International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 221-224
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)

A REVIEW ON SPEECH EMOTION, RECOGNITION-ADVANTAGES, DISADVANTAGES, DIFFERENT TYPES & TECHNIQUES

Poonam Kamboj
Computer Science & Engineering
SVIET, Banur, Punjab, India

Er.Amarinder Kaur
Assistant Professor, CSE
SVIET, Banur, Punjab, India

Abstract—**Emotion Recognition is rapidly developing as a major aspect of human-computer interaction. Emotions are recognized efficiently by having a look at the facial expressions and simultaneously listening to the speech. However, Emotion Recognition solely based on speech signals has many applications in real time like the one discussed in which discusses about novel toys responding emotionally to the users. Speech recognition is a process used to recognize speech spoken by a speaker and has remained in the arena of investigation for additional than five periods since 1950s. Voice communication is the most real method of transmission used by individuals. Speech detection is an important and emerging technology with great potential.**

Keywords—**Speech Sentiment detection, voice transmission and technology**

I. INTRODUCTION

Emotional voice recognition purposes at robotically identifying the sensitive or physical state of a humanoid being from his or her speech. The expressive and somatic states of a speaker are recognized as emotional features ofvoice and are comprised in the so-called paralinguistic topographies. However the sensitive state doesn't adjust the verbal content, it is an important factor in humanoid transmission, because it delivers response info in many applications as it is outlined next [1] voice detection aims at involuntarily classifying the expressive or physical condition of a human being via his or this woman speech.

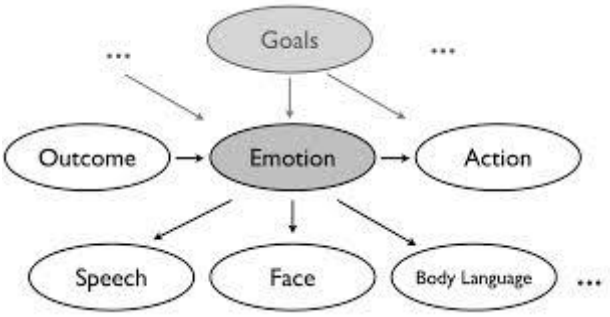

Fig. 1. Types Of Emotions

A speaker has unlike stages through speech that are recognized as emotional aspects of speech. Talking is perhaps the usually proficient way to agree with each other. This too means that speech could be a helpful border to collaborate with machineries. Speaker detection is also a challenging mission and is widely used in many speech aided applications [2]**.**

II. TYPES OF EMOTIONS

Emotions can be usefully divided into two broad types or classes—basic sentiment episodes and
energetic emotion-cognition connections or emotion schemas. Failure to make and retain the difference among these two types of emotion involvements may be the biggest source of misunderstandings and misconceptions in current emotion science.

1) *Basic positive emotions:* The basic positive emotions of interest and joy (e.g. an infant's interest activated by the human face and joy activated by the familiar face of her mother are equally essential to existence, development, and growth. However, their construction and time course may differ significantly once each additional. The infant's involvements of joy might be relatively brief by comparison with

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 221-224
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)

experiences of interest. The basic optimistic emotion of attention inspires play in early growth and thus may have short or relatively long duration.

2) *Basic undesirable sentiments:* Basic unwanted emotions (sorrow, anger, and revulsion, panic) typically run their course mechanically and stereo typically in a transitory time distance. The basic emotion of fear (or a fear-action episode) was described rather exactly in the original human archives: "A man who blunders upon a viper will jump aside: as trembling takes his knees, paleness his braveries; His backbones and backs not here".
3) *Basic or important emotions:* The discrete emotions of disgrace, guilt, and then dislike (sometimes named the communal or self-conscious sentiments) and the pattern of emotions in love and add-on may be measured basic in the intelligence that they are important to human evolution, normative development, human mentality, and real version. After language gaining, the sentiments connected to the self-concept or self-consciousness remain characteristically sentiment schemas that include higher order cognition (e.g., around personality and self-other relations) and must culture-related cognitive components [3].

III. ADVANTAGES OF SPEECH RECOGNIZE

Speech sentiment detection aims to mechanically classify the animated state of a humanoid being after his or her voice. It is founded happening in-depth examination of the cohort instrument of speech sign[10], extracting certain features which surround emotional information from the speaker's speech, and taking suitable pattern detection methods to recognize emotional states. Like characteristic pattern recognition systems, speech sentiment detection system covers four main units: voice input, feature mining, feature selection, cataloging, and sentiment output. Meanwhile a humanoid cannot categorize easily natural sentiments; it is difficult to suppose that machineries can offer advanced preciseorganization. A characteristic usual of sentiments covers 300 emotional states which are disintegrated hooked on six primary sentiments like anger, contentment, blues, surprise, fear, unbiased. Success of speech emotion recognition depends on naturalness of database.

IV. DISADVANTAGE OF SPEECH RECOGNIZE

Real-time emotion recognition is a great experiment for current practice as apparently stresses concerning heftiness and correctness are very high. Additionally, passable response times are indispensable for human-computer communication, as it becomes confusing for the user if she has to on hold and there is no direct reaction. Thus, the recognition process needs to be very fast to improve usability. Note that we are worried here individual with the examination phase of the classifier, as training can be done offline and is not obligatory to be particularly fast. The principal problem we faced is the fast segmentation of the continuously incoming audio sign into expressive, consistent sections. We originate a voice action detection with no in-between pauses longer than 1000 ms to be a good compromise between speed and accuracy. Interruption in the voice activity approximate phrase breaks, however the subsequent sections may not be linguistically complete. However, this segmentation requires no further information and is therefore very high speed. Furthermore, involuntary linguistic division byspeech recognition, besides being time-consuming, is static actual error-prone on impulsive dialogue, which could simply have negative influence on the emotion recognition, too[4].

V. APPLICATIONS OF SPEECH EMOTIONS

It was not astonishment that irritation was recognized as the most important sentiment for request hubs. Taking into explanation the position of anger and shortage of data for some other emotions we decided to generate a recognizer that can differentiate among two states: "agitation" which contains anger, happiness and panic, and "calm" which contains usual state and sorrow. To create the recognizer we used a corpus of 56 telephone mails of variable length (from 15 to 90 sec.) stating mostly normal and angry emotions that were logged through eighteen non-professional performers. These words were repeatedly split into 1-3 second chunks, which were then measured and branded by persons. They were used for making recognizers using the methodology developed in the first study [5]. The goal mouth of the expansion of this system was to create an emotion recognizer that can process telephone excellence speech memos (8 kHz/8 bit) and could be used as a slice of a result support system for prioritizing voice messages and assigning a proper agent to respond the message.

Fig.2. Speech Emotions

VI. RELATED WORK

Typical features are the pitch, the formants, the spoken tract cross-section extents, the Mel-frequency cepstral coefficients, the Teager energy operative based structures, the strength of the voice signal, and the signal rate. The third goal is to review appropriate methods in order to categorize speech into expressive states. We inspect separately classification methods that exploit timing information as of which that disregard it. Organization methods constructed on hidden Markov models,

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 221-224
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)

artificial neural networks, linear discriminant examination, and support vector machines (SVM), k-nearest devices are reviewed [6] classified in four sections namely HAPPY, SAD, FEAR and IRRITATED. Here are two leading units in this investigation work specifically Training and Testing. The exercise has remained done happening the basis of wave files providing for each group. Every user have distinct feature set hence when the operator is recognized their topographies has been removed for all groups and has been saved into the database. The testing unit categorizes the working out set of information with the help of BACK PROPAGATION NEURAL NETWORK (BPNN) classifier. Beforehand[11] organization of sentiments, firstly feature mining of speech signal is done using MFCC.[7]The master CPU averagely allocates primeval training usual to slave CPUs so that they could almost independently run serial SMO on their individual exercise set. As it accepts the strategies of bumper and shrink, the rapidity of the parallel training algorithm is amplified, which is presented in the experimentations of similar SMO based on the data set of MNIST. The experiments indicate that the similar SMO procedure has decent presentation in solving large scale SVM**.** [8]Related to emotional speech files used in expressive speech detection are also fleetingly discussed. On dissimilar structures used in the task of sentiment detection from voice is presented. The position of selecting dissimilar organization models has been discussed along with the appraisal. The significant issues to be reflected for further sentiment recognition investigation in common and in exact to the Indian situation have been underlined where ever necessary. [9]a novel dual formulation of the QCQP as a second-order cone software design problematic, and show in what way to exploit the method of Moreau-Yolinda regularization to yield a formulation to which SMO methods could be applied. We existing investigational effects that show that our SMO-based algorithm is meaningfully more effectual than the general-purpose inner point approaches available in current optimization toolboxes. [8]

VII. WHY EMOTION RECOGNIZES?

There are many ways that humans display their emotions. The most natural way to display sentiments is by means of facial terminologies. In the previous 20 years at this time has been much investigation on knowing sentiment through facial expressions recognizing the emotion through facial appearance showed in live video. The technique uses all of the chronological info showed in the video.[12] The logic behind hand using all of the temporal information is that any emotion being showed has exclusive temporal design. Several facial appearance research works classified each frame of the video to a facial appearance founded on certain set of features calculated for that time surround.

VIII. TECHNIQUES OF VOICE RECOGNITION

The aim of sentiment recognition scheme is to allow Human–Computer Communication (HCI). Furthermore, there are numerous areas in humanoid processor communication that could professionally use the ability to understand emotion. Understanding emotion can alsoplay[14] important role in brainy rooms and emotional computer tutoring. In this paper, current text self-governing emotion detection technique.

The scheme contains of four phases:

· *Speech achievement,*
· *Properties extraction,*
· *Properties selection*
· *Classification.*

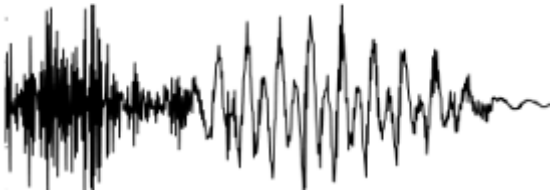

Fig. 3 Speech Signal (Time Domain)

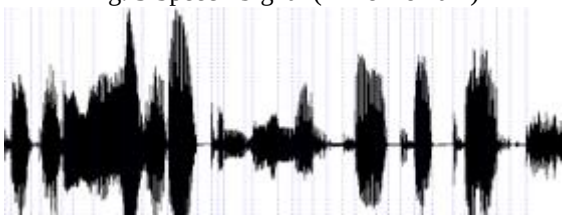

Fig.4. Speech Signal Spectrum(Frequency Domain)

The above figure defines the speech signal spectrum according to time and frequency domain in Mel- frequency co-efficient cepstral.

The examined expressive classes are fear, sad, irritated and happy. The MFCC features are removed from the voice signal for additional organization. It is essential to select the best feature for actual sentiment appreciation of any organization and so MFCC, which is one of the ghostly features, is used. Then, the BPNN is used for classification [9].

IX. CONCLUSION

Processing of emotions from speech helps to assure spontaneity in the presentation of existing voice schemes. Considerable quantity of work in this area is done in the recent past. A list of statistics assortments was providing including all obtainable info about the files such as the types of sentiments, the language, etc. Yet, there are immobile certain patent difficulties since the substantial from radio or TV is held under a limited agreement with broadcasters. First, decoding of emotions in speech is complex process that is prejudiced by national, social, and intelligent features of subjects. People are not perfect in decoding even such obvious emotions as irritation and pleasure. Additional, anger is the maximum recognizable and easier to portray emotion. It is also the greatest important sentiment for commercial. But irritation has frequent variants (for example, hot anger, cold anger, etc.).

X. REFERENCES

[1] "SPEECH EMOTION REVIEW PAPER (1)." .

[2] V. Garg, H. Kumar, and R. Sinha, "Speech based Emotion Recognition based on hierarchical decision tree with SVM, BLG and SVR classifiers," *2013 Natl. Conf. Commun. NCC 2013*, 2013.

[3] C. E. Izard, "Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues," *Annu. Rev. Psychol.*, vol. 29, no. 6, pp. 997–1003, 2012.

[4] S. Harbich and M. Hassenzahl, "Affect and Emotion in Human-Computer Interaction," *Affect Emot. Human-Computer Interact. SE - Lect. Notes Comput. Sci.*, vol. 4868, no. March, pp. 154–162, 2008.

[5] V. Petrushin, "Emotion in speech: Recognition and application to call centers," *Proc. Artif. Neural Networks Eng.*, pp. 7–10, 1999.

[6] J. Kaur and A. Sharma, "SPEECH EMOTION-SPEAKER RECOGNITION USING," vol. 3, pp. 308–312, 2014.

[7] P. Peng, Q. L. Ma, and L. M. Hong, "The research of the parallel SMO algorithm for solving SVM," *Proc. 2009 Int. Conf. Mach. Learn. Cybern.*, vol. 3, no. July, pp. 1271–1274, 2009.

[8] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech: A review," *Int. J. Speech Technol.*, vol. 15, no. 2, pp. 99–117, 2012.

[9] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," *Twenty first Int. Conf. Mach. Learn. ICML 04*, vol. 69, no. 1, p. 6, 2004.

[10] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, 2011.

[12] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, "Speech emotion recognition: Features and classification models," *Digit. Signal Process. A Rev. J.*, vol. 22, no. 6, pp. 1154–1160, 2012.

[13] K. Han, D. Yu, and I. Tashev, "Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine," *Fifteenth Annu. Conf. ...*, no. September, pp. 223–227, 2014.

[14] A. Ingale and D. Chaudhari, "Speech Emotion Recognition," *Int'l J. Soft Comput. Eng.*, vol. 2, no. 1, pp. 235–238, 2012.