Development of Isolated Marathi Words Emotional Speech Database

V. B. Waghmare
Research Student, Department of Computer Science & IT, Aurangabad (MS), India

R. R. Deshmukh
Professor, Department of Computer Science & IT, Aurangabad (MS), India

P. P. Shrishrimal
Research Student, Department of Computer Science & IT, Aurangabad (MS), India

G. B. Janvale
Assistant Professor, SCIT, Symbiosis International University, Pune, India (M.S.)

ABSTRACT
The research work describes the procedure and the development of Isolated Marathi Emotional Speech Database. The database consists of samples, collected from 50 speakers including males and females who simulated the emotions producing by the Marathi utterances which are used in everyday communication and are interpretable in all applied emotions. The speech samples were enhanced by spectral subtraction method and distinguished by the various real life situations. The recorded speech samples were categorized in three basic categories i.e. Happy, Sad and Angry.

General Terms
Speech Signal Processing, Speech Recognition, Emotion Recognition, Spectral Subtraction, Emotional Speech Database.

Keywords
Speech Databases, Emotion, Speech, Emotion Recognition, Human Computer Interaction.

1. INTRODUCTION
Speech is the most effective and common way of communication between humans. Human beings have long been motivated to create computer system that can understand and talk like humans. In this direction, researchers are trying to develop systems which can analyze, classify and recognize the speech signals [1]. However, the emotion expressed by speech is one of the major influencing factors for the low recognition accuracy achieved during the development of speech based systems.

The emotions are expressed by humans through speech and various actions like crying, yelling, dancing, laughing, stamping, and many more ways. However, when it comes to speech human emotions affects the tone and the speaking style of the person. The researchers around the globe are now trying for detection of emotions in Speech. In human computer interaction (HCI), many researchers are exploring the depth in the area of emotion detection from speech. The emotional speech differs from the normal speech in terms of its pitch, loudness, timbre, speech rate, and pauses. The designing and development of emotional speech databases is one of the major challenges for the researchers who are working in the area of emotion recognition and studying the effect of emotions on speech recognition or speech synthesis system. The emotional speech database can help in the development of robust automatic speech recognition (ASR) and for robotics. The emotional speech database might be helpful to overcome the challenge of development of various robust speech applications whether it is speech recognition system, speech synthesis system or development of new interfaces using speech. For Indian languages the work has not yet reached to the level where it can be used as real communication tool, as a lot of work has already been done in other languages of developed countries; thus the focus of this work is on Marathi language [2].

This paper gives the details about the development of isolated emotional Marathi speech database. The database was developed after the studying various available emotional speech databases in different languages [3].

The paper describes the procedure followed for the development of isolated Marathi emotional speech database. Section II describes what is meant by emotions. In section III, we have discussed the selection of emotion for the development of the database. Section IV describes the selection of isolated words from the selected emotions. Section V describes the speech data collection; the recording procedure followed for the development Isolated Marathi Emotional Speech Database is explained in section VI. Sections VII describes the enhancement technique used for the removal of any background noise from the collected samples and section VIII and IX gives the conclusion and the future work respectively.

2. EMOTIONS
The most important thing while talking about emotional speech is what is meant by emotion? A number of definitions of emotions have been proposed since1884 when William James first tried to define or give the answer of it. The emotion has been defined as “an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism” [4]. The natural emotion means the emotion which are expressed spontaneously when a series of event occur to which the brain responses accordingly. The artificial or acted emotions means to mimic the natural emotions which are similar to those emotions expressed without the occurrence of events to which the brain responses spontaneously.

There are many arguments over the selection of natural or real emotion and acted or artificial emotion. To capture the natural emotions it is very difficult as these emotions are responses to the internal or external stimulus received by the brain. No one can predict how the brain of different person will react to an event so it becomes difficult to capture the natural emotions and their classification. We did require emotions but due to the above mentioned problem we were unable to capture the natural emotion so we developed database of acted or

artificial emotions which are mimicked. We developed the artificial emotional Marathi speech database and performed the experiment for emotion recognition on the developed database [5, 6].

3. CHOICE OF EMOTIONS

The selection of emotions is the most important aspect while doing research in the domain of emotion recognition. The emotions can be classified in 5 different types i.e. conceptions, sensations, reflexes, involuntary expressions and voluntary expressions. The emotions that come under the category of conceptions, sensations, reflexes and involuntary expressions are biologically adapted. These emotions are transmitted to the human through reproduction or in other words by hereditary characteristics. The voluntary types of expressions are due to cultural adaptions; such types of emotions are transmitted or evolved through interaction and day to day experiences which are faced during the life span [7].

The emotions selected for the development of the said database was on the basis of most commonly observed emotions and those which were earlier used by the researchers working in the field of emotion recognition. We selected the three Basic and most commonly observed emotions i.e. Happy, Sad and Angry. Another reason for the selection of these emotions is keeping in mind about the comparison of our research results with the result of earlier research work carried by the researchers around the globe which will help us to create a benchmark for the research in emotion recognition for Marathi language. The emotion recognition research conducted for different languages have also used the above said three emotions along with neutral, afraid and surprise.

4. SELECTION OF WORDS FOR SELECTED EMOTIONS

Selection of isolated words for capturing the specified emotions is a very important aspect while developing the emotional speech database. Another problem while working for Marathi language is its dual meaning nature. A same word or sentence can have different meaning in different scenarios. It was a major concern that was taken care of while selecting the words; a single word can be spoken in different scenarios and it can express different emotions.

For the development of the isolated words emotional speech database we developed a word sets for the selected basic three emotions i.e. happy, sad and angry emotions. The words were selected after observing the people around us for a long duration and the way how the emotions were expressed using single or isolated words in real life scenarios.

Finally eight words expressing each selected basic emotions were selected from a list of more than fifty words. For the selection we did considered the dual meaning criteria to avoid the confusion that may arise for categorizing the emotions into their respective category.

The table I, II and III represents the words selected for the development of the isolated word emotional speech database along with the respective transliterations and IPA (International Phonetic Alphabet).

Table I: Happy words in Marathi Language along with Transliteration and IPA

Devanagari	Transliterated (Translated In English)	IPA
अरेव्वा	Arewaa (Oh Good)	/ ərevʊa /
कितीछान	Kiti Chan (How good)	/ kiʈi/ʧʰəaɳə /
कितीगोड	Kiti Goad (How Sweet)	/ kiʈi/gəodɻə/
मस्त	Mast (Good)	/məsəʈə/
खतरनाक	Khatarnaak (Fantastic)	/kʰəʈərəɳəakə/
ह्याट रे	Hyaat re (Wow)	/həjəaʈərae/
जबरदस्त	Jabardast (Very Good)	/dʒəbərəɖəsəʈə/
व्वाव्वा	Waa waa (Wow)	/ʊvaʊvə/

Table II: Sad words in Marathi Language along with Transliteration and IPA

Devanagari	Transliterated (Translated In English)	IPA
आरे देवा	Are Dewa (Oh God)	/əre/ /ɖevə/
अरेरे	Arere (Ohh)	/ ərere/
अबब	Ababa (Ohh)	/əbəbə/
हे काय झाल	He kaay zal (What Happen)	/he//kajə//dʒʱəalə/
आई गं	Aai ga (Oh mother)	/ai//gə/
धत्तेरेकी	Dhattereki	/ɖʱəʈəerəeki/
जाऊदे रे	Jau de re (Let it be)	/dʒəauɖəe//rəe/
नाही ग जमत	Nahi ga jamat (Its not possible)	/ɳəahəi//gə/dʒəməʈə/

Table III: Angry words in Marathi Language along with Transliteration and IPA

Devanagari	Transliterated (Translated In English)	IPA
गप रे	Gapp re (Just Shutup)	/gəpə/ /rəe/
आयला	Aayla ()	/ajələə/
च्यामारी	Chyamari ()	/ʧəjəaməarəi/
चल निघ	Chal Nigh (Get out)	/ʧələ/ /ɳəigʱə/
व्हय्घर्री	Vhay Ghari (Go to your home)	/ ʊəhəjəgʱərəi/
हट्ट	Hutt (Leave me alone)	/həʈʈə/
मुस्काड फोडीन	Muskaad Fodin (Will Slap your face)	/məusəkaɖə//pʰəoɖəiɳə/
नकोय मला	Nakoy Mala (I Don't Want)	/ɳəkəojə/ /mələ/

5. SPEECH DATA COLLECTION

In this section, the steps followed for developing speech database is described. Firstly, the recording media was chosen to capture the speech signal. The Data samples were recorded using a two different microphones and laptop using Praat software for recording the speech signals.

5.1 Speakers Selection

The speech data was collected from the native speakers of Marathi language. The selected speakers were resident of Marathwada region of Maharashtra state.

5.2 Data Collection

The speakers were asked to speak total 24 words in the three basic emotions with 3 utterances of every word. The speech data was collected from people belonging to the Marathwada regions. The speakers were selected on the basis of the educational qualification and their native language

5.3 Data Collection Statistics

We collected speech samples from 50 speakers. The 50 speakers were categorized according to the gender. We have collected the data from 25 males and 25 females in the age group of 21 to 40. The database consists of 150 utterances of each word selected from the list of isolated words. The database consists of in all 3600 utterances of 24 emotional words in Marathi language.

6. RECORDING PROCEDURE FOLLOWED

The selected words were recorded from speakers using two different headsets. The headsets used were Sennheiser PC350 and PC360 which are different in terms of technical specifications. The reason for selecting these specific headsets was that they are having noise cancelling facility. The distance of the both the microphone from speaker was same.

The data was recorded in normal environment. We used PRAAT software for recording the speech samples. The main strength of PRAAT is its graphical user interface, the functionalities like spectral analysis, pitch analysis, formant analysis, intensity analysis, other functionalities for drawing the Cochleagram, Spectrogram, speech signal plots and most important that it is open source.

The speakers were asked to pronounce the each word while keeping in mind the emotion in which the word is categorized. The recording was carried until the speaker does not pronounce the word correctly in the proper emotional scene. The recording sessions usually lasted for 30 minutes as mimicking the emotions correctly was difficult.

In some scenarios we discussed with them there past events in which they expressed the selected emotions so that we can record the correct utterances in exact emotional state.

The speech data was recorded with a sampling frequency of 16000 Hz, 16 bit in mono audio format. The files were saved with .wav (dot wav) file extension. As the data was recorded in a normal environment the recorded samples consisted of background noise which was later enhanced.

7. SPEECH SIGNAL ENHANCEMENT

The recorded speech contained some background noise. The noisy speech samples were processed for the removal of noise. The speech signal enhancement is very important before we can extract the feature for developing the recognition system. The isolated Marathi emotional speech database needs to have good speech samples without background noise. The speech samples should have good quality and good intelligibility for increased recognition accuracy. There are various speech signal enhancement techniques available like: spectral subtraction, subspace based algorithm, adaptive filtering technique (like LMS algorithm, RLS algorithm, Kalman filter) and adaptive comb filtering.

The spectral subtraction speech signal enhancement technique was implemented while enhancing the speech samples for the developed speech database.

In Spectral subtraction, an average signal spectrum and average noise spectrum are estimated in parts of the recording and subtracted from each other, so that average signal-to-noise ratio (SNR) is improved [8]. It is assumed that the signal is distorted by a wide-band, stationary, additive noise, the noise estimate is the same during the analysis and the restoration and the phase is the same in the original and restored signal.

In the signal domain the model is described by equation 1:

$$y(n) = x(n) + d(n) \qquad (1)$$

Where, 'x' is the speech signal,'d' is the noise and 'y' noisy speech.
In the frequency domain the noisy speech model equation is expressed in 2:

$$y(j\omega) = x(j\omega) + d(j\omega) \qquad (2)$$

Where, y(jω), x(jω) and d(jω) are the Fourier transforms of the noisy signal y (n), x (n) and d (n) respectively.
As the statistic parameters of the noise are not known, thus the noise and the speech signals are replaced by equation 3:

$$\hat{x}(j\omega) = y(j\omega) - \hat{d}(j\omega) \qquad (3)$$

The Noise spectrum estimate $\hat{d}(j\omega)$ is related to the expected noise spectrum E [$|\hat{d}(j\omega)|$] which is usually calculated using the time-averaged noise spectrum $\hat{d}(j\omega)$ taken from parts of the recording where only noise is present. The noise estimate is given by equation 4:

$$\hat{d}(j\omega) = E\left[\left|d(j\omega)\right|\right] \cong \left|\bar{d}(j\omega)\right| = \frac{1}{K}\sum_{i=0}^{K-1}\left|d_i(j\omega)\right| \qquad (4)$$

Where $d_i(j\omega)$ is the amplitude spectrum of the ith frame of the K frames of noise. Noise estimate in kth frame may be obtained by filtering the noise using the first order low-pass filter of equation 5:

$$\hat{d}(j\omega) = \left|\tilde{d}_k(j\omega)\right| = \lambda_n \cdot \left|\tilde{d}_{k-1}(j\omega)\right| + (1-\lambda_n)\cdot\left|d_k(j\omega)\right| \qquad (5)$$

Where $\tilde{d}_k(j\omega)$ the smoothed noise estimate in the ith frame, λ_n is the filtering coefficient. To obtain the noise estimate, the part of the recording containing only noise that precedes the part of containing speech signal needs to be analyzed [9].

The enhanced speech samples after the removal of background noise using Spectral subtraction were stored separately. The original copy of the speech samples were retained after obtaining the noise free speech samples.

The figure 1 (a) represents the waveform of the speech sample for happy emotion having background noise and figure 1 (b) represents the spectrogram for the speech sample for happy emotion having background noise. The figure 2 (a) represents

the waveform of the speech sample for happy emotion after removal of the background noise and figures 2 (b) represents the spectrogram for the speech sample for happy emotion after removal of the background noise.

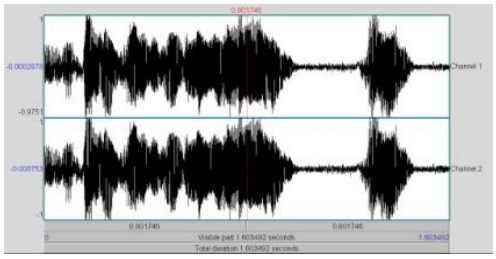

Fig. 1 (a) Waveform of the speech sample for happy emotion having background noise

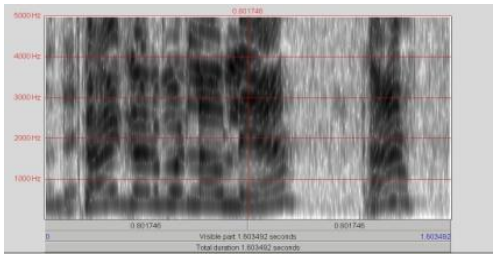

Fig. 1 (b) Spectrogram of speech sample for happy emotion with background noise

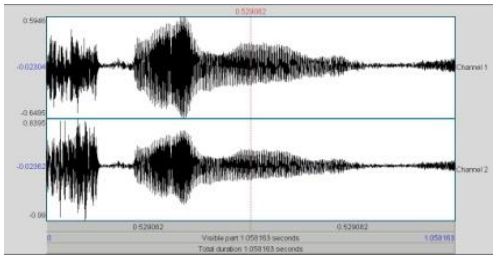

Fig. 2 (a) Waveform of speech sample for happy emotion without background noise

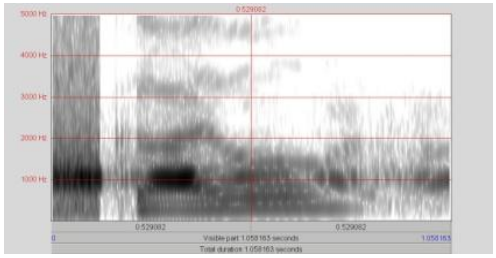

Fig. 2 (b) Spectrogram of speech sample for happy emotion without background noise.

8. CONCLUSION

In this paper, we presented the steps followed by us to develop an isolated word emotional Marathi speech database. The developed isolated emotional words speech database is first of its type for Marathi language. The developed speech database would help the researchers willing to work in the field on emotion recognition from speech in Marathi language. The developed database can also help in the development of robust automatic speech recognition system which can understand the utterances spoken in different emotional states

9. FUTURE WORK

The database will be used to develop emotion recognition system from speech for Marathi language and compare the results of the emotion recognition with the results of researchers who developed and worked on various emotional speech databases in different language. To standardize and increase the size of the emotional speech database for Marathi language and to correctly recognize the emotions from speech in Marathi language

10. ACKNOWLEDGMENTS

The authors would like to thank the University Authorities for providing the infrastructure to carry out the research. This work is supported by University Grants Commission, New Delhi as Major research project

11. REFERENCES

[1] Pukhraj Shrishrimal, R. R. Deshmukh, Vishal Waghmare, (2012, July) “Indian Language Speech Database: A Review". International Journal of Computer Application (IJCA) Vol 47, No.5 pp.17-21

[2] Ganesh B. Janvale, Vishal Waghmare, Vijay Kale and Ajit Ghodke "Recognition of Marathi Isolated Spoken Words Using Interpolation and DTW Techniques" ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol I Advances in Intelligent Systems and Computing Volume 248, 2014, pp 21-29

[3] Vishal B Waghmare, Ratnadeep R Deshmukh, Pukhraj P Shrishrimal (2012, July) "A Comparative Study of the Various Emotional Speech Databases". International Journal on Computer Science and Engineering, Vol 4, issue 6, pp. 1236-40

[4] Klaus R. Scherer, "What are emotions? And how can they be measured?" (2005) Trends and developments: research on emotions, Social Science Information Vol 44 – no 4, pp. 695–729.

[5] Vishal B Waghmare, Ratnadeep R Deshmukh (2014, February) “Development of Artificial Marathi Emotional Speech Database” in proceeding of 101st Indian Science Congress, Jammu, Indian

[6] Vishal B. Waghmare, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal, Ganesh B. Janvale, “Emotion Recognition System from Artificial Marathi Speech using MFCC and LDA Techniques”, Fifth International Conference on Advances in Communication, Network, and Computing – CNC 2014 (22-23 Feb 2014), Organized by ACEEE, Chennai, pp

[7] Wagner J.,Jonghwa Kim, Andre E. (2005) "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification" IEEE International Conference On Multimedia & Expo, Amsterdam. pp. 940 - 943.

[8] Philipos C. Loizou, “Speech Enhancement: Theory and Practice”, CRC Press. June 7, 2007.

[9] Saeed V. Vaseghi, “Advanced Digital Signal Processing and Noise Reduction”, Second edition, John Wiley & sons Ltd, pp. 333-354.