Emotion Speech Recognition System For Isolated Words In Marathi Language

Asmita Gajarishi[1], Shital Sukre[2], Dhanashri Rachetti[3], Pratik Kurzekar[4]

Department of Computer Engineering, Institute of Knowledge College of Engineering, Pune-412208 (MS) India
(Email: asmitagajarishi1996@gmail.com)[1]
(Email: shitalsukre1994@gmail.com)[2]
(Email : arpitapatil91@gmail.com)[3]
(Email :pkkurzekar@iokcoe.ac.in)[4]

Abstract: *Recognition of emotions from speech is one of the most important sub domains in the field of affective computing. Sometimes, a person speaks the sentence while stay in some emotion which makes the tone of speech changes the meaning of the sentence completely. Speech signal consist not only the words and meaning but also it consist of the emotions. 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. When it comes to human speech emotions affects the tone and the speaking style of the person. The research in this area is needed to overcome these problems of emotion recognition from speech. However the problem is usually deals with the following basic emotion categories: Happy, Sad, Angry, Afraid, Surprise, Neutral. From the literature survey for the proposed study, it is observed that there is no proper emotional speech corpus in any of the Indian languages for carrying out the research on emotional speech processing. No any standard emotional speech database as well as the real life emotional speech database is available in the context of Indian languages. It is also observed from the literature that excitation source information is not thoroughly investigated for the purpose of emotion recognition task. Most of the researchers have used frame-wise spectral features extracted from entire utterance for speech emotion classification. Most of the existing emotion recognition systems are developed using only gross prosodic features extracted from the entire utterances. This paper would be helpful for the researchers to find the brief overview of emotion speech recognition systems developed in different languages around the world and the purpose and approach of the research.*

Keywords: Speech recognition system, Emotion recognition system, Types of speech recognition system, Speech Database.

I. INTRODUCTION

Speech is the most commonly and widely used form of communication between humans. There are various spoken languages which are used throughout the world. The communication among the human being is mostly done by vocally, therefore it is natural for people to expect speech interfaces with computer [1]. Since early 1960's researchers are trying to develop system which can record, interpret and understand human speech. The use of speech for interacting with the computer may help the developing nations as the language technologies can be implemented for the e-governance system. Speech recognition (SR) means translation of spoken words to the text or commands. Development of Speech recognition systems has attained new heights but robustness and noise tolerant recognition systems are few of the problems which make speech recognition systems inconvenient to use [2]. Many Research projects have been completed and currently in progress around the world for the development of robust speech recognition systems. The paper presents the review of the continuous speech recognition systems.

II. CLASSIFICATION OF SPEECH RECOGNITION SYSTEMS:

The speech recognition systems can be classified in different type depending upon different classes. The speech recognition system can be classified on the basis of type of utterances, vocabulary size and speaker dependency.

A. Classification on the basis of utterances:

1) Isolated Words:

Isolated word recognizers usually require each utterance to have quite on both sides of the sample window. It accepts single words or single utterance at a time. These systems have "Listen/Not-Listen states", where they require the speaker to wait between utterances.

2) Connected Words:

Connected word systems are similar to isolated words, but it allows separate utterances to be 'run-together' with a minimal pause between them.

3) Continuous Speech:

Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. Recognizers with continuous speech capabilities are some of the most difficult to create because they utilize special methods to determine the utterance boundaries.

4) Spontaneous Speech:

At a basic level, it can be thought of as a speech that is natural sounding and not rehearsed. An ASR system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being run together, "ums" and "ahs", and even slight stutters [3].

B. Classification on the basis of Vocabulary size:

1) Small Vocabulary:

The speech recognition systems which can recognize limited and given set of vocabulary (i.e. few hundred words or sentences) are known as limited vocabulary speech recognition system.

2) Medium Vocabulary:

The speech recognition system which can recognize a considerable number of vocabularies (i.e. few from few hundred up to few thousands of words or sentences) such systems are known as medium vocabulary speech recognition system.

3) Large Vocabulary:

The speech recognition system which can recognize a large number of vocabularies (i.e. more than few thousands of words or sentences) such systems are known as large vocabulary speech recognition system [4].

C. Classification on the basis of Speaker mode:

1) Speaker Dependent:

Speaker dependent speech recognition systems learn the unique characteristics of a single person's voice, in a way similar to voice recognition. The system is trained on the basis of the training dataset and it may use templates.

2) Speaker Independent:

In speaker-independent speech recognition systems there is no training of the system to recognize a particular speaker and so the stored word patterns must be representative of the collection of speakers expected to use the system. The word templates are derived by first obtaining a large number of sample patterns from a cross-section of talkers of different sex, age-group and dialect, and then clustering these to form a representative pattern for each word.

3) Speaker Adaptive:

In speaker adaptive speech recognition systems the uses the speaker dependent data and adapt to the best suited speaker to recognize the speech and decrease the error rate by adaption [5].

III. EMOTION RECOGNITION SYSTEM

Emotion recognition systems focus on modeling of spectral as well as the prosodic features such as formants, pitch, loudness, timbre, speech rate and pauses which contain the linguistic and semantic information [6]. However the problem usually deals with the following basic emotion categories: Happy, Sad, Angry, Afraid, Surprise, Neutral.

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. When it comes to speech human emotions affects the tone and the speaking style of the person. The research in this area is needed to overcome these problems of emotion recognition from speech [7].

Speech emotion is one of the important things in human expression. A number of definitions of emotions have been proposed from1884 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" [8]. 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 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 emotion and there classification. The research required emotional speech but due to the above mentioned problem we were unable to capture the natural emotion so we developed databases 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 [9, 10].

A. Artificial Emotion

The artificial emotions are the emotions which are not real but they mimicked or acted by an individual. Natural emotions are responses to the internal or external stimulus to the brain. Any person will react to an event so it becomes the natural emotion. The problem is that it's not at all possible to capture the natural emotions so for the study the best source was the acted or artificial emotions which are mimicked.

Everyone cannot mimic the emotion accurately however; the professional actors and actress can mimic the emotions to most probable real emotion [11]. The researchers have tried to recognize the emotions from movies, TV shows, plays and other sources where a person mimics the emotions. In movies, TV Shows, Plays and other sources the actors and actress performs the act and the complete the emotional tasks so it gets easy to carry out the emotion recognition process on such type of data.

B. Natural Emotion

The human speech contains the information and some meaning which relates the emotion which contains not only the linguistic content but also contains some emotions of the speaker even though the emotion does not alter the linguistic content. There are many arguments over the selection of natural or real 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 emotions.

C. Real Life Emotions

In real life human can express their emotion by many ways such as by facial expression, by yelling, by touch. Speech is one of the important outcomes of the emotional state of human beings. A speech signal is produced from the contribution of the vocal tract system excited by excitation source signal.

D. Whispered Emotional Speech

The whispered speech can also carry emotional information like prosodic features, including short time energy and speak rate, voice quality parameters, formant, and spectrum to analyze the differences between emotions. One could perceive others feelings at that moment when someone whispered. Nowadays, with the widespread of the cellular phone, people whisper so as to reduce the amount of speech being spell out; to public safety, whispered speech are often encountered for criminal analysis; and to laryngectomees, whisper is the only means of articulation. With the help of acoustic features like endpoint detecting, abstraction of formant frequencies and the corresponding bandwidths the whispered speech signal can be measured. [12]

E. Mood Extraction

Mood extraction from speech is one of the difficult tasks. The Sentiment analysis (SA) plays a vital role in natural language processing. The sentimental analysis can be done with the help of tasks to classify the different moods such as Happy, Sad, Frustrated, Angry, Depressed, Temper etc for domain-specific sentence-level mood extraction [13].

IV. LITERATURE REVIEW OF EMOTION RECOGNITION FROM SPEECH

A. Emotional Speech Databases for English Language

During the study we observed that maximum work for emotion recognition from speech has been carried out for English language. We studied various emotional speech databases developed in the English language. The Table 1 shows the work done so far for emotion recognition for English and the developed Emotional Speech corpora.

Table 1 English Emotional Speech Corpora

Sr. No.	Emotion	Number of Speakers	Type of Database	Purpose and approach
01	Depression and neutral [14]	22 patient and 19 healthy persons	Simulated	Recognition. Prosody variations are analyzed with respect to the speech samples of depressed and healthy people
02	Anger, disgust, fear, joy, neutral, sadness and surprise [15]	8 actors (2 per language)	Simulated	Synthesis. Emotional speech is recorded in 4 languages (English, Slovenian, Spanish, and French).

03	Anger, boredom, joy and surprise [16]	51 children	Elicited	Recognition. Recorded at the university of Maribor, in German and English.
04	Anger, fear, happiness, neutral and sadness [17]	40 native speakers	Natural	Recognition Two broad domains of emotions are proposed based on prosodic features.
05	Different natural emotions [18]	125 TV artists	Natural	Recognition. It is known as Belfast natural database and is used for several emotion processing

Since 1996 the work for emotion recognition from speech for English languages has been started. Different English emotional speech corpora's are developed with different parameters and with the different emotional states. Two simulated type of emotional speech databases are developed in English language in which the recorded emotional speech samples are collected by few patients facing the problem of depression and few healthy persons with two emotional categories like depression for patients and neutral for the healthy person. The developed databases was used for experiment in which the prosody variations were used to analyze recognition task with respect to the speech samples of depressed and healthy speakers and amplitude modulation, formants, power distribution were used to analyze depressed and suicidal speech respectively. The synthesis was carried out for 7 emotions like Anger, disgust, fear, joy, neutral, sadness and surprise.

A database in English language has been developed in which 8 actors simulated the emotions. The database development work was carried out in four different languages and similar work has been done for 128 TV actors simulating different natural emotions, the developed database is named as Belfast Natural Database which has been used for several emotion processing applications.

B. German Emotional Speech Corpora

In European languages, German language is difficult to understand by the machine. The researchers have focused for work on Emotional Speech recognition for German. The researchers found that the emotion recognition for German language is very difficult task. Table 2 shows six databases that are developed in German language.

Table 2 German Emotional Speech Corpora:

Sr. No.	Emotion	Number of Speakers	Type of Database	Purpose and approach

01	Anger, Boredom, disgust, fear, joy, neutral and sad [19]	10 Actors	Simulated	Synthesis
02	Different elicited emotions [20]	51 School children (21M+30 F)	Elicited	Recognition. Children are asked to spontaneously react with Sony AIBO pet robot. Around 9 & 1/2 hours of effective emotional expressions of children were recorded
03	Anger, Boredom, disgust, fear, joy, neutral and sad [21]	10 Actors (5M+5F)	Simulated	Recognition. About 800 utterances are recorded using 10 neutral German sentences.
04	Soft, modal and loud [22]	Single Actor	Simulated	Synthesis. Di-phone based approach is used for emotional speech synthesis
05	Anger, Boredom, disgust and worry [23]	6 Native Speakers	Simulated	Recognition. Affective bursts and short

| | | | | emotional non – speech segments are analyzed for discriminating the emotions. |

Since 2000 the emotional speech database development for German language has been started. At very first the simulated type of database was developed in which the recordings were taken from 10 Germany based actors simulating the emotions like anger, boredom, disgust, fear, joy, neutral and sad; later the same researcher has extended their database with another 800 utterances recorded by 10 speakers 5 male and 5 female actors/actresses using 10 neutral German sentences.

In 2003, a new type of simulated emotional speech database was developed in German language for di-phone based approach in which three emotional categories were taken like soft, model and loud which were uttered by a single actor.

In 2006 one elicited type of database was developed in which the different elicited emotions were recorded from 51 School children, 21 male and 30 female children speakers. Children were asked to spontaneously react with Sony AIBO pet robot. Around 9&1/2 hours of effective emotional expressions of children have been recorded.

C. Chinese Emotional Speech Corpora

In Chinese language the emotional speech database development started in 2001, till date 4 databases are developed in Chinese language. In Chinese language the phonation are somewhat difficult to understand by the machine. Table 3 shows the detail information about the developed Chinese emotional speech databases along with their types, purpose and approaches.

Table 3 Chinese Emotional Speech Corpora

Sr. No.	Emotion	Number of Speakers	Type of Database	Purpose and approach
01	Antipathy, anger, fear, happiness, sad and Surprise [24].	Two actors	Simulated	Recognition.
02	Anger, disgust, fear, joy, sad and Surprise, 60 Utterances per emotion per speaker were recorded	12 Actors	Simulated	Recognition. Log frequency power coefficients are used for emotion recognition using HMMs.

	[25]			
03	Anger, happiness, neutral and sad, 721 short utterances per emotion were Recorded [26]	Native TV actors	Simulated	Recognition.
04	Anger, fear, joy, neutral and sad, 288 sentences per emotion were recorded [27]	9 Native speakers	Elicited	Recognition. Phonation, articulation and prosody are used to classify 4 emotions.

Three artificial emotional speech databases were developed in Chinese. At the very first in 2001, a simulated type of emotional speech database was developed in Chinese language. The database was developed for Native TV speakers expressing 4 basic emotions like anger, happiness, neutral, and sad. The developed database consisted of 721 short utterances per emotion. In 2003, another simulated type of database was developed for 12 actors expressing 6 emotions like anger, disgust, fear, joy, sad and surprise with 60 Utterances per emotion per speaker were recorded. On this database Hidden Markov Model (HMM) was implemented by using Log frequency power coefficients for detection of Chinese speech emotion.

The third simulated type of emotional speech database was developed in 2006, for 2 Chinese actors consisting speech samples with antipathy, anger, fear, happiness, sad and surprise emotional categories.

Only one elicited type of emotional speech database is developed in Chinese for native speakers for which the speech samples were recorded from 9 native Chinese speakers with anger, fear, joy, neutral and sad emotions, 288 sentences per emotions were recorded and the classification of 4 emotions was done by using phonation, articulation and prosody features which plays a vital role in the recognition of Chinese emotional speech.

D. Spanish Emotional Speech Corpora

In Spanish language only two simulated type of emotional speech databases are developed. The database development started in 1999. Table 4 shows the details of two Spanish emotional speech databases.

Table 4 Spanish Emotional Speech Corpora

Sr. No.	Emotion	Number of Speakers	Type of Database	Purpose and approach
01	Desire, disgust, fear, fury (anger), joy, sadness, and surprise [28].	8 Actors (4M+4F)	Simulated	Synthesis. Acoustic modeling of Spanish emotions is studied. Rules are used to identify significant behavior of emotional parameters.
02	Anger, disgust, happiness, and sadness, 2000 phones per emotion are considered [29]	Single actor	Simulated	Synthesis. Pitch, tempo, and stress are used for emotion synthesis

In 1999 a simulated type of emotional speech database in Spanish language was developed by using a single actor simulating 4 emotions like anger, disgust, happiness and sadness, 2000 phones per emotion were considered in the database. The features like pitch, tempo and stress were used for emotion synthesis.

Later in the year 2000, another simulated type of database was developed for 8 actors/actresses in which 4 male and 4 female speakers were expressing the 7 emotions like desire, disgust, fear, fury (anger), joy, sadness, and surprise. Emotional speech synthesis was carried out by doing acoustic modeling of Spanish emotions. Rules were used to identify significant behavior of emotional parameters.

F. Current Status of Indian Languages in Speech Emotion domain:

As we have seen from above sections, significant efforts have been made in the development of emotional speech databases all over the world in different languages. Most of the researchers have used standard databases for their study. Many databases have been developed in different languages like English, Russian, Japanese, German, Deutsch, Swedish, French and others; apart from this as it comes to Indian languages only few emotional speech databases has been designed and developed till date. In this section we look over those emotional speech databases and the work done so far by using those databases [30].

At Indian Institute of Technology (IIT) Kharagpur lot of work is going on for emotion recognition in Indian languages. Three major emotional speech corpuses has been designed and developed for Indian language.

In 2011 the researchers from Indian Institute of Technology (IIT) Kharagpur has developed the simulated emotion speech corpus for Hindi language. The speech corpus is named as Indian Institute of Technology Kharagpur Simulated Emotion Hindi Speech Corpus (IITKGP-SEHSC).

The database is recorded using professional artists from Gyanavani FM radio station, Varanasi, India. The speech corpus is collected by simulating eight different emotions using neutral emotions which were emotion free text prompts. The emotions taken in the database are anger, disgust, fear, happy, neutral, sad, and sarcastic and surprise. Emotion classification has been performed on the IITKGP-SEHSC using prosodic and spectral features. Mel frequency cepstral coefficients (MFCCs) were also used to represent spectral information. Energy, pitch and duration were used to represent prosody information. The average emotion recognition performance using prosodic and spectral features were found to be around 77% and 81% for female speech utterances. The quality of the emotions expressed in the database was evaluated using subjective listening tests. The emotion recognition

performance using subjective listening tests was observed to be around 74%. The results of subjective listening tests are grossly on par with the results obtained using prosodic analysis of the database [31].

Later in the year 2012, a new semi natural database was introduced by IIT Kharagpur for categorizing speech emotions. This semi natural database is named as Graphic Era University Semi Natural Emotion Speech Corpus (GEU-SNESC). The database was collected from the dialogs of Hindi movies and the expressions of emotions which are close to the real and practical situations. The conversations in movies are generally acceptable by listener as natural as he/she can easily categorize these emotions by listening dialogs spoken by the speaker.

This database was collected from Hindi movies for single (male+female) and multi (male+female) speakers. The emotions recorded in this database are sadness, anger, happiness and neutral. For single speaker, one popular male/female actor/actress and for multi-speaker, multiple male and female actors/actresses were considered. Male and female dialogs are separately extracted from movies to collect desired emotions. For each emotion, different clips were collected from various Hindi movies and then combined to form a single file of 15 minutes. From the databases, 70% of the

data was used for training the emotion recognition model and 30% was used for testing the trained models. Initially the audio was extracted from the video with the help of Adobe Audition, in which the sampling rate of 16 KHz and mono channel with 16 bit resolution was chosen. After that different speech sentences without background music were separated carefully to be a part of the database based on the contextual emotion present. Then these sentences were edited with the help of a software tool known as wave surfer for removing the longer silence regions between the words without affecting the emotions [32].

In the same year 2012 the same researchers have extended their work and they have developed Simulated Telugu emotion speech database (IITKGP-SESC) which was developed in Telugu language [33]. They extracted various features and tried various classification methods on the developed database. Average emotion recognition performance was calculated for male and female speaker which was observed to be around 65.3% and 72% respectively.

V. CONCLUSIONS

We discussed the literature review conducted for the research in the area of emotion recognition from speech. The study was carried out for the various emotional speech databases developed, the study of systems using Prosodic

features for emotion recognition and lastly the research conducted for Indian language. It was observed during the study the majority of the work in being carried out for English language. The first attempt of work for the English language was done in 1996.

During the study it was observed that 16 databases are available for English language. Similarly the work for development of emotional speech database and recognition/analysis of the databases has also been done for German, Italian, Chinese, Japanese, French, Swedish and other languages.

The work being conducted for Indian languages is limited to Hindi and Telugu languages. No attempts were made for the development of emotional speech database for any other Indian languages. The ignorance for the other Indian languages motivated for carrying the research for emotion recognition from Marathi speech.

ACKNOWLEDGMENT

REFERENCES

[1] Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal," Continuous Speech Recognition System: A Review", Asian Journal of Computer Science and Information Technology 4: 6 pp. 62 – 66,2014.

[2] Chao Huang, Eric Chang, Tao Chen "Accent Issues in Large Vocabulary Continuous Speech Recognition (LVCSR)", Microsoft Research China, MSR-TR-2001-69, pp.1-27,2001.

[3] Santosh K. Gaikwad, Bharti Gawli, Pravin Yannawar, "A Review of Speech Recognition Technique", International Journal of Computer Applications (0975–8887) Volume 10, No.3, November 2010.

[4] M. A. Anusuya, S. K. Katti, "Speech Recognition by Machine: A Review", International Journal of Computer Science and Information Security (IJCSIS), Vol. 6, No. 3, pp. 181-205, 2009.

[5] X. D. Huang, "A Study on Speaker - Adaptive Speech Recognition", Proc. DARPA Workshop on Speech and Natural Language, pp. 278-283, February 1991.

[6] 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

[7] Yu Zhou, Yanqing Sun, Lin Yang, Yonghong Yan, "Applying articulatory features to speech emotion recognition", 2009 International Conference on Research Challenges in Computer Science, 978-0-7695-3927-0/09, IEEE 2009

[8] Ganesh Janvale, Vishal Waghmare, Vijay Kale and Ajit Ghodke, "Recognition of Marathi Isolated Spoken words Using Interpolation and DTW techniques", ICT and critical infrastructure: proceeding of the 48th Annual of Computer Society of India Vol I. Advances in Intelligent system 3-319-03107-1_3, Print ISBN 978-3-319-031066 Online ISBN 978-3-319-03107-1, January 2014.

[9] 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

[10] 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.

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

[12] Gong Chenghui, Zhao Heming, Zou Wei, Wang Yanlei, Wang Min, "Preliminary Study on Emotions of Chinese Whispered Speech" International Forum on Computer Science-Technology and Applications, 978-0-7695-3930-0/09, IEEE 2009 pp. 429 – 433.

[13] Neethu Mohandas, Janardhanan P. S. Nair, Govindaru V., "Domain Specific Sentence Level Mood Extraction from Malayalam Text" 2012

International Conference on Advances in Computing and Communications IEEE 2012 pp 78-81.

[14] D. C. Ambrus, "Collecting and recording of an emotional speech database". Tech.rep. Faculty of Electrical Engineering, Institute of Electronics, Univ. of Maribor.(2000)

[15] M. Alpert, E. R. Pouget R. R. Silva "Reflections of depression in acoustic measures of the patient's speech", Journal of Affective Disorders, 66, 59–69,(2001)

[16] A. Batliner, C. Hacker, S. Steidl, E. Noth, D. S. Archy, M. Russell, M.Wong "You stupid tin box children interacting with the Aibo robot: a cross-linguistic emotional speech corpus" In Proc. language resources and evaluation (LREC 04),Lisbon. (2004).

[17] R. Cowie, E. Douglas-Cowie, "Automatic statistical analysis of the signal and prosodic signs of emotion in speech" In Fourth international conference on spoken language processing ICSLP 96, Philadelphia, PA, USA, October 1996 (pp. 1989–1992). (1996).

[18] R. Cowie, R. R. Cornelius "Describing the emotional states that are expressed in speech" Speech Communication, 40, pp 5–32 (2003).

[19] F. Burkhardt, W. F. Sendlmeier "Verification of acoustical correlates of emotional speech using formant synthesis" In ITRW on speech and emotion, Newcastle,Northern Ireland, UK, Sept. 2000 (pp. 151–156). (2000)

[20] A. Batliner, S. Biersacky, S. Steidl "The prosody of pet robot directed speech: Evidence from children" In Speech prosody 2006, Dresden (pp. 1–4). (2006)

[21] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss "A database of German emotional speech" In Interspeech. (2005)

[22] M. Schroder, M. Grice "Expressing vocal effort in concatenative synthesis" In International conference on phonetic sciences ICPhS 03, Barcelona (2003).

[23] M. Schroder "Experimental study of affect bursts" Speech Communication, 40(1– 2). Special issue on speech and emotion (2003)

[24] C. H. Wu, Z. J. Chuang, Y. C. Lin "Emotion recognition from text using semantic labels and separable mixture models" ACM transactions on Asian language information processing (TALIP), 5, 165–182 (2006).

[25] T. L. Nwe, S. W. Foo, L. C. D'Silva, "Speech emotion recognition using hidden Markov models". Speech Communication, Vol. 41, 603–623, 2003.

[26] F. Yu, E. Chang, Y. Xu, H. Y. Shum "Emotion detection from speech to enrich multimedia content" In Proc. IEEE Pacific Rim conference on multimedia, Beijing (pp. 550–557) (2001).

[27] J. Yuan, L. Shen, F. Chen "The acoustic realization of anger, fear, joy and sadness in Chinese" In International conference on spoken language processing (ICSLP 02), Denver, Colorado, USA, Sept. 2002 (pp. 2025–2028) (2002).

[28] I. Iriondo, R. Guaus, A. Rodrguez, P. Lzaro, N. Montoya, J. M. Blanco, D.Bernadas, J. M. Oliver, D. Tena, L. Longhi "Validation of an acoustical modeling of emotional expression in Spanish using speech synthesis techniques" In ITRW on speech and emotion, New Castle, Northern Ireland, UK, Sept. 2000.

[29] J. M. Montro, J. Gutterrez-Arriola, J. Colas, E. Enriquez, J. M. Pardo "Analysis and modeling of emotional speech in Spanish" In Proc. int. conf. on phonetic sciences (pp. 957–960) (1999).

[30] Aditya Bihar Kandali, Aurobinda Routray, Tapan Kumar Basu, "Emotion Recognition from Speeches of Some Native Languages of Assam Independent of Text and Speaker", National Seminar on Devices, Circuits & Communication (NASDEC2-08), November, 2008

[31] Koolagudi S.G., Reddy, R., Yadav, J., Rao, K.S. "IITKGP-SEHSC: Hindi Speech Corpus for Emotion Analysis", IEEE International Conference on Devices and Communications (ICDeCom) pp 1 - 5, 2011

[32] Shashidhar G. Koolagudi, Swati Devliyal, Anurag Barthwal and K. Sreenivasa Rao "Emotion   Recognition from Semi Natural Speech Using Artificial Neural Networks and Excitation Source Features", Springer-Verlag Berlin Heidelberg, pp. 273–282, 2012.

[33] Shashidhar G. Koolagudi, Anurag Barthwal, Swati Devliyal and K. Sreenivasa Rao "Real Life Emotion Classification from Speech Using Gaussian Mixture Models", Springer-Verlag Berlin Heidelberg pp. 250–261,2012.