृती प्रतिष्ठानने, घेतली आहे.	80.27	155.01	102.56	154.78
शिवसेनेस, कमजोर करण्याचे काम, काँग्रेस पक्ष करू शकला नाही.	105.66	177.88	112.45	171.39
शिवसेनेचे एकत्रीकरण करणे, असे ज्या, चळवळ्यांना वाटते.	101.56	161.98	115.78	159.14

AN ALGORITHM ON THE BASIS OF WORD
Here I define an algorithm which is used in this analysis process.

ALGORITHM
1) Select sound waveform and convert into pitch tier
2) Apply Stylization - Reduce pitch value
3) Find Start, End, Min and Max of each word.
4) Increase Start, End and Min By constant factor
5) Decrease Start, End and Max By constant factor

ALGORITHM ON THE BASIS OF SENTENCE
Here I define an algorithm which is used in this analysis process.

ALGORITHM
6) Select sound waveform and convert into pitch tier
7) Extract the pitch points for modification
8) Divide the sentence into Start, Middle and End
9) Apply Stylization - Reduce pitch value
10) Increase Pitch values of all pitch points by constant factor
11) Find pitch points to alter Start, End, Min and Max of each sentence.
12) Modify Start, End and Min By constant factor
13) Replace pitch Tier for resynthesize of new pitch

IV. CONCLUSION
This research paper presented an analysis of different emotions based on pitch contour with the goal of finding the emotion from the neutral emotion. An algorithm for emotion conversion been described

which consists of simple rules for converting pitch points. The alignment of pitch value by linguistic rules has not been considered, the future work will focus on the linguistic rules for emotion conversion. But the proposed system has scope for further refinements. The pitch factor has been considered for our experiment; the effect of other factors like Spectrum, Duration etc can be further investigated. The experiment has been performed on small utterances and not adequate in terms of numbers. The Marathi database should be enhanced to achieve the perfectness.

REFERENCES

[1.] Sangramsing N.kayte "Marathi Isolated-Word Automatic Speech Recognition System based on Vector Quantization (VQ) approach" 101th Indian Science Congress Jammu University 03th Feb to 07 Feb 2014

[2.] Monica Mundada, Sangramsing Kayte, Dr. Bharti Gawali "Classification of Fluent and Dysfluent Speech Using KNN Classifier" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 9, September 2014

[3.] Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" International Journal of Computer Science and Information Technologies (IJCSIT)

[4.] Monica Mundada, Sangramsing Kayte "Classification of speech and its related fluency disorders Using KNN" ISSN2231-0096 Volume-4 Number-3 Sept 2014

[5.] 5)Amit Sharma, Frederic Kaplan. Pierre Dillenbourg, "Use of Prosody features for Conversational Analysis"

[6.] Sangramsing Kayte, Monica Mundada "Study of Marathi Phones for Synthesis of Marathi Speech from Text" International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-10) October 2015.

[7.] Carlo Drioli, G.Tisato, P.Cosi, and Fabio Tesser, "Emotions and Voice Quality: Experiments with Sinusoidal Modeling"

[8.] Rong J, Li G, Chen Y-P P. Acoustic feature selection for automatic emotion recognition from speech. J Inf Proc Manag, 2009,

[9.] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte " Performance Calculation of Speech Synthesis Methods for Hindi language IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 13-19e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[10.] Sangramsing Kayte, Monica Mundada, Santosh Gaikwad, Bharti Gawali "PERFORMANCE EVALUATION OF SPEECH SYNTHESIS TECHNIQUES FOR ENGLISH LANGUAGE " International Congress on Information and Communication Technology 9-10 October, 2015

[11.] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis Systems for Marathi Language" OSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 5, Ver. I (Sep –Oct. 2015), PP 76-81e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197 www.iosrjournals.org

[12.] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis System for Hindi" International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015

[13.] Paul Boersma and David Weenink. (2009, November) Praat: doing phonetics by computer. [Online]. HYPERLINK "http://www.fon.hum.uva.nl/praat/"

[14.] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Implementation of Marathi Language Speech Databases for Large Dictionary" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 40-45e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[15.] Sangramsing Kayte, Dr. Bharti Gawali "Marathi Speech Synthesis: A review" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 6 3708 – 3711