Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 5s (2025)
https://www.jneonatalsurg.com

Stress Recognition in Speech – A Survey of The State of The Art

L Lavanya[1], N Vasavya[2]

[1]Assistant Professor, Department of CSE(AI & ML), Sri Venkateswara College of Engineering, Karakmbadi Road, Tirupati. Email ID: lavanya.l@svce.edu.in

[2]Assistant Professor, Department of CSE(AI & ML), Sri Venkateswara College of Engineering, Karakmbadi Road, Tirupati. Email ID: vasavya.n@svce.edu.in

.Cite this paper as: L Lavanya, N Vasavya, (2025) Stress Recognition in Speech – A Survey of The State of The Art. *Journal of Neonatal Surgery*, 14 (5s), 793-798.

ABSTRACT

Stress recognition from speech seeks an humongous attention among the researchers and from the industrial sides like call centres for recognizing the customer's intension over speech. Recognizing stress using visual is easier when compared with recognition of stress from speech signal since Lombard effect affects the normal speech heavily. In this paper a detailed survey has been made on the research works that are carried out only to recognize the stress from speech signal. This paper also addresses the databases that are considered only for stress recognition. The speech signals of the databases cited in this paper consists of the speech signals that are only intended to recognize the level of stress from the speech signal. A detailed Table has been cited which holds the core part of each and every research work carried on recognition of speech signal.

Keywords: *Survey, Stress Recognition, Stress Speech Database, Features recognizes stress.*

1. INTRODUCTION

Speech Emotion Recognition (SER) is the process of recognizing the emotions from a speaker's multimedia (voice) data. The applications of SER have a wide range of use including emotion recognition through voice data in call-centre services which helps the concern to decide on the responding authority for the incoming call, on vehicles to extract the emotion of the driver to avoid accident probability. It also can be applied on the medical field where the extract the present condition of patient's disorder and in E-tutors to analyse the mood of the user in order to change the mode of listening.

The features for speech recognition can be categorized into Qualitative, Spectral, Continuous and TeagerEnergy Operator (TEO) based features.With respect to emotional part in a speech, it is highly related with the features like pitch, zero crossing rate and energy which comes under the category of prosody features. These features mentioned above comes under the category which are related in terms of rate of articulation in the speech, total energy in the speech, spectral and fundamental frequency (f_0). The emotions perceived and the overall quality in voice data have high level of impact with the emotions. The quality of voice can be classified as voice level, voicepitch, temporal and feature boundary structures.Spectral analysis-based features come under short representation of time over the speech signal. Over voice data, emotional content has higher level of relation with spectral energy distribution. From all the references in the literature it can be perceived that the emotions such as happiness and anger have higher level with respect to energy bounded with higher frequencies and on the other hand emotions such as sad have less energy bounded with low frequency.

Spectral based features such as Mel-frequency cepstral coefficients (MFCC) and Linear PredictiveCepstral Coefficients (LPCC)are the most common features for emotion detection from speech signal. The base of speech signal is due to the airflow in vocal tract system in non-linear manner. Muscle tension is the another factor that affects the speech sound produced by the speaker when he tends to speech under stressed conditions.

Hence, on addressing this scenario it is evident that the non-linear features of a speech is much essential factor in detecting the emotions. Teager Energy Operator (TEO) proposed by Teager and Kaiser is the evidential proof for recognizing emotions through similar kind of features which is discussed above. Among the emotions, stress plays a vital role in real world scenarios such as the psychological mood of a driver in order to avoid major accidents, psychological state of a person in order to guide them through proper counselling, a child's psychological state to improve their parents and other acquaintance, etc.

A number of approaches were identified to recognize the stress from the raw audio signal after pre-processing. One such dedicated model for recognizing stress from speech signal is TEO-CB-Auto-Env which uses pitch related features for recognizing the stress. They have used HMM and SUSAS database as the sub components for speech recognition. Features such as Critical Bank Filter, Auto correlation and area of the envelope are used for stress recognition.

2. DETAILED ANALYSIS ON STRESS RECOGNITION

In this paper a detailed analysis has been made on the research articles that recognizes stress from the speech signal. Table 1 holds the databases that holds the speech signal that are intended to recognize the stress. It consists of every database name, the language of speech signal it possess, the subjects that are used to generate the speech signal along with their nativity, the classes of stress that can be inferred from it and its references. In Table 2, every contribution made on stress recognition has been addressed in row wise. Every tuple in the table refers to a research contribution.

2.1 Stress Speech Databases

The most important criteria to be taken into account when it comes to stress speech databases is the degree of naturalness in the databases. A wrong choice of database with low quality or immaterial choice may lead to wrong classification process. In some of the databases reported in this section possess some different levels in stress classification [23].

2.1.1. Real world Stress Vs Simulated:

Usage of real-world speech data is highly recommendable in classification of speech data. However, it is not legal and moral to use such speech data since it is prohibited for usage of research purpose. That too when it comes to stress related data the legal and moral as well as personal concerns are high when compared to other speech datasets. On the other hand, these stress emotions can also be depicted from the sound laboratories. These speech data are there in most of the databases which are listed here. In [24] it is found that the acted emotions are highly exaggerated when it compared with the real emotions. In most of the databases, the emotions will be expressed by the actors who are well versed in their own phenomena of producing emotions in speech. The databases are further cbe classified into to types, local language and common language. English is one of the common languages and it will be always be depicted by the native speakers. In some databases non-native speakers also been used to reduce the exaggeration in speech.

Table 1: Collection of Databases on Stress

Database Name	Language	Subjects	Stress Classes	Reference
SUSAS	English	32 Various	Lombard effect, Stress, Task based Stress	[1]
SUSC-0	English	18 Non-native	Stress	[1]
SUSC-1	English	20 Native	Stress	[1]
DLP	English	15 Native	Stress	[1]
ORESTEIA	English	29 Native	Stress	[2]
SOQ	English	6 Soldiers	5 Stress Levels	[3]
Lost Luggage	Various	109 Passengers	Stress	[4]
	English, German	100 Native	Stress, Task based Stress	[5]
	English	4 Drivers	Task based Stress	[6]
Cockpit	English	8 aircraft pilots	Task based stress	[7]
EMO-DB	English	six female and male speakers	Stress	[11]

2.2 Features for Stress Recognition

Features plays a vital role in stress recognition from speech data. Extraction of liable features from the raw speech data which will be the key feature for recognizing stress from it are all addressed as a major issue in stress recognition. Via patter recognition schema it will be achieved in some areas of research. However, since pattern recognition acts independent with the nature of the problem and problem domain, proper features are to be denoted in prior to the recognition of stress.

There are four major issues that are considered in stress recognition the first is the level of analysis that has been made for feature extraction. The second is which type of features are to be extracted. For example, pith, frequency, zero crossing rate, etc. the thirst analysis comes with the pre and post processing steps followed for feature extraction. And the final issue states which type can be integrated with acoustic for improving the overall performance of classification.

Author	**YEAR**	**Inference**	**Features used**	**Database**	**Classifiers used**	**Compared Algs.**	**Performance Measures**	**Ref**
Scherer, K.R	**2003**	Use of features derived from multi resolution analysis of speech and TEO for classifcation of driver's speech under stressed conditions	**TEO, Features at utterance level**	**[6]**	**HMM**	FHMM ARHMM HMDT HMM M-HMM SVM ANN	**% of errors**	**[6]**
Rahurkar, et.al	2002.	Weighted TEO band critical frequencies are used	TEO-CB-AutoEnv	**[9]**	**HMM**	self	**% of errors**	**[9]**
Besbes, et al.	2017	Comparison of Effective features among MFCC, GFCC	MFCC, GFCC	**SUSAS**	OAO, OAA and OC-SVM	Self	**% of Accuracy**	**[10]**
Surekha, et al.	2017	Feature Fusion	TEO-CB-Auto-Env with MFCC	**SUSAS,** Berlin German database	**GMM**	Self	**% of Classification accuracy**	**[11]**
Yogesh, C. et al.	2017	Inferred 3rd order derivation of bispectral features	28 BispectralFeatrues	**SUSAS, BES**	**ELM kernel, KNN, PNN, GRNN,**	Self with different classifiers	**% of recognition rate**	**[12]**
Yogesh, C., et al.	2017	Inferred 3rd order derivation of bispectral features	28 BispectralFeatrues	**SUSAS, BES**	**ELM kernel, KNN, PNN, GRNN,**	BBO, PSO	**% of recognition rate**	**[13]**

Author	**YEAR**	**Inference**	**Features used**	**Database**	**Classifiers used**	**Compared Algs.**	**Performance Measures**	**Ref**
Yogesh, C, et al.	2017	Inferred 3rd order derivation of bispectral features	28 BispectralFeatures	**SUSAS, BES**	**ELM kernel, KNN, PNN, GRNN,**	BBO, PSO	**% of recognition rate**	**[14]**
Sahar E. , et al.	2000	Examining the impact of Linear power Spectrum and Fourier Transform	MFCC Features	**SUSAS**	**-**	Self	**% of recognition rate**	**[15]**
Sahar , et al.	1998	Modelling HMM to recognize stress using the features such as duration of speech	Pitch Contour, Duration of Speech, Spectral Structure Average.	**SUSAS**	**HMM**	Self	**% of Accuracy**	**[16]**
Aswathi, et al.	2017	Representation of speech in terms of sparse is considered for finding stress in speech	**TEO-CB-Auto-Env**	**IIT Gauhatti**	**WEKA**	Other Classifiers	**% of accuracy**	**[17]**
Zhou, G., , et al.	1998	TEO-CB-Auto-Env	**TEO based Features**	**SUSAS**	**HMM**	MFCC	**Classification rate**	**[18]**
Author	**YEAR**	**Inference**	**Features used**	**Database**	**Classifiers used**	**Compared Algs.**	**Performance Measures**	**Ref**
Yakoumaki, , et al.	2014	extended adaptive Quasi-Harmonic Model – eaQHM id proposed to classify emotional speech	**Amplitude and Frequency**	**SUSAS**	**HMM, GMM**	Self	**Classification rate**	**[19]**
Reddy, G. G.	2004	Each speech signal is recognized either as noisy or harmonic	**Pitch**	**SUSAS**	**HNM**	Self	**Classification rate**	**[20]**

		and classification is made						
McAulay, , et al.	1986	The speech signal is considered as sinusoidal model and analysis on speech recognition has been carried out	**Pitch, Amplitude and Frequency**	**Own**	**-**	Self	**Classification rate**	**[21]**
Sagayama,, et al.	1981	Frequency and intensity of speech are recognized for modelling respective speech form	**Frequency and intensity of every sine model**	**Self**	**-**	Self	**Classification rate**	**[22]**

3. DISCUSSION AND CONCLUSION

In this paper a detailed analysis on existing methodologies proposed for addressing stress recognition from speech signal has been carried with the intension of finding the impact of stress recognition and its limitations. From Table 2, it is found that in Recognition of Stress with different contours, it has the potentiality to differentiate between Stress and Lombard Effect using the combination of the speech signal frequency, amplitude and phase. However, it does not possess the capability to differentiate between other speech models such as Happiness, anger, sad etc. Meanwhile in TEO, differentiation between Stress and other models of speech is possible but differentiation between Stress and Lombard Effect is still lacking due to the similarities in terms of Pitch with Lombard Effect and Stress.The future work can be enhanced with a model that addresses both the limitations of sinusoidal and TEO.

REFERENCES

[1] Hansen, J.H.L., 1996. NATO IST-03 (formerly RSG. 10) speech under stress web page. Available from: http://cslr.colorado.edu/rspl/stress.html

[2] McMahon, E., Cowie, R., Kasderidis, S., Taylor, J., Kollias, S., 2003. What chance that a DC could recognise hazardous mental states from sensor outputs? In: Tales of the Disappearing Computer, Santorini, Greece.

[3] Rahurkar, M., Hansen, J.H.L., 2002. Frequency band analysis for stress detection using a Teager energy operator based feature. In: Proc. Internat. Conf. on Spoken Language Processing (ICSLP ’02), Vol. 3, pp. 2021–2024

[4] Scherer, K.R., 2000b. Emotion effects on voice and speech: paradigms and approaches to evaluation. In: Proc. ISCA Workshop on Speech and Emotion, Belfast, invited paper.

[5] Scherer, K.R., 2003. Vocal communication of emotion: a review of research paradigms. Speech Comm. 40, 227–256.

[6] Fernandez, R., & Picard, R. W. (2003). Modeling driver’s speech under stress. *Speech Communication, 40*, 145–159.

[7] Luig, Johannes, and Alois Sontacchi. "A speech database for stress monitoring in the cockpit." *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 228.2 (2014): 284-296.

[8] Sabo, Róbert, and Jakub Rajčáni. "Designing the database of speech under stress." *Journal of Linguistics/Jazykovednýcasopis* 68.2 (2017): 326-335.

[9] Rahurkar, Mandar A., et al. "Frequency band analysis for stress detection using a Teager energy operator-based feature." *Seventh International Conference on Spoken Language Processing*. 2002.

[10] Besbes, Salsabil, and ZiedLachiri. "Classification of speech under stress based on cepstral features and one-class SVM." *2017 International Conference on Control, Automation and Diagnosis (ICCAD)*. IEEE, 2017.

[11] http://emodb.bilderbar.info/index-1024.html

[12] Yogesh, C. K., Hariharan, M., Yuvaraj, R., Ngadiran, R., Yaacob, S., &Polat, K. (2017). Bispectral features and mean shift clustering for stress and emotion recognition from natural speech. *Computers & Electrical Engineering*, *62*, 676-691.

[13] Yogesh, C. K., Hariharan, M., Ngadiran, R., Adom, A. H., Yaacob, S., &Polat, K. (2017). Hybrid BBO_PSO and higher order spectral features for emotion and stress recognition from natural speech. *Applied Soft Computing*, *56*, 217-232.

[14] Yogesh, C. K., Hariharan, M., Ngadiran, R., Adom, A. H., Yaacob, S., Berkai, C., &Polat, K. (2017). A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal. *Expert Systems with Applications*, *69*, 149-158.

[15] Bou-Ghazale, S. E., & Hansen, J. H. (2000). A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Transactions on speech and audio processing*, *8*(4), 429-442.

[16] Bou-Ghazale, S. E., & Hansen, J. H. (1998). HMM-based stressed speech modeling with application to improved synthesis and recognition of isolated speech under stress. *IEEE Transactions on Speech and Audio Processing*, *6*(3), 201-216.

[17] Aswathi Varsha K T K, S.Lalitha, “Stress Recognition using Sparse Representation of Speech Signal for Deception Detection Applications in Indian Context” IEEE Conference, 2017.

[18] Zhou, G., Hansen, J. H., & Kaiser, J. F. (1998). A new nonlinear feature for stress classification. In *Third IEEE Nordic Signal Processing Symposium*.

[19] Yakoumaki, T., Kafentzis, G. P., &Stylianou, Y. (2014). Emotional speech classification using adaptive sinusoidal modelling. In *Fifteenth Annual Conference of the International Speech Communication Association*.

[20] Reddy, G. G. SPEECH ANALYSIS-SYNTHESIS FOR SPEAKER CHARACTERISTIC MODIFICATION.

[21] McAulay, R., &Quatieri, T. (1986). Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *34*(4), 744-754.

[22] Sagayama, S., &Itakura, F. (1981). A composite sinusoidal model applied to spectral analysis of speech. *Electronics and Communications in Japan (Part I: Communications)*, *64*(2), 1-10.

[23] M. You, C. Chen, J. Bu, J. Liu, J. Tao, Getting started with susas: a speech undersimulated and actual stress database, in: EUROSPEECH-97, vol. 4, 1997, pp. 1743–1746.

[24] C. Williams, K. Stevens, Emotions and speech: some acoustical correlates,J. Acoust. Soc. Am. 52 (4 Pt 2) (1972) 1238–1250