An Epitomization of Stress Recognition from Speech Signal

Veena Narayanan[1], S Lalitha[2], Deepa Gupta[3]

[1,2]Department of Electronics & Communication Engineering
[3] Department of Mathematics
Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India
E-mail: veenaswaraj6@gmail.com, s_lalitha@blr.amrita.edu, g_deepa@blr.amrita.edu
**Corresponding author E-mail: s_lalitha@blr.amrita.edu*

Abstract

Detection of stress from speech signal is gaining large attention recently. The emergence of new methods and techniques for feature extraction and classification paved the way to different solutions to detect different stress conditions using human speech and led to an increase in the accuracy of stress recognition. A large number of parameters are proposed for the characterization of stress in speech. Similarly numerous classifiers and machine learning algorithms are investigated for stress classification and regression. In this treatise, a recital on the commonly used databases, stress conditions, different feature extraction methods and classifiers along with some of the statistical measures as well as compensation techniques for stress detection are presented in this article. After thorough illustration of existing methodology for the task, future prospects for the work are elaborated.

Keywords: *Stress Recognition; Feature Extraction; Statistical Measures; MFCC; LPCC*

1. Introduction

Stress is the response of an individual to numerous factors like multitasking, emotional state, deprivation of sleep, deception, emergency situation, high workload, deceiving or depressing environment. The day to day changing lifestyle and behavior of people has led to a stress imbalance and hence there is a need for maintaining a proper stress balance.

Stress can be detected from different measures. Facial recognition, heart rate variability, fingertip temperature, galvanic skin response, electrocardiogram, blood volume pulse and speech signal are some of the measures used to mark the stress as shown in Fig.1. Facial recognition is done by selecting facial features from digital image. The eye gaze, head movements and facial expressions are the visual cues considered [1]. The heart beat variability indicates the changes in the heart beat time interval and it can be measured using electrocardiogram and blood pressure [2]. The fingertip temperature indicates if a person is stressed or not. The warmer fingertip indicates that the person is relaxed and a cooler fingertip indicates that the person is stressed or tensed [3]. The galvanic skin response is the resistance offered by the skin when a person is under stress [4]. The blood pulse volume is the change in volume of blood that happens with heart rate variability and each heartbeat [5]. It can be measured using a photo plethysmograph. Speech is one of the fastest and most natural forms of human response. Hence among all these measures, it is widely used to detect different stress conditions. Stress also causes alteration in the speech production system and this arises due to glottal abnormalities. These abnormalities can affect the speech recognition system thereby reducing its efficiency and recognition rate. Thus for an efficient human machine interaction and better performance of speech recognition system, there is a need to detect the stress. A wide research has been done on speech emotion recognition to improve the human computer interaction[6]. Most researches are done using Berlin Emotional speech database comprising different emotions [7]. A detailed review of the databases comprising 32 emotional speech databases including English, German, Spanish, Dutch, Russian, Sweden and Chinese are presented in [8]. A survey on the emotional speech datasets, features and classifiers is done[9]. The recognition of stress along with emotion recognition can improve the recognition results.

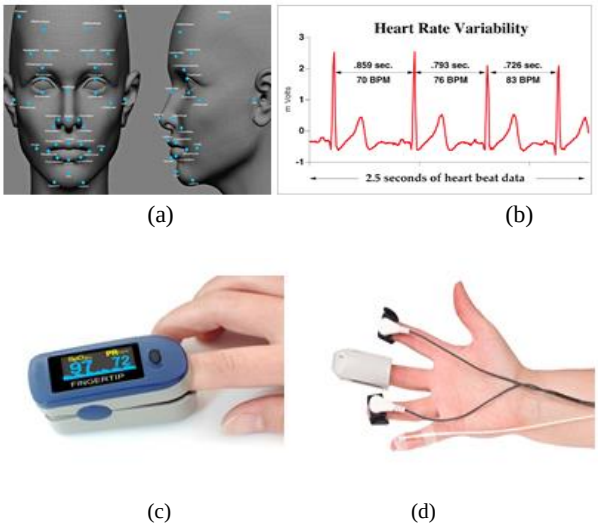


Fig. 1: Various modalities for stress Detection (a) Facial recognition [www.theguardian.com] (b) Heart rate variability [www.armbeep.com] (c) Fingertip temperature[www.quora.com](d)Blood pulse volume [www.biofeedback-tech.com]

Stress detection from speech signal is having a lot of applications. It is used in psychology to monitor the different stress levels of patients with different stress conditions and provide necessary treatments. The safety and security of a system can be established by monitoring the different stress levels of pilots, deep sea divers and military officials facing law enforcement. Stress detection is also useful in speaker identification, deception detection and identification of threat calls in few cases of crime[10]. Thus, all these factors and applications suggest that there is a need to explore this area of stress detection so that it will provide insights into many existing speech related issues.

The article is organized in five sections. The first section explains the general methodology of speech based stress recognition. Section II provides detailed information about different existing features, feature extraction methods and classifiers used for stress

detection. The various database used and their details are mentioned in Section III. In Section IV, the discussion and challenges of stress recognition are explained and section V includes the future directions of the article.

2. General Framework of Stress Detection using Speech

Stress detection is usually carried out in three stages which are preprocessing, feature extraction and classification as shown in Fig 2. Windowing and framing are the common preprocessing techniques used. Other methods are analog to digital conversion which converts the analog speech signal to digital signal and end point detection in which the silence portions in the speech signal are removed. Zero Crossing Rate and Short Time Energy are the two methods used for end point detection. Higher frequencies in speech signal are enhanced using pre-emphasis techniques[11]. A variety of speech features including acoustical, prosodic, biomechanical and glottal features can be extracted from the speech signal. Prosodic features convey the tone and rhythm of a signal. Features can also be classified as temporal features and spectral features. Temporal features are the time domain features that have an easy physical interpretation and are simple to extract. Some of the examples of temporal features are zero crossing rate, maximum amplitude and short term energy of the signal. Spectral features are the frequency based features obtained from the frequency domain[12]. The examples of spectral features are MFCC, LPCC, spectral centroid, spectral flux, fundamental frequency and spectral crest factor. These features after extraction are given to the classifiers for classification.

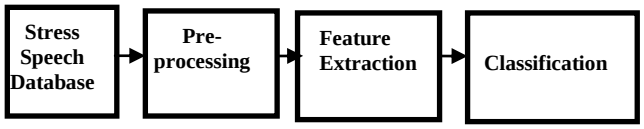


Fig. 2: Block diagram of Stress Detection

In most of the cases, the classifiers make decisions based on the training data and they are tested using testing data. The classifier makes a model from the training data and predicts the target values of the test data. Some of the commonly used classifiers are Support Vector Machine (SVM), Naïve Bayes, Artificial Neural Network (ANN), Hidden Markov Model(HMM), Vector Quantization (VQ) and Gaussian Mixture Model(GMM) classifier[13]. A single classifier system and multiple hybrid classifier system can be used for classification [9]. Many machine learning algorithms are used to support the detection of stress in speech signal. These algorithms are used to classify different speech samples under different stress conditions. Statistical and Qualitative analysis are also performed for achieving the discrimination capability of each features considered. To compensate for the noise and other degrading factors present in speech, various compensation techniques are also used. All these techniques are illustrated in section IV.

3. Database Used

The creation of suitable stress database is a crucial part in the investigation of stress conditions. Most of the existing databases are elicited, acted or natural speech. The databases were recorded from speakers who were facing an examination or an interview, aircraft controllers and pilots under stress, speakers under emergency situations or recordings from speakers who were made to do multiple tasks at the same time [10]. Speech recordings were taken from passengers reading words during rollercoaster ride [14]. But, maintaining a consistent stress level in speech database is difficult to achieve using spontaneous stress speech. Therefore, the acted stress speech databases are created. The acted speech is usually read and not spoken spontaneously. The speakers are made to speak with different stress or emotions and the speech is recorded. Many existing speech database are self -recorded and are not easily available. However, speech under simulated and actual stress (SUSAS) and simulated stress speech database (SSD) are accessible for researches and hence they are most frequently used for stress detection. Thus, the details of these databases are elaborated.

3.1. Speech under simulated and actual stress (SUSAS)

Speech under simulated and actual stress database was created in the university of Colorado-Boulder and it includes speech samples with stress and emotions[15]. The creation of database was directed by Professor John H. L. About 16000 utterances were taken from 32 speakers of age 22 to 76 years of which 13 are female and 9 are male. There are five domains in this database. First domain is based on talking styles and include angry, soft, loud, clear, question, fast and slow. Second domain is based on Psychiatric data comprising speech under depression, anxiety and fear. Third domain includes the noisy speech causing Lombard effect. Other two domains are based on fear tasks and computer response tasks. The database is made up of aircraft communication words which were obtained under stressed conditions.

3.2. Simulated Stressed Speech Database (SSD)

The SSD consists of 33 Hindi keywords recorded from fifteen adult speakers of which 10 are male and 5 are female. The speakers were non-professionals. The database was created by Sumitra Shukla. It considered angry, happy, neutral, sad and lombard stress conditions. Noise is played through headphones to record the Lombard speech. The speakers were asked to think about a situation where they can act these stress conditions. Recording were done in two sessions with a one week time gap. The sampling rate at which the speech was recorded was 16 kHz with a sampling resolution of 16 bits/sample. The database includes about 3100 speech files in which 620 files were present for each stress class[16]. The ability of automatic stress recognition system to identify stress was studied first.

4. Feature Extraction and Classification

The review analysis of stress detection from speech signal is carried out for a period of 20 years starting from 1996 to 2016. The evolution of different features, features extraction methods and classification techniques are studied. The idea about how different features contribute to stress and how the stress conditions are classified are being analyzed from the survey. Some of the bottlenecks associated with stress detection are also understood from the survey.

4.1 Work on SUSAS

The researches done using SUSAS database are discussed first. This database was created to help researchers to analyze, model and develop new speech algorithms for addressing the stress and emotion related issues. The classification rate that can be achieved using SUSAS database with different stress conditions are also discussed in this section.

The work of *Hansen et al*. in [17] conducted using SUSAS database found out that the involvement of stress in speech causes changes in speech produced and as a result, there is degradation in the performance of the speech processing algorithms used. The parameter extracted were mel, cross correlation mel, delta mel, delta-delta mel and the autocorrelation mel parameters. The stress conditions considered were angry, clear, fast, loud, lombard, normal, cond50, cond70, soft and slow. Cond50 and cond70 are

the speech taken when a high workload computer response task is performed. Classification of features was done using a neural network classifier. The illustration of how stress affect the speech production model is done by visualizing the shape of the vocal tract, analyzing the area of acoustic tube and by observing the variations in the speech parameters. The autocorrelation mel was discovered to be one of the most useful feature for separating stress conditions and a classification rate 79% for in-vocabulary and 46% for out of vocabulary test set were achieved.

Tin Lay New et al. in [18] used angry, loud, neutral, clear and Lombard stress conditions. They considered both linear and non-linear features of speech signal. The linear feature used is the short time log frequency power coefficients (LFPC) and non-linear features used are time domain as well as the frequency domain LFPC features. The features are extracted. The classifier used was HMM model containing two Gaussian mixtures for every state. The results obtained revealed the highest average accuracy of 85% for LPFC features. Among the non-linear features the frequency domain LFPC features gave better performance than the time domain LFPC feature.Unlike earlier stress categories in [17] and [18], the work done by *Ling He et al.* in [19] used SUSAS database for high, low and neutral stress conditions. Non-linear Teager energy operator features are calculated from the bands of discrete wavelet transform (DWT), critical bands and then the wavelet packet bands. Probabilistic Neural Network(PNN) together with Multilayer Perceptron Neural Network(MLPNN) were used to classify the different stress classes. The TEO, perceptual wavelet packet analysis along with PNN classifier achieved the best performance score of 93.67%.
An automatic stress recognition system was proposed by *Salsabil Besbes* and *Zied Lachiri* in [20] which was based on kernel classification. The neutral , lombard, angry and loud classes were used in the study. The acoustic features along with Gammatone Frequency Cepstral Coefficients were extracted. The prosodic and spectral features which include MFCC, PLP, LPCC, Pitch and energy was derived. The multiclass SVM methods based on One Against One (OAO), Directed Acyclic Graph (DAG) and One Against All (OAO) using linear, Gaussian and polynomial kernel were employed. The experimental result conveyed that the Gaussian kernel rendered the best performance with an accuracy of 98.12% for OAA and 98.79% for DAG. They also mentioned that the same experiments can be conducted in future by using one class SVM.

The reseachers *G .Senthil Raja* and *S. Dandapat* in [21] extracted six features from database comprising angry, question, Lombard and neutral stress conditions. The features are MFCC, Reflection Coefficients (RC), Linear Prediction Coefficients (LPC), Arc-Sin Reflection Coefficients(ARC) , Linear Prediction Cepsral Coefficients (LPCC) and Log Area Ratios (LAR). These features are largely used in speaker recognition. The speaker recognition results are evaluated using Gaussian mixture model and VQ classifier. The attainment of speaker recognition is enhanced by using four compensation techniques. The compensation techniques used are Compensation by the Removal of Stressed Vectors (CRSV), Speaker and the Stressed Information based Compensation(SSIC), Combination of both MFCC and Sinusoidal Amplitude features (CMSA) and Cepstral Mean Normalization (CMN). The maximum average classification rate of 92.57 is achieved for MFCC, LAR and ARC using vector quantization classification and CMSA compensator. Maximum speaker identification(SI) rate of 85.7% was achieved for neutral condition and 60.31% for question speech. For angry speech 30.15% is the maximum SI result achieved using SSIC and for Lombard , it is 52.38% using CRSV. F-ratio values were also computed to evaluate the speaker and also the stress information.

The work of *H .Patro et al.* in [22] deals with the evaluation of sinusoidal frequency features (SFF), the Mel Frequency Cepstral Coefficients(MFCC), sinusoidal amplitude feature (SAF) and Cepstral Coefficients (CC). These features were extracted using the speech under simulated emotion (SUSE) corpus. Inorder to evaluate which feature can distinguish the feature classes more accurately and precisely, statistical analysis is performed. The statistical measures used are Kolomogorv-Smirnov (KS) test, F-ratio test, probability density characteristics and feature discrimination measure (FDM). Vector quantization and Gaussian mixture models are used for classification. The results showed that SAF achieved maximum recognition rate for both classifiers (83.57% for GMM and 87.18% for VQ) and the features SFF, MFCC and CC followed. The statistical techniques used exhibited almost the same performance for FDM and KS test and for F-ratio test, SFF achieved the best performance.

4.2 Work on SSD

A wide research on stress detection is done using simulated stress speech database also. An illustration of the works done using SSD database and the classification accuracy achieved are discussed. The features that can be extracted from this database and the challenges encountered while using this database are also understood.

Sumitra Shukla et al. investigated the features based on spectral slope for the stressed speech classification [23]. They used simulated stress speech database with Lombard, neutral, angry and sad classes. The tilt of the spectrum is observed for each stress class and the relative displacement in the formant peak is derived from the cepstrally smoothed log spectrum and LPC. Then the results are compared with the MFCC features. The rank level, feature level and score level feature combination techniques are used to combine the features. The classification result showed that the MFCC achieved a performance rate of 53.15%, RFD taken from LPC and the cepstrally smoothed spectrum achieved 51.66 and 52.40% performance rate respectively. Thus RFD and MFCC were equally capable to discriminate stress. The combination of cepstrally smoothed log spectrum derived RFD and MFCC achieved the maximum average classification rate of 59.53% indicating an increase in performance when combining techniques are used. The authors in [24] also analyzed how the human and automatic stressed speech processing tasks is affected by stress. Thirteen MFCC features were extracted from speech under each stress conditions. The average performance of 59.44% was achieved for human stress classification and for automatic stress classifier, 54.65% was achieved for VQ and 56.02% for HMM. For stressed speech recognition, 99.60% performance is achieved by human stressed speech recognition and 82.42% and 76.79% for VQ and HMM in automatic stressed speech recognition. The deterioration in the performance of automatic processing system raise the necessity for the progress of new techniques to handle the information of stress.

Suman Deb and *Samarendra Dandapat* in [25] explored how the breathiness components can affect the speech under stress. The stress conditions angry, neutral, Lombard, sad and happy were considered. The features extracted includes Glottal to the Noise Excitation Ratio(GNER), the Harmonic Energy(HE), the Harmonic Energy of Residues(HER), Amplitude Perturbation Quotient (APQ), Period Perturbation Quotient (PPQ), Harmonic to Noise Ratio(HNR) and Harmonic to Signal Ratio(HSR). Hidden Markov Model was used for classification with 80% of the utterance for training and remaining 20% for testing the model. From the experimental result, it was observed that the breathiness features gave a classification rate of 59.4% and MFCC feature gave a classification rate of 66% and both of them when combined together gave a better classification rate of 72.8%. A better performance for breathiness feature for the class angry, happy and lombard compared to sad and neutral were noticed. The same authors also proposed the classification of stressed speech using Harmonic Peak to Energy Ratio (HPER) in [16]. HPER is a new feature that has the capability to characterize the breathiness levels

and hence stress conditions in speech signal. The same stress conditions mentioned in [20] were considered. This feature was analyzed with other MFCC, TEO-CB and TEO-BB-Auto-Env features and LPC features. Statistical measures were used to estimate the discrimination capability of HPER among different stress conditions. The F-score and t-score are the statistical measures that are computed using the mean and variance estimated from the pdf characteristics of HPER features. Support vector machine and Binary cascade multiclass classification approach were used for classification. Classification was performed using five-fold cross validation method. The confusion matrix for different features was obtained and maximum accuracy (88%) was obtained using the grouping of MFCC and HPER features. The HPER, LPC, MFCC and TEO-CB-Auto-Env features attained an individual accuracy of 84.6%, 64.6%, 81.4% and 67.6% respectively. Thus, this work established the potential of HPER feature for stress speech classification.

The design of the evolutionary algorithm for searching an optimum filter bank was proposed by *Leandro D Vignolo et al.* in [26]. The spline function was used to shape the filter bank in chromosome codification, so that the chromosome instead of holding the parameters of filter bank, will hold the spline parameters. This approach gives Evolutionary Spline Cepstral Coefficients (ESCCs). The stress speech database and the FAU Aibo emotion corpus were used for experiment. Classification was done using SVM using polynomial kernel in which 80% of the instances were used for training and remaining 20% for evaluation. The results obtained for final classification test for Hindi corpus was higher (91.31%) than FAU Aibo corpus (42.50%). The study did not consider the impact of noise on filter bank shape and was confined to speech signals that were clean.

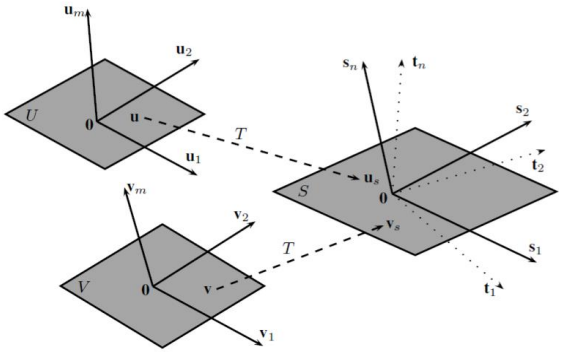


Fig.3: Linear Transformation on Higher Dimensional Speech Subspace S where U is the neutral speech and V is the stress speech subspace [27]

Bhanupriya and *S. Dandapat* in [27] stated that there exists a speech subspace that contains the properties of neutral speech signals as well as stressed speech signal. The experiment was conducted for neutral , happy, sad, Lombard and angry stress conditions. The neutral and stress speech are parameterized as the non-linear TEO-CB-Auto-Env feature. The speech signal holding neutral and the stress conditions are described and analyzed using a linear transformation as shown in Fig.3. The deviation in the speech signal properties for both the neutral and stress conditions are observed on a subspace whose dimension is higher compared to the subspace dimension of original neutral and stress speech. The parameterization of speech signal was done using bandpass Gabor filter bank in which the Gabor filter is a Gaussian modulated cosine pulse. The linear transformation matrix is created using supervectors that are obtained from the HMM model. Each state of HMM is modelled using Gaussian mixtures. A total of 264 supervectors were obtained and PCA with gaussian, polynomial and exponential kernel was used to determine the orthonormal supervectors. The results obtained indicate that stress compensation can be achieved using linear transformation on speech subspace. They also concluded that speech subspace and stress speech subspace are related linearly.

4.3 Work on Other Databases

Apart from SUSAS and SSD, other databases which are self-recorded are also used in few researches. Since these databases are not easily accessible, they are not widely used. Experiments are also conducted by combining other modalities with speech. The work done using other database and combined modalities are also discussed here.

Laszlo Czap and *Judit Maria Pinter* conducted experiments using Hungarian Speech Database which consists of read text that was recorded in average user environment like offices, home and laboratory with a sampling rate of 16 kHz and resolution 16 bits/second. They considered suprasegmental features[28] to show whether the nature of syllable is stressed or unstressed. The suprasegmental features used are tone, intonation, speech rate, pause, rhythm and tonality, volume and stress and the word stress and sentence stress can be considered using these features. The energy of the syllables were used for stress detection. The energy of stationary state of a vowel is used to denote the energy of a syllable. The HMM is trained using the database and from the first state, energy of reference vowel was derived. Feature extraction gives the current energy of vowel. The ratio of the current vowel energy to the reference vowel energy gives the relative intensity. The stressed and unstressed nature of a syllable was understood by comparison of the amplitude of actual vowel with that of average vowel. The results showed that the method have to be studied further.

The researchers *Hindra Kurniawan et al.* in [4] conducted the experiment by recording speech signals, facial expression and the resistance offered by the skin called the skin conductance or Galvanic Skin Response (GSR).



(a) (b)

Fig. 4: (a) The wire connector sensor (b) GSR device [4]

The speech features extracted were pitch, energy, MFCC, and then the Relative Spectral Transform Perceptual Linear Perception (RASTA-PLP). The GSR is obtained using a sensor as shown in Fig.4. The speech and GSR features were fused together and the class output was obtained using the classifier. Four classifiers, namely SVM, decision tress, K-means and GMM were used.The SVM classifier outperformed all other classifiers by giving an accuracy of 92% for speech features and 70% for GSR features. Thus they derived a conclusion that speech is the better indicator of stress than the GSR.

Victoria Rodellar Biarge et al. in their work in [29] uses a different database. The database used in their work is obtained from the recordings of individuals with contradictory versus self-consistent opinions [30]. Contradictory opinions are fabricated artificially and self-consistent opinions expresses thoughts and feelings naturally. They mentioned that the neural activity can be related to the voice that may cause alteration in the production of speech. Features extracted involved the estimation of biomechanical, glottal and acoustical properties of voice. Statistical approaches like relative entropy, Receiver Operator characteristics (ROC), student's t-test and Wilcoxon are used to

achieve better feature selection. The feature reduction was carried out by Principal Component Analysis (PCA) and classification is done by SVM using linear, Radial Basis Function (RBF), quadratic and polynomial kernel. The results obtained proved tremor to be the most significant parameter to characterize the stress.

Table I : Existing works on stress detection using SUSAS and SSD Database

Database	Authors	Features	Count of stress classes	Classifiers	Accuracy (%)	Best classified stress conditions
SUSAS	J. H. L. Hansen and B. D. Womack[1996]	Mel, delta-mel, CC-mel, AC-mel [17]	11	NN	79.0	Cond50/70,Normal. Soft
	Tin Lay Nwe, Say Wei Foo and L. C. De Silva[2003]	LFPC[18]	5	HMM	85.0	Anger and Neutral
	L. He, M. Lech, N. C. Maddage and N. Allen[2009]	TEO[19]	3	PNN	93.67	Not reported
	Senthil Raja, G. & Dandapat, S[2010]	MFCC, ARC, LAR[21]	4	VQ	92.57	Neutral and Question
	S. Besbes and Z. Lachiri[2016]	Gammatone frequency cepstral coefficients[20]	4	SVM	98.79	Neutral and Angry
SSD	Shukla, S.; Dandapat, S.; Prasanna, S.R.M. [2011]	RFD+MFCC[23]	4	HMM	59.53	Sad
	S. Shukla, S. R. M. Prasanna and S. Dandapat [2011]	MFCC[24]	5	VQ	82.42	Sad and Angry
	S. Deb and S. Dandapat [2015] [2016]	Breathiness+MFCC[25]	5	HMM	72.8	Angry, Happy , Lombard
		HPER+MFCC[16]	5	SVM	88	Sad
	Leandro D. Vignolo, S.R. Mahadeva Prasanna, Samarendra Dandapat, H. Leonardo Rufiner, Diego H. Milone. [2016]	ESCC[26]	5	SVM	91.31	Sad and Neutral
OTHER DATA-BASES	L. Czap and J. M. Pintér[2015]	Suprasgmental Features Vowel energy, Intensity [28]	3	HMM	Not reported	Not reported
	H. Kurniawan, A. V.Maslov and M. Pechenizkiy [2013]	MFCC, RASTA-PLP, Pitch[4]	Not reported	SVM	92.6	Not reported
	Rodellar-Biarge, V., Palacios-Alonso, D., Nieto-Lluis, V., and Gómez-Vilda, P [2015]	Acoustical, glottal and biomechanical parameters[29]	Not reported	SVM	90-Female 80-Male	Not reported

The existing methods of stress detection using SUSAS, SSD and other databases are tabulated in Table I. The accuracy achieved is taken as the performance metric. The stress class Neutral and Anger achieved better recognition rate than other stress classes in majority of the work done using SUSAS [18,20]. In the work done using SSD, in almost all cases, Sad stress condition achieved higher classification rate. The understandings and observations attained from the review and the challenges that arise during stress detection from speech signal are discussed in next section.

5. Discussion and Challenges

From the works discussed so far, it is observed that SUSAS database is capable of giving better classification rate. The best performance for SUSAS was achieved using Gammatone frequency cepstral coefficients [20] and multiclass SVM. The non-linear TEO features calculated using DWT in [19] and classified using PNN and MLPNN also achieved better performance. It is observed from [21] that the use of compensation techniques can increase the Speech recognition rate. From the survey, it is found that the use of spectral features can give better results and SVM can outperform other classifiers. The survey conducted using SSD indicate that the best results were obtained in the work that used evolutionary algorithm to obtain ESCC [26]. It is also observed that the classification rate can be increased by using the combination of breathiness and MFCC features [16,25].

The higher classification achieved in SUSAS may be due to the use of speech under actual stress. In simulated stress speech database, there is a possibility of the speaker being not able to convey the actual stress condition. Also, SUSAS database

contains five domains and consider angry, soft, loud, clear, question, fast, slow, fear, Lombard and anxiety stress conditions while SSD considers only five stress conditions which are angry, happy, Lombard, sad and neutral. Thus it is understood from the review that, speech under actual stress comprising more number of stress conditions can provide better results compared to speech under simulated stress. Also, the performance of the classifier cannot be evaluated from individual research. The accuracy and recognition rate of classifiers also depend on the database, preprocessing techniques and the features used. Thus the features and feature combinations that gives the best result has to be discovered. Stress detection from speech signal face few challenges.

- One of the main challenges is the availability of proper stress database. So far, SUSAS and SSD are the only databases used in most of the work. Thus, there is a need to create more stress speech database that is accessible.
- There still exists non-uniformity in the categorization of stress conditions. The stress condition Angry, Happy, Sad, Lombard, Question, Loud and Clear are generally used for stress detection and other stress conditions like fast, soft, slow and fear are not being investigated much.
- Also, the expression of stressed speech differs based on the origin or nativity. The individuals of certain place may be able to convey the stress through speech more accurately than the people of some other place.
- Another challenge is the need to find the best feature combination that characterizes stress more precisely.

Thus, these challenges in stress detection have to be dealt with proper solutions. Other existing classifiers like deep neural networks, decision tree classifiers and features that can accurately detect and portray the stress conditions in speech have to be discovered. The development of new algorithms like evolutionary algorithm and tools for feature extraction and classification based

on attribute selection that gives better accuracy and better classification rate should be encouraged.

6. Conclusion and Future Scope

The need for the management of chronic stress in individuals raised the concept of stress detection. To facilitate a better understanding of stress detection from speech signal, a detailed survey on different approaches for stress detection from speech signal conducted by different researchers are presented in this paper. Discussion range from the comparison of different database used in the study and the different feature extraction methods used to obtain the features from the speech signal. A detailed survey on emotional speech database is presented in [8].Therefore, only stress related database especially SUSAS and SSD are discussed here. The various statistical measures, feature selection schemes and compensation techniques used are also listed. The use of different classifiers and machine learning algorithms along with the accuracy achieved for each classifier using different features and feature combinations are also provided. Based on the survey, the challenges and gaps in stress detection using speech are understood.

The future directions of stress detection are listed below.

- There is a need to search for other features extraction methods that can be used for stress speech detection. The use of wavelet transform in feature extraction is not being explored much. Most of the researches are based on Fourier transform and the fact that the wavelet transform possess the time localization property make it superior for the analysis of non-stationary signals. The wavelet transforms show good time resolution for high frequency and remarkable frequency resolution for slowly varying functions. Thus, the use of wavelet based transforms such as Complex Wavelet Transform, Multi Resolution Analysis (MRA), Dual wavelet, the Fast Wavelet Transform (FWT), Morlet wavelet, Wavelet Packet Decomposition (WPD) and Wavelet Energy Feature (WEF) need to be explored for feature extraction from stressed speech.
- The use of deep neural networks for stress recognition has to be explored.
- The majority of the reported work is on emotional stress. A future direction could be towards isolated stress detection of high, low and medium stress levels.
- There is a need to discover a proper characterization of stress condition. Thus, the stress condition other than Angry, Happy, Sad, Lombard, Question, Loud and Clear have to be explored. Multimodal stress conditions that give the best accuracy and recognition rate have to be investigated.
- Detection of stress from speech signal by considering the origin and nativity of the speaker. Also stress from multilingual Indian languages can be explored and compared.

Another future scope lies in using a group of parameters such as skin conductance, heart rate variability, finger-body temperature and blood volume pulse along with speech signal for detection of stress

References

[1] H. Gao, A. Yüce and J. P. Thiran, "Detecting emotional stress from facial expressions for driving safety," *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, 2014, pp. 5961-5965

[2] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," *2011 Computing in Cardiology*, Hangzhou, 2011, pp. 85-88.

[3] G. Shivakumar and P. A. Vijaya, "Emotion Recognition Using Finger Tip Temperature: First Step towards an Automatic System," *International Journal of Computer and Electrical Engineering* vol. 4, no. 3, pp. 252-255, 2012.

[4] H. Kurniawan, A. V.Maslov and M. Pechenizkiy, "Stress detection from speech and Galvanic Skin Response signals," *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, Porto, 2013, pp. 209-214

[5] J. Xie, W. Wen, G. Liu, C. Chen, J. Zhang and H. Liu, "Identifying strong stress and weak stress through blood volume pulse," *2016 International Conference on Progress in Informatics and Computing (PIC)*, Shanghai, 2016, pp. 179-182

[6] S.Lalitha, S. Patnaik, T. Arvind, V. Madhusudhan, and S.Tripathi, "Emotion Recognition through Speech Signal for Human-Computer Interaction," in Electronic System Design (ISED), *2014 Fifth International Symposium on*, 2014, pp. 217-218.

[7] Lalitha, S., Geyasruti, D., Narayanan, R., Shravani, M.: "Emotion detection using MFCC and Cepstrum features*"in Procedia Comput. Sci*. 70, 29–35 (2015)

[8] D. Ververidis and C. Kotropoulos, "A State of the Art Review on Emotional Speech Databases", *in Proc. 1st Richmedia Conference, Lausanne*, Switzerland, pp. 109-119, October 2003.

[9] C.N. Anagnostopoulos T. Iliou and I. Giannoukos "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011 " *Artificial Intelligence Review* pp. 1-23, 2012

[10] Hansen, J., Patil, S, "Speech under stress: Analysis, modeling and recognition". *In: Müller, C. (ed.) Speaker Classification 2007. LNCS (LNAI)*, vol. 4343, pp. 108–137. Springer, Heidelberg (2007)

[11] Rashmi Makhijani Urmila Shrawankar Dr. V. M. Thakare "Speech Enhancement using Pitch Detection Approach for Noisy Environment" *International Journal of Engineering Science and Technology (IJEST)* Vol. 3 No. 2 PP. 1764-1769 Feb 2011.

[12] M .P. Kesarkar And P.Rao,"Feature Extraction for Speech Recognition" ,*Credit Seminar Report, Electronic Systems Group*, EE. Dept, IIT Bombay, 2003.

[13] Sunny Sonia, S David Peter, K. Poulose Jacob, "Performance of different classifiers in speech recognition", *International Journal of Research in Engineering and Technology*, vol. 2, no. 4, pp. 590-597, Apr. 2013.

[14] L. He, M. Lech, C. Namunu, et al., “Study of empirical mode decomposition and spectral analysis for stress and emotion classification in natural speech”, *Biomedical Signal Processing and Control 6* (2011) 139–146.

[15] Hansen, J., Bou-Ghazale, S., 1997. “Getting started with SUSAS: A speech under simulated and actual stress database*”. In: Proceedings of the Eurospeech*, Rhodes, Greece, Vol. 5, pp. 2387–2390, 1997.

[16] Suman Deb ,S Dandapat “Classification of speech under stress using harmonic peak to energy ratio” ,*Computers and Electrical Engineering*, Volume 55 Issue C, October 2016, Pages 12-23.

[17] J. H. L. Hansen and B. D. Womack, "Feature analysis and neural network-based classification of speech under stress," in *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 4, pp. 307-313, Jul 1996.

[18] Tin Lay Nwe, Say Wei Foo and L. C. De Silva, "Classification of stress in speech using linear and nonlinear features," *Acoustics, Speech, and Signal Processing*, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on, 2003, pp. II-9-12 vol.2.

[19] L. He, M. Lech, N. C. Maddage and N. Allen, "Neural Networks and TEO Features for an Automatic Recognition of Stress in Spontaneous Speech," *2009 Fifth International Conference on Natural Computation*, Tianjin, 2009, pp. 227-231.

[20] S. Besbes and Z. Lachiri, "Multi-class SVM for stressed speech recognition," *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Monastir, 2016, pp. 782-787.

[21] Senthil Raja, G. & Dandapat, S. “speaker recognition under stressed condition” *Int J Speech Technol* (2010) 13: 141

[22] H. Patro, G. S. Raja and S. Dandapat, "Classification of Stressed Speech using Gaussian Mixture Model," *2005 Annual IEEE India Conference - Indicon*, 2005, pp. 342-346.

[23] Shukla, S.; Dandapat, S.; Prasanna, S.R.M. “Spectral slope based analysis and classification of stressed speech”. *Int. J. Speech Technol*. 2011, 14, 245–258.

[24] S. Shukla, S. R. M. Prasanna and S. Dandapat, "Stressed speech processing: Human vs automatic in non-professional speakers scenario," *2011 National Conference on Communications (NCC)*, Bangalore, 2011, pp. 1-5.

[25] S. Deb and S. Dandapat, ‘‘A novel breathiness feature for analysis and classification of speech under stress,’’ *in Proc. 21st Nat. Conf. Commun. (NCC)* , 2015, pp. 1–5

[26] Leandro D. Vignolo, S.R. Mahadeva Prasanna, Samarendra Dandapat, H. Leonardo Rufiner, Diego H. Milone. “Feature optimisation for stress recognition in speech”, *Pattern Recognition Letters*, Volume 84. 1–7, 2016.

[27] B. Priya and S. Dandapat, "Linear transformation on speech subspace for analysis of speech under stress condition," *2015 Twenty First National Conference on Communications (NCC)*, Mumbai, 2015, pp. 1-6.

[28] L. Czap and J. M. Pintér, "Intensity feature for speech stress detection," *Proceedings of the 2015 16th International Carpathian Control Conference (ICCC)*, Szilvasvarad, 2015, pp. 91-94.

[29] Rodellar-Biarge, V., Palacios-Alonso, D., Nieto-Lluis, V., and Gómez-Vilda, P. “Towards the search of detection in speech-relevant features for stress”, *Expert Systems*, 32: 710–718, 2015.

[30] Frank, M.G. and P. Ekman, “Appearing truthful generalizes across different deception situations”, *Journal of Personality and Social Psychology*, 86, 486–495, 2004.