Speech Emotion Recognition In Marathi Language Using Deep Belief Network For Robotics Application

[1]Ms.M.D.Kawade, [2]Mr.A.S.Ufade

[1]Assistant Professor, [2]Assistant Professor
[1]Information Technology, SNJB's LS KBJ COE, Chandwad,Nashik,India
[2]Electronics and Telecommunication, MET,IOE BKC,Nashik, India.

Abstract : For a **Robot** to Plan their actions autonomously and interact with people, recognizing their emotion is crucial & required. Emotions have always been known to play a very important and complex part in human behavior. For affective human-robot interaction, In the proposed method, emotion will be classified into Common or Basic classes: Angry, sad, happy, & neutral .Some **uncommon classes** are: Antipathy (feeling of intense dislike), approved, attention, prohibition. There are some important Research concern in Speech Emotion Recognitions are the term Emotion itself is Uncertain and subjective, Emotion is an individual mental state that arises spontaneously rather than through unaware efforts .there are no standard speech corpora for comparing performance of research approaches used to recognize emotions. Ideally SER should be robust to process real life & noise speeches. Feature extraction from speech and processing are possible in many different ways. Here we will use particular feature which will be suitable for one particular language .ideally speech feature should be independent of language and personality. But various researches show that different people have different way to show reactions. The speech processing involves three main steps i.e. pre-processing, feature extraction and pattern recognition. Speech emotion recognition is nothing but an application of the pattern recognition system in which patterns of derived speech features such as Pitch, Energy, MFCC are mapped using classifier.

IndexTerms **- Speech Emotion Recognitions,Communication,Mood.**

–

I. INTRODUCTION

Speech is most important mode of communication in human being. Apart from information sharing speech also convey information about human emotional state. Mood and emotion are two different words in psychology .As per literatures more than 140 emotions are available and out of this count researchers have work only on few common class or basic emotions such as Happy, Angry, sad, neutral etc. The speech signal is the fastest and the most natural method of communication between humans. This fact has motivated researchers to think of speech as a fast and efficient method of interaction between human and machine. The task of speech emotion recognition is very challenging for the following reasons. First, it is not clear which speech features are most powerful in distinguishing between emotions. The acoustic variability introduced by the existence of different sentences, speakers, speaking styles, and speaking rates adds another obstacle because these properties directly affect most of the common extracted speech features such as pitch, and energy contours [1].

People have been speaking to each other for thousands of years. In recent time, Human -machine interaction (HMI) has become a growing area of innovation in industry as well as academic field [2]. Speech is one of the fundamental ways of communication known to mankind. A speech signal is a logical arrangement of sounds. Our brain performs a complex set of analyses of auditory input (i.e. sounds). It converts the sounds into some conceptual ideas and thoughts which forms the basis of instructions, commands, information & entertainment.

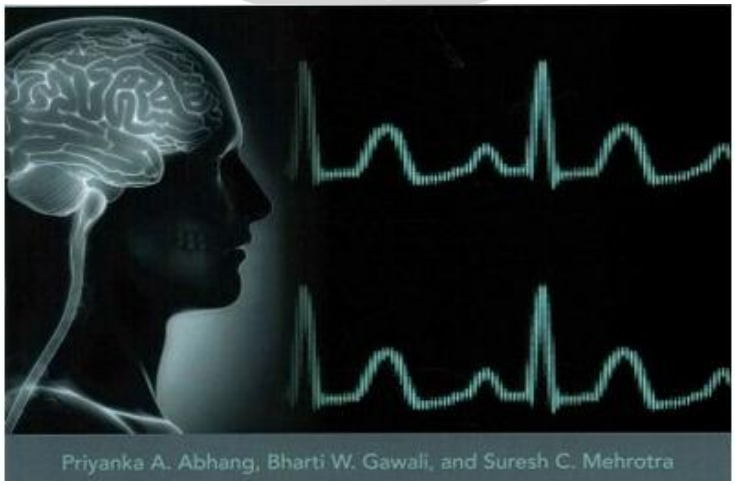

Figure 1: View of Speech Produce

Automatic recognition is often studied in sense of identifying emotion among some fixed set of classes. Speech emotion recognition is a kind of analyzing vocal behavior. The speech processing involves three main steps i.e. preprocessing, feature extraction and pattern recognition. In case of speech signal, vowels carry the most of the informative part. Vowels are mainly voiced part of the spoken words. Therefore it is customary to separate out voiced part from unvoiced part of the information spoken and proceed further with signal processing on only voiced part.

For an effective and natural HMI, emotion recognition plays a vital role. Emotions reflect the mental state of the person through speech, facial expressions, body postures and gestures and also other physical parameters like body temperature, blood pressure, muscle action, etc. The mental state of the person indirectly affects the speech produced by the person. E.g. in human-human interaction, speech rate is faster in case of anger/ joy and pitch range is also wider while in case of sadness, speech is slower with Lower pitch range. Therefore, emotion detection in speech is advantageous in various applications [9].

Challenges Ahead (Motivation);

- Feature set : it is not clear which speech features are most powerful in distinguishing between emotions.
- Representation: There may be more than one perceived emotion in the same utterance .it is very difficult to determine the boundaries between these portions.
- Speaker and culture dependencies : a certain emotions expressed is highly depend on speaker and his or her culture, personality, environment etc. (different people have different style to show emotions.)

II. LITERATURE SURVEY

There is hug scope for work on emotion recognitions in regional languages as very few researchers have attempted this. Marathi is one regional language which is spoken by majority of group. In India this 4th highest spoken language after Hindi, Bengali, Telugu.

Various researches contributed through their research on Language specific emotion extraction & recognitions. E.g. Since Chinese is a tonal language and features proper to English are not always applied to mandarin.[4].by Bu chen .In Marathi Language we have total 13

Vowels and 36 consonants. An important issue in the design of a speech emotion recognition system is the extraction of suitable features that efficiently characterize different emotions. Since pattern recognition techniques are rarely independent of the problem domain, it is believed that a proper selection of features significantly affects the classification performance. Four issues must be considered in feature extraction. The first issue is the region of analysis used for feature extraction. While some researchers follow the

ordinary framework of dividing the speech signal into small intervals, called frames , from each which alocal feature vector is extracted, other researchers prefer to extract global statics from the whole speech utterance. Another important question is what the best feature types for this task are, e.g. pitch, energy, zero crossing, etc.? A third question is what is the effect of ordinary speech processing such as post-filtering and silence removal on the overall performance of the classifier? Finally, whether it suffices to use acoustic features for modeling emotions or if it is necessary to combine them with other types of features such as linguistic, discourse information, or facial features.

Key Point from Literature Survey:

- Speech emotional library is the foundation of Speech Emotion Recognition.(SER).
- Speech databases are available in different languages but not officially reported in any Indian languages.
- Quality of feature extraction is directly proportional to accuracy of SER.
- From the studied literature we can comment that Deep Belief network is proved better for SER.
- Researchers do not agree on the parameters of speech responsible for emotion.[8]
- Acted emotions does not match with spontaneous emotions.
- Most of the researchers have worked on common Emotions like anger , happy, neutral , sad.

III. PROPOSED SYSTEM

Feature extraction from speech and processing are possible in many different ways. Here we will use particular feature which will be suitable for one particular language .ideally speech feature should be independent of language and personality. But various researches show that different people have different way to show reactions. Hence first step in this research will be to find particular language specific feature identification. Once this is done second stage here will be to build the database for specific language. Development of database is another important challenge in this work. Available databases are there in different languages and number of reported databases in required language is not officially reported by any organization or group. third stage will be feature extraction from developed database .the fundamental or base of this entire work is mainly depend on two important things one is good database second is selection and quality of feature Extraction .once features will be extracted Deep networks will be trained with the features. Now in testing phase the input speech signal will be applied to the system and system will detect the emotion from the speech. various features used in various available literature are MFCC, pitch, Energy, Glottal Velocity volumes and many more .Speech corpora used for Developing emotional speech can be mainly divided in to 3 types acted/simulated ES Database, Elicited/induced ES database, natural

Emotional Speech Database.

The speech processing involves three main steps i.e. preprocessing, feature extraction and pattern recognition.

Speech emotion recognition is nothing but an application of the pattern recognition system in which patterns of derived speech features such as Pitch, Energy, MFCC are mapped using classifier.

The speech processing involves three main steps i.e. pre-processing, feature extraction and pattern recognition. Speech emotion recognition is nothing but an application of the pattern recognition system in which patterns of derived speech features such as Pitch, Energy, MFCC are mapped using classifier. Different combinations of emotional features give different emotion detection rate. The researchers are still debating for what features influence the recognition of emotion in speech. From the studied literature
We can conclude that Deep Belief network is proved better for SER. Quality of feature extraction is directly proportional to accuracy of SER .Speech emotional library is the foundation of SER. Speech databases are available in different languages but not officially reported in any Indian languages.

Let's consider we have input database D contains a set of Speech signal s which belongs to E number of emotions. Every Emotion is with represented by W number of speech signal and every signal is represented by with a length V.

$s \in \{e_i ; 1 \leq i \leq E\}$ ----------1.1

$e_{i= s_n^i; 1 \leq r \leq W}$----------1.2

The Problem is to find the emotional class e of the input signal s to find the state of the person.

Technique:

- The speech processing involves three main steps i.e. preprocessing, feature extraction and pattern recognition.
- Speech emotion recognition is nothing but an application of the pattern recognition system in which patterns of derived speech features such as Pitch, Energy, MFCC are mapped using classifier.

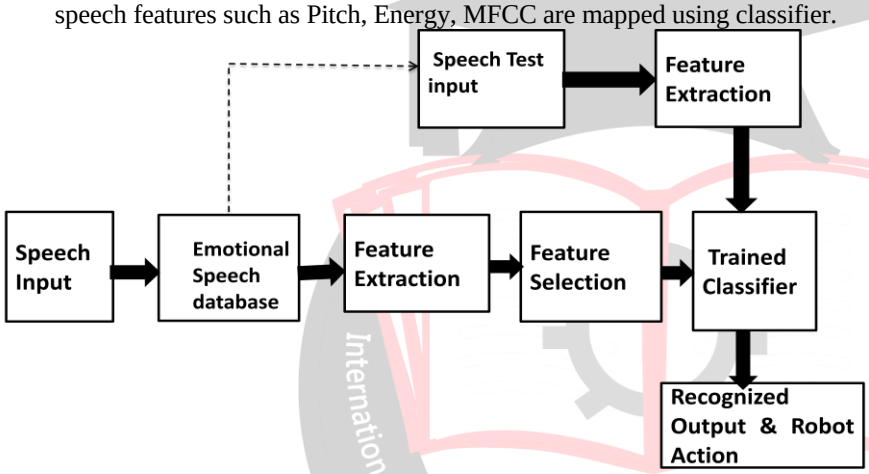

Figure 2: Structure of Speech Emotion Recognition

3.1Objectives:

1) Development of Database: Develop Speech Emotional Database for Marathi, Hindi, and English Languages.

2) Feature Identification and Extraction: Find out the Speech Features (time domain And Spectral Domain) which are suitable for particular language through Experimentation and comparing Results with available literature?

3) Implement Deep Belief Network and required number of Hidden layers. (Experimentation)

4) Train the DBN and test the performance with test input signal verify the results obtain.

5) Use this output to perform the application on some Robotics Platform.

3.2 Deep Belief network (DBN)

In machine learning, a deep belief network (DBN) is a generative graphical model, or alternatively a class of deep neural network, composed of multiple layers of latent variables ("hidden units"), with connections between the layers but not between units within each layer.
When trained on a set of examples without supervision, a DBN can learn to probabilistically reconstruct its inputs. The layers then act as feature detectors. After this learning step, a DBN can be further trained with supervision to perform classification
DBNs can be viewed as a composition of simple, unsupervised networks such as restricted Boltzmann machines (RBMs) or auto encoders, where each sub-network's hidden layer serves as the visible layer for the next. An RBM is an undirected, generative energy-based model with a "visible" input layer and a hidden layer and connections between but not within layers. This composition leads to a fast, layer-by-layer unsupervised training procedure, where contrastive divergence is applied to each sub-

network in turn, starting from the "lowest" pair of layers (the lowest visible layer is a training set). Teh's observation that DBNs can be trained greedily, one layer at a time, led to one of the first effective deep learning algorithms.:6 Overall, there are many attractive implementations and uses of DBNs in real-life applications and scenarios (e.g., electroencephalography, drug discovery).

3.3. Application

- Man-Machine Interface such as web movies , computer tutorials.
- In car board system to detect drivers emotions.
- Diagnostic tools for therapist.
- Automatic translation tools
- Aircraft cockpit.
- Call center and mobile application
- Emotional analysis of telephone conversation between criminal.
- Call analysis in emergency services like ambulance, fire brigade to evaluate genuineness of request.

IV. ACKNOWLEDGMENT

The authors would like to thank the "Department of Computer Engineering and Information Technology" at SNJB's LS KBJ COE,Chandwad,nashik, Savitribai Phule Pune University" for facilitating the development of the paper ,making available resources.

REFERENCES

[1] R. Banse, K. Scherer, Acoustic profiles in vocal emotion expression, J. Pers. Soc. Psychol. 70 (3) (1996) 614–636.

[2] Guihua Wen, Huihui Li, Jubing Huang, Danyang Li, and Eryang Xun "Random Deep Belief Networks for Recognizing Emotions from Speech and Communications, Cloud and Big Data Computing, Signals" Hindawi Computational Intelligence and Neuroscience Volume 2017, Article ID 1945630, 9 pages doi.org/10.1155/2017/1945630.

[3] Kasiprasad Mannepalli, Panyam,Narahari Sastry & Maloji Suman, " Design and development of Fractional Deep Belief Networks for speaker emotion recognition" International Journal of Speech Technology, ISSN 1381-2416 Int J Speech Technology DOI 10.1007/s10772-016-9368-y.

[4] Bu Chen, Qian Yin,, Ping Guo," A Study of Deep Belief Network Based Chinese Speech Emotion Recognition" 2014 Tenth International Conference on Computational Intelligence and Security, 978-1-4799-7434-4/14 $31.00 © 2014 IEEE DOI 10.1109/CIS.2014.148.

[5] M. Srikanth B, D. Pravena, and D. Govind", Tamil Speech Emotion Recognition Using Deep Belief Network(DBN)" Springer International Publishing AG 2018.S.M. Thampi et al. (eds.), Advances in Signal Processing and Intelligent.Recognition Systems, Advances in Intelligent Systems and Computing 678 DOI 10.1007/978-3-319-67934-1 29.

[6] Shashidhar G.Koolagudi ,K.Sreenivasa Rao "Emotion Recognition from a Speech : a Review " Springer science plus business Media LLC 2011,International journal of Speech technology 2012, DIO10.1007/s10772-011-9125-1.

[7] E.M.Albornoz ,M.Sanchez "Spoken emotion recognition using deep learning." 19th iberoamerican congress on pattern recognition (CIARP 2014) nov 2014.

[8]Wu li,Yanhui Zhang ,yingzi Fu," Speech Emotion Recognition in E learning system Based on Affective Computing" international conference on natural computation (ICNC 2007) IEEE computer society.

[9] Surabhi Vaishnav, Saurabh Mitra ,"Speech Emotion Recognition: A Review", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, Volume: 03 Issue: 04 | Apr2016.

