A Novel Approach for Features Extraction Towards Classifying Normal and Special Children Speech Emotions in Urdu Language

***Maria Andleeb, Najmi Ghani Haider, Saman Hina, Syed Abbas Ali**

Department of Computer Science & Information Technology, N.E.D. University, Pakistan.

Summary
Spoken utterances play significant role in identifying the emotional states of speakers. However, extracted features add sense in spoken utterances which leads to provide speaker emotions. In this paper, a novel approach is presented for feature extraction toward classifying normal and special children speech emotion using spoken utterances in Urdu language. Eleven different features presents in this paper using thresholding technique on the extracted features implements the proposed algorithm namely: frequency, pitch, rate of zero passages, rate of acceleration, formant frequencies, intensity, log power, log energy, Mel Frequency Cepstrum Coefficients (MFCC), Linear Prediction Cepstral coefficient (LPCC) and Relative Spectral Transform - Perceptual Linear Prediction (Rasta PLP) with four different emotions (Angry, Happy, Neutral and Sad) to classify speech emotion of normal and special children. Experimental results evident that the proposed algorithm shows 100% accuracy with reduce error rate in Normal speech emotion and special speech emotion category for Angry, Neutral, Sad and Happy and Sad emotions respectively.

Key words:
Feature Extraction, Speech Emotion, Normal and special children, Urdu Language, classification accuracy.

I. Introduction and Related Work

The system which identifies the physical state as well as emotional state of the human being from his or her voice is called emotion recognition system [1]. The emotional as well as physical state of the human being is identified through speech emotion recognition. The four main modules of SER are shown in Fig.1, [2]:

Fig. 1 Classification of emotion

The early research work was based on utilizing small feature set (10-100) of prosodic features in particular pitch, intensity and duration, while in the recent research work , LLD's features such as shimmer, jitter, Harmonic to noise ratio (HNR), cepstral measurements have been used extensively [9,10]. Along with the pitch, energy and formant, rhythm and sentence duration was also included in [11] Yang et.al and Jin et al. Moreover MFCC was accompanied by Non-Uniform perceptual linear predictive features and LPCC in the feature set [11]. Some of the prosodic features and vocal expressions are associated with mental illness such as depression, Tourette syndrome [12]. In [13], Cooperative learning i.e. a novel method for highly efficient exploitation of unlabeled data is proposed. Cooperative learning is based on the idea of sharing the labeled work between machines and human efficiently in such a way that instances predicted with high confidence value are subject to machine labeling and those with low confidence value are human labeled. An experiment was performed on two emotions recognition tasks with two different sets and two emotion recognition tasks with two different sets and two emotion recognition tasks. The results of this experiment shows that in all test cases the cooperative learning outperforms individual active and semi supervised learning techniques. This method efficiently reduces the need for human annotations. The characteristics of depressed speech for the purpose of automatic classification are investigated in [14]. This research analyzed the effect of different speech features from over 40 depressed subjects plus 40 control subjects (both male and females), on classification results. ANOVA is used to find the characteristic of depressed speech statistically and the results are linked to Gaussian Mixture Model (GMM) and support Vector Machine (SVM). Likewise, classifier configuration employed in emotion recognition from speech is used to address the questions i.e. "how speech segments should be selected?", "What features provide good discrimination?", and "What benefits feature normalization might bring given the speaker-specific nature of mental disorder?" The database of 23 depressed and 24 control subjects was obtained from audio-video data collection. Investigation in [15] deals with the problem of detecting depression from the recordings of subjects using speech processing and machine learning. It concluded that the features from harmonic model improve the performance

Manuscript received July 5, 2017
Manuscript revised July 20, 2017

of detecting depression from spoken utterances than other alternatives. The recordings used in this experiment consisted of 148 subjects, including 98 females and 50 males. Similarly another research suggested several objective voices acoustic measures from 35 physician referred patients affected by depression can be obtained reliably, over the telephone. Spectral features are extracted from these frames also called segmental features. Pros and cons of some spectral features DSCC, MFCC, LPCC, ZCR, Jitter, shimmer, LPCC. Also called as acoustic features, they are referred to as suprasegmental features. Attributes of acoustic features are pitch, intensity, formant, energy, spectral tilt. The teaser energy operators are the nonlinear features that are extracted from stressed speech [16-17]. The most common classifier for speech emotion recognition was HMM [18-21]. HMM consist of first order Markov chains whose states are hidden from the observer and they are the stochastic processes. For the state of stress recognition HMM classifier is used. The most recent classifier to be used is the SVM as assessed in many papers [22-25]. The most important reason to use SVM over GMM and HMM is that it provides the global optimality of training algorithm and the existence of excellent data dependent generalization bond [26].

Automated speech analysis combined with machine learning was tested to predict the later psychosis onset in youths at Clinical High risk (CHK) for psychosis. It interviewed 34 CHR youths (11 females) and assessed quarterly for 2.5 years. Convex hill classification algorithm was used for speech features feeding. The speech features included the measure of semantic coherence and two syntactic markers of speech complexity that is the use of determiners and maximum phase length. 100% accuracy was shown with these speech features for predicting later psychosis development [27]. Likewise, an algorithm was formulated in which the automatic segmentation of speech signal was the basis to detect voiced segments. In order to estimate the pitch and pitch changes, the spectral matching approach was used. Evaluation of the performance of algorithm was carried out on speech database, in which the electroglottographic signal is included [28]. The acoustic model trained with the neural network is used to build a system with split temporal context features. Knowledge of the speech language pathologist with constrained HMM encoded was also used. Phonological error pattern is improved by 33% when compared with standard HMM decoders. Small corpus of speech disordered samples of Australian Children was used in this study but this led to the achievement of 94% accuracy [29]. Similarly, the two most commonly used classifiers that are Gaussian Mixture Model- Universal Background model (GMM-UBM) and Support Vector Machine (SVM) is used to analyze the linear prediction cepstrum. The hybrid strategy of GMM-UBM and SVM is used to investigate the performance gain in detection of articulation disorder from children speech [30]. Psychometric properties of autism spectrum disorder-comorbid for Children (ASD-CC) was developed and evaluated in this study. Factor structure of the Korean version of ASD-CC was figured out using confirmatory factor analysis (CFA). The internal consistency and test reliability were measured. Finally the inter correlation was computed between the Korean child behavior checklist (K-CBCL) and ASD-CC [31]. In another study the treatment option suggested are limited to psychosocial therapies for the core symptoms of autism. For the autism spectrum disorder, it is suggested that5 Risperidone and aripiprazole are the only FDA approved medications [32]. Behavior consistent with autism spectrum disorder (ASD) is described by 19 items [33]. To identify the social-emotional/ behavioral problems and delay / deficits in emotional competence the brief infant toddler social emotional assessment (BITSEA) is designed which is a 42 item screener.

II. Materials and Methods

The data evaluated in this study were collected from 200 normal and 200 special children of age group 10-13 years of both genders. An Urdu language sentence with four different emotions (Angry, Happy, Neutral and Sad), suitable to be uttered is chosen for this study. The following ITU recommendations have been used for corpus recording with specifications: SNR>= 45dB and bit rate 24120 bps. The selected sentence was "**I have to play**". Windows 10 built in sound recorder has been used for recording the speaker's utterances with 48 KHz sampling rate and sensitivity of 56dB±25dB.

1) Speech Emotion and Language Disorder Recognition System Model:

Speech emotion and language disorder recognition system is consists of steps of speech processing which are shown in Fig. 2.

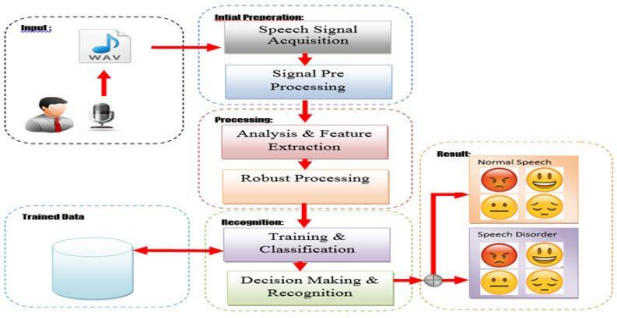

Fig. 2 System flow of speech emotion and language disorder recognition system

In signal acquisition, microphone is used to captures the speech. The analog signal to digital signal is carried out by sampling it 16000 times a second as shown in Fig. 3.

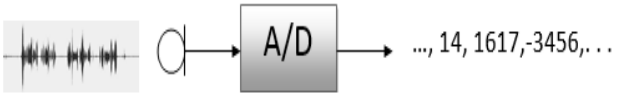

Fig. 3 Signal Acquisition

In order to improve the efficiency of subsequent feature extraction and classification stages. The preprocessing stage in speech recognition system is used. The general preprocessing pipeline is depicted in Fig. 4.

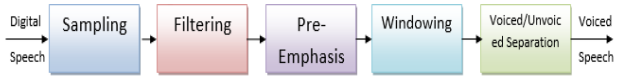

Fig. 4 Preprocessing pipeline of speech recognition system

After preprocessing step, digitization of the speech signal is the first and foremost procedure to enable it to be processed by computer system. The time and value discrete signal is the result of sampling and quantization applied on time continuous speech signal. Nyquist Shannon sampling theorem states that sampling frequency of at least 2fmax is needed to sample the band limited time continuous certain signal frequency fmax, so the signal can be reconstructed by its time discrete signal. Recognition accuracy is highly affected when the sampling frequency is used along with the speech vector size. A A6kHz sampling frequency is sufficient for the task of speech recognition as there is a low bandwidth of human speech mostly between 100Khz and 8Khz on sampled values. Quantization is done to have the value discrete signal so a significant reduction of data takes place. Usually 8 or 16 bits samples are encoded by the speech recognition system, depending on the available processor power. Sampled values with a higher bit resolution are preferable in case of sufficient processing power. The speech signal quality is improved by the stage of noise reduction or de-noising that is the speech degraded by noise, aims to improve the intelligibility of signal. Speech signal corrupted by noise has the following three categories; a) Microphone related noise, b) Electrical noise, c) Environmental noise. Moreover, there are three fundamental classes of noise reduction algorithms;

- Filtering Techniques: Adaptive Wiener filtering and the spectral subtraction methods are the prominent algorithms that are based on filtering techniques.
- Speech Restoration: The inducing of missing spectral component of the nonverbal sounds by adding noise to increase intelligibility is referred to as spectral restoration.
- Speech Model Based: Noise+ harmonic model of speech is used by a de noising technique which is referred to as harmonic decomposition.

The improvement in signal to noise ratio is the criteria to measure the quality of noise reduction system and the best measure is the improvement in recognition performance.

2) Feature Extraction:

Features were extracted for speech emotion recognition of special children. Following text contains description of extraction of each feature and specific equations of each feature are given in. First feature is frequency which starts with the loading of the speech signal. Then, the analog signal is converted into numeric data. After loading process, maximum and minimum frequency is set and Fast Fourier transform (FFT) of windowed signal is performed as shown in equations (1) and (2). This process is followed by the cepstrum calculation of the frequency. After extracting frequency, Next feature is pitch contour in which signal acquisition and signal processing is same as described for extraction of frequency. After this, FFT is applied on the processed signal and Fourier Transform (DFT) is taken of the FFT signal using specific formula described in

Table 1 in equations (3) and (4). Then the log of FFT is calculated and the real cepstrum is calculated. This will give the absolute value of filtered log of DFT and finally, real cepstrum pitch is extracted as shown in equation (5). For the intensity feature, the magnitude of the DFT signal with log 10 is filtered signal Y as shown in Equation (6). The conversion of the magnitude into decibels and transpose result of decibel gives out intensity as shown in Equation (7). Another feature, ‘rate of acceleration’ was extracted by processing of the signal which involves the time instant calculation as shown in

Table 1. Then the speed of the time instant was calculated and the derivative of speed gives velocity. After this, the gradient of velocity by 0.01 yields the acceleration as shown in Equations (8) and (9), finally the average rate of acceleration is calculated by taking the mean of acceleration as mentioned in equation (10) . Then, equation (11), (12), (13) shows the extraction of formant frequencies.

Extraction of log power feature was obtained by preprocessing steps that involves the setting of the sampling rate and sampling window size. Equations (14), (15), (16) were used for the calculation of log power. Another feature, log energy was extracted by preprocessing that involves the same procedure as log power extraction. Average energy is calculated by applying windowing on signal x according to window size and frame size save as ‘result’ as shown in equation (17) in

Table 1.

Similarly, equation (18) was used for the extraction of 'rate of zero passages', after that three feature extraction techniques that is MFCC, LPCC and Rasta PLP are used to extract spectral features. Equation (19), (20) was used for Mel Frequency Cepstrum Coefficient (MFCC), equation (21), (22) for linear Prediction Cepstrum Coefficient (LPCC).For last feature Rasta PLP, processing starts by signal acquisition and signal processing involves setting of the speech signal and sampling rate of samples. Then setting of the Rasta Default to 1 is carried out. If it is set to 0 then start by calculating PLP. It is followed by setting the model order to whose default value is 8. Feature extraction involves few steps. First is to compute the power spectrum and frame energy of the input signal. Second is the critical band analysis performance. Then Rasta filtering is performed and LPCC is converted into LPC. Finally smoothing of the LPC gives out Rasta PLP.

Table 1: Feature description and its relevant equations

Extracted Feature	Equations of Feature extraction	Equation Number
Frequency	ms1=fft(x.*hamming(length(x))) ; f=fs/(ms1+fx-1);	(1) (2)
Pitch	dfty=abs(fft(x)); dft=log10(dfty); rcp=real_ceps(16:length(real_ceps)).	(3) (4) (5)
Intensity	M=20*log10(abs(Y(1:length(x)))+eps); i = mag2db(M).	(6) (7)
Rate of acceleration	x1=(0.4*t.^4)+(10.8*t.^3)-(64.4*t.^2)-(28.2*t)+4.4; v=diff(x1); acc=gradient(v,0.01)	(8) (9) (10)
Formant frequencies	ncoeff=2+Fs/1000; a=lpc(ncoeff); r=roots(a);	(11) (12) (13)
Log Power	windowsize = sampling_rate*fs; Average_Energy = sum(result.^2); Lp= Average_Energy/windowsize	(14) (15) (16)
Log Energy	Average_Energy = sum(result.^2);	(17)
Rate of Zero passages	RZP = sum(abs(sign(y1)-sign(y2))/2)/windowsize.	(18)
Mel Frequency Cepstrum Coefficient	dctm = @(N, M)(sqrt(2.0/M) * cos(repmat([0:N-1].',1,M) ..... .*repmat(pi*([1:M]-0.5)/M,N,1))); CL = @(N, L)(1+0.5*L*sin(pi*[0:N-1]/L))	(19) (20)
LP cepstrum Coefficients	[x_lpc] = lpc(x, p); lpcc= fft(x_lpc, N)	(21) (22)

[Where ms1=windowed signal; fft=Fast Fourier Transform; f=frequency; x=input signal; fs=sampling frequency; fx= input frequency; dft=discrete fourier transform; rcp=real cepstrum pitch; M=Magnitude; Y=Filtered signal; i= intensity; mag2db= Magnitude to decibel; x1=speed; T= Time instant; v=velocity; diff=derivative; acc=acceleration; N=Number of Coefficients; lpc= Linear Prediction Cepstrum; lpcc=Linear Prediction cepstrum coefficient; r=roots; sum= sum of frame size of windowed signal; Lp= Log power; RZP= Rate of Zero passages; y1=Maximum frequency; y2= Minimum frequency; dctm= discrete cosine transform matrix; M= No of filter bank channels; L=Length of channel; CL=Cepstrum filter; p= prediction paths; N= Number of shifts.]

III. Theory/Calculation:

After extracting all features, a new approach is implemented on the training data set. The algorithm proposed for this new approach is elaborated in Fig.5. This algorithm will be used to detect the language disorder that classifies the emotions that are angry, happy, neutral and sad accordingly.

Algorithm : The proposed framework to detect the language disorder and classify the emotions

1. **Extract all features of input test speech**
2. **Classify test sample by classification function**
3. **Set the threshold for each feature**
4. **// Calculate the maximum category score position**
5. ***if result of position rn=1***
6. **Set category to normal and write the percentage**
7. ***else***
8. **Set category to special and write the percentage**
9. ***end if***
10. **// Calculate the maximum emotion score position**
11. ***if result of position rn=1***
12. **Set emotion to Angry and write the percentage**
13. ***else if rn=2***
14. **Set emotion to Happy and write the percentage**
15. ***else if rn=3***
16. **Set emotion to Neutral and write the percentage**
17. ***else***
18. **Set emotion to Sad and write the percentage**
19. ***end if***
20. **Save category and emotion as label results**
21. **Extract percentage to display.**
22. **Extract all features of input test speech**
23. **Classify test sample by classification function**
24. **Set the threshold for each feature**

```
// Calculate the maximum category score position
if result of position rn=1
Set category to normal and write the percentage
else
Set category to special and write the percentage
end if
// Calculate the maximum emotion score position
if result of position rn=1
Set emotion to Angry and write the percentage
else if rn=2
Set emotion to Happy and write the percentage
else if rn=3
Set emotion to Neutral and write the percentage
else
Set emotion to Sad and write the percentage
end if
Save category and emotion as label results
Extract percentage to display.
```

Fig. 5 Language disorder detection Algorithm

Furthermore, the whole system flow diagram for recognition and classification of speech emotion is shown in Fig.6. In the system flow, the thresholding of the extracted features is done before data file preparation. The process of data file preparation for each sample file is applied for decision making of the algorithm. It starts by loading the sample file speech.mat. After loading of file, the file name, file path categories and emotions were extracted by path. Then, the file name is added in speech. mat. In addition to this information, files containing information of emotions and categories are also added in speech.mat. Finally, the thresholding feature extraction is performed using equation (23).

Features= allfeatures_extraction(wav_file); (23)

A function for decision making is implemented in equation (24);

Function [Category, Category_Percentage, Emotion, Emotion_Percentage] = hybrid_Decision_making(wav_file) (24)

The test samples were classified by using hybrid classification function and then finding the maximum category score position. The process of hybrid classification is described as follows;

Step#1: If Result of position n = 1

Set category = Normal and Category_Percentage=Categories_output(1).

Else category = Special and Category_Percentage=Categories_output(2).

Step#2: After the category is decided as normal or special, the emotion is labeled on the category according to the classification accuracy. The maximum score position is to be estimated in equation (25)

[rn,cn]=find (strcmp (Emotion_outputs, max(Emotions_output))) (25)

Step#3: If Result of position n = 1

Set Emotion = Angry and Emotion _Percentage= Emotions_output(1).

Else If Result of position n = 2

Set Emotion = Happy and Emotion _Percentage= Emotions _output(2).

Else If Result of position n = 3

Set Emotion = Neutral and Emotion_Percentage= Emotions_output(3).

Else

Set Emotion = Sad and Emotion _Percentage= Emotions _output (4).

End of If condition.

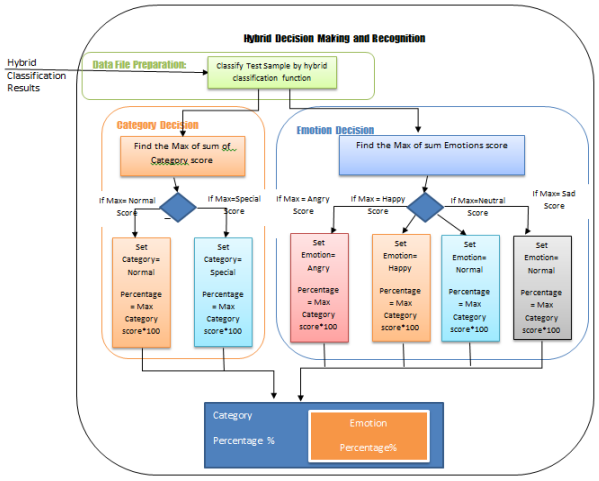

Fig.6 Proposed strategy for decision making and recognition

IV. Evaluation and Implementation of Algorithm

The performance of the proposed classification model is shown by the help of confusion matrix which is depicted in Fig.7. The total of 92 samples is tested and there are two classes i.e. “Output class” and “Targeted class”. Label 1 is the “Normal Category” and Label 2 is the “Special” category. Out of 92 samples 48 are the normal speech

samples and 44 are the special speech samples. Green blocks are accurate hits/class and Red blocks are error/miss per class. Our results are complete hits which are shown by Grey boxes and shows 100% accuracy as the targeted samples gives accurate output. Similarly testing is performed on 92 samples to test the classification accuracy of emotions shown in Fig.8. 23 samples of each emotion is taken. Label 1 is the "Angry" emotion, Label 2 is the "Happy", Label 3 is the "Neutral" emotion and Label 4 is the "sad" emotion. In this case all the emotions samples which are tested shows 100% accuracy as shown by grey boxes.

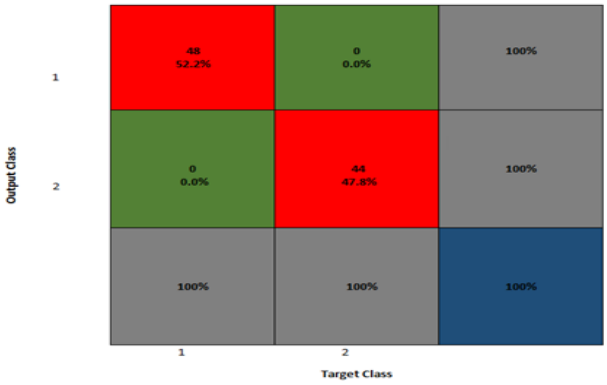

Fig.7 Confusion Matrix of testing the Normal and Special category

Fig.8 Confusion Matrix for Emotions

After implementation of proposed algorithm on the training data set, accuracy for classification of each emotion is achieved which is shown in Table 2. For 'Normal' category, it can be seen that accuracy is 100% for emotion 'Angry'; however in case of 'Special' category its accuracy is lower for the same emotion. Another interesting score is for emotion 'Sad', which has achieved 100% accuracy for both 'Special' and 'Normal' category.

Table 2: Classification Accuracy of categories and emotions on training set

Category	Emotions	Classification Accuracy (%)
Normal	Angry	100%
	Happy	66.7%
	Neutral	100%
	Sad	100%
Special	Angry	66.7%
	Happy	100%
	Neutral	83.33%
	Sad	100%

Table 3: Error rate of categories and emotions on training set

Category	Emotions	Error Rate (%)
Normal	Angry	0%
	Happy	33.33%
	Neutral	0%
	Sad	0%
Special	Angry	33.33%
	Happy	0%
	Neutral	16.7%
	Sad	0%

For analyzing the impact of algorithm, error rate was also analyzed which is shown in Table 3 for each category and emotions. It is shown that the Classification accuracy of the algorithm is 100% for the category Normal and Emotion Angry; Neutral and Sad and for the category Special and in case of emotions 'Happy' and 'Sad'. In 'Normal' category the error rate is shown to be 33.33% for the 'happy' emotion and for 'Special' category the error rate is 33.33% and 16.7% for the angry and neutral emotions respectively. No error rate was observed for any other emotion on training set.

V. Conclusion

This paper proposed feature extraction approach to extract eleven different features for classifying the emotion based spoken utterances of normal and special children in Urdu language namely: pitch, rate of acceleration, formant frequencies, intensity, log power, log energy, Mel Frequency Cepstrum Coefficients (MFCC), Linear Prediction Cepstral coefficient (LPCC) and Relative Spectral Transform - Perceptual Linear Prediction (Rasta PLP). Experimental frame work were comprised on collected data corpus of 200 normal and 200 special children spoken utterances of age group 10-13 years for both gender in four different emotions (Angry, Happy, Neutral and Sad). The proposed algorithms were evaluated in MATLAB tool on 92 samples for 'Special' and 'Normal' categories against each emotion in term of classification accuracy and error rate. Demonstrative experiments show 100% classification accuracy and reduce error rate in Normal speech emotion category for Angry, Neutral and

Sad emotions and in Special emotion Category for Happy and Sad emotions on both training and test data set. Authors are focusing on developing experiments with large data samples and evaluating proposed approach against traditional machine learning classifiers.

References

[1] Ramakrishnan, Dr. Mahalingam, "Recognition of Emotion from Speech: A Review", InTech China and Europe, pp. 121-138, 14 March 2012.

[2] Swati Pahune, Nilu Mishra, "Emotion Recognition through Combination of Speech and Image Processing: A Review", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 3, pp. 134– 137, February 2015.

[3] Schuller B, Batliner A, Seppi D, Steidl S, Vogt T, Wagner J, Devillers L, Vidrascu L, Amir N, Kessous L, Aharonson V , "The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functional". In: Proceedings of INTERSPEECH, pp 2253–2256, 2007.

[4] Schuller B, Rigoll G, " Recognising interest in conversational speech–comparing bag of frames and supra-segmental features". In: Proceedings of INTERSPEECH, pp 1999–2002, 2009

[5] LuggerM,Yang B , "An incremental analysis of different feature groups in speaker independent emotion Recognition", In Proceedings of international congress phonetic sciences, pp 2149–2152, 2007.

[6] Firoz Shah A, Vimal Krishnan VR, Raji Sukumar A, Jayakumar A, Babu Anto P, "Speaker independent automatic emotion recognition from speech: a comparison of MFCCs and discrete wavelet transforms", In: Proceedings of international conference on advances in recent technologies in communication and computing, pp 528–531, 2009.

[7] Neiberg D, Elenius K, Laskowski K , "Emotion recognition in spontaneous speech using GMMs". In Proceedings of INTERSPEECH conference, pp 809–812, 2006.

[8] Vlasenko B, Schuller B,WendemutA, RigollG, Frame vs, "Turn-level: emotion recognition from speech considering static and dynamic processing". In Proceedings 2nd international conference on affective computing and intelligent interaction, pp 139–147, 2007.

[9] LuggerM,Yang B , "An incremental analysis of different feature groups in speaker independent emotion recognition", In Proceedings of international congress phonetic sciences, pp 2149–2152, 2007.

[10] Jin Y, Zhao Y, Huang C, Zhao L, "Study on the emotion recognition of whispered speech" , In Proceedings of global congress on intelligent systems, pp 242–246, 2009.

[11] Zhou Y, Sun Y, Yang L, Yan Y, "Applying articulatory features to speech emotion recognition", In Proceedings of international conference on research challenges in computer science, pp 73–76, 2009.

[12] Calder J, Lawrence AD, Young AW , "Neuropsychology of fear and loathing" , Nat Rev Neurosci vol. 2, pp.352–363, 2001.

[13] Zixing Zhang, Eduardo Coutinho, Jun Deng, Björn Schuller, "Cooperative Learning and its Application to Emotion Recognition from Speech", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 1, \pp 115-126, Jan 2015.

[14] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, G. Parker, M. Breakspear, "Characterising Depressed Speech for Classification", in proceedings of Conference of the International Speech Communication Association (InterSpeech), Florence, Italy 2013.

[15] Asgari, I. Shafran, L. B. Sheeber, "Inferring clinical depression from speech and spoken utterances", in proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1 – 5, Reims, France, 21-24 Sept. 2014.

[16] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: resources, features, and methods", Speech Communication, vol. 48, no.9, pp. 1162-1181, Apr 2006.

[17] G. Zhou, J. Hansen and J. Kaiser, "Nonlinear feature based classification of speech under stress", IEEE Trans. on Speech and Audio Processing, vol. 9, no. 3, pp. 201-216, Mar 2001.

[18] Yun S,Yoo CD, "Speech emotion recognition via a max-margin framework incorporating a loss function based on the Watson and Tellegen's emotion model". In Proceedings IEEE international conference on acoustics, speech and signal processing, pp 4169–4172, 2009.

[19] Yu W, " Research and implementation of emotional feature classification and recognition in speech signal", In Proceedings of international symposium on intelligent information technology application, pp .471–474, 2008.

[20] Fu L, Mao X, Chen L , "Speaker independent emotion recognition using HMMs fusion system with relative features" . In Proceedings of 1st international conference on intelligent networks and intelligent systems, pp 608–611, 2008.

[21] Nogueiras A, Moreno A, Bonafonte A, Mariño JB, " Speech emotion recognition using Hidden Markov models". In Proceedings of EUROSPEECH", pp 2679–2682, 2001.

[22] Vlasenko B, Schuller B,WendemutA, RigollG, Frame vs , "Turn-level: emotion recognition from speech considering static and dynamic processing" . In Proceedings 2nd international conference on affective computing and intelligent interaction, pp 139–147, 2007.

[23] Luengo I, Navas E, Hernaez I, "Feature analysis and evaluation for automatic emotion identification in speech. IEEE Trans Multimed, vol.12, pp.490–501, 2010.

[24] Schuller B, Batliner A, Steidl S, Seppi D ,"Emotion recognition from speech: putting ASR in the loop", In Proceedings of IEEE international conference on acoustics, speech and signal processing, pp 4585–4588, 2009.

[25] Wang Y, Du S, Zhan Y, "Adaptive and optimal classification of speech emotion recognition. In: Proceedings of 4th international conference on natural computation, pp 407–411, 2008.

[26] Anagnostopoulos CN, Vovoli E , " Sound processing features for speaker-dependent and phrase-independent emotion recognition in Berlin Database". In: Papadopoulos GA, Wojtkowski W, Wojtkowski G, Wrycza S, Zupancic J (eds) Information systems development, pp 413–421, 2010.

[27] G. Bedi, F. Carrillo, G.A. Cechi (et.al), "Automated analysis of free speech predicts psychosis onset in high risk youths", NPJ Schizophrenia, 26 August 2015.

[28] N. Vanello, A Guidi, C. Gentili, s.Werner, G. Bertschy, G. Valenza and E.P Scibingo, "Speech Analysis for model state

characterization in bipolar patients,", Engineering in Medicine and Biology society (EMBC), in Proceedings of IEEE Annual conference, Association, pp. 2104-2107, 2012.
[29] L. Ward, A. Stefani, D.Smith, A.Duenser, J. Frenzyne, B. Dodd, A. Morgan, "Automated screening of speech development issues in Children by identifying phonological error patterns", Interspeech, pp. 2661-2665, September 2016.
[30] N.Ramou and M.J Guerti, "Automatic detection of articulations disorder from Children's speech preliminary study", Commun-technol Electron, Springer, vol.59, pp 1274-1279, November 2014.
[31] K-M chung, D-Jung, "Validity and reliability of the Korean version of autism spectrum disorders comorbid for children (ASD-CC)", Research in autism spectrum disorders, vol.39, pp. 1-10, July 2017.
[32] M. Deflippis and K.D. Wagner, "treatment of autism spectrum disorder in children and adolescent", psycho-pharmacology bulletin, vol.46, pp. 18-41, 15th August 2016.
[33] I.G. Kiss, M.S. Feldman, R.C.Sheldrick, A.S Carter, "Developing Autism Screening Criteria for the Brief Infant Toddler Social Emotional Assessment (BITSEA)", Journal of Autism and development disorder, vol.47, pp.1269-1277, May 2017.

Maria Andleeb Siddiqui received her B.E degree in Telecommunication Engineering in 2010 from NED University of Engineering and Technology, Karachi, Pakistan. She received M.E degree in Computer and Information **system** engineering in 2013 from NED University of Engineering and Technology. Currently she is enrolled in PhD in the same University. Her research area is speech emotion recognition. She is lecturer in Usman Institute of Technology.

Najmi Ghani Haider Professor, Department **of** Computer Science & Information Technology supervisor for research scholar, his research directions include Machine learning, robust speech recognition and image processing.

Syed Abbas Ali received his B.E. and M.Engg. degrees in Computer Systems & Electrical Engineering from the N.E.D University of Engineering & Technology, Pakistan, in 1999 and 2002 respectively. Also, he received Ph.D. degree in Computer Science from N.E.D University of Engineering & Technology, Pakistan, in 2015. Since 2005, he has been working as an Assistant Professor in Department of Computer & Information Systems Engineering, Faculty of N.E.D University. His current research interests include Machine Learning & Speech Recognition, Speech Emotion Recognition and Computational Intelligence..