Performance of the Vocal Source Related Features from the Linear Prediction Residual Signal in Speech Emotion Recognition

Rajesvary Rajoo[1, 2] and Rosalina Abdul Salam[1]
[1]*Faculty of Science and Technology, Universiti Sains Islam Malaysia, Bandar Baru Nilai, Negeri Sembilan, Malaysia*
[2]*Faculty of Science and Technology, Nilai University, No 1, Persiaran Universiti, Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia*
rajes_e@nilai.edu.my

***Abstract*—Researchers concerned with Speech Emotion Recognition have proposed various useful features associated with their performance analysis related to emotions. However, a majority of the studies rely on acoustic features, characterized by vocal tract responses. The usefulness of vocal source related features has not been extensively explored, even though they are expected to convey useful emotion-related information. In this research, we study the significance of vocal source related features in Speech Emotion Recognition and assess the comparative performance of vocal source related features and vocal tract related features in emotion identification. The vocal source related features are extracted from the Linear Prediction residuals. The study shows that the vocal source related features contain emotion discriminant information and integrating them with vocal tract related features leads to performance improvement in emotion recognition rate**

***Index Terms*—Linear Prediction Analysis; Speech Emotion Recognition; Vocal Source Features; Vocal Tract Features**

I. INTRODUCTION

Voice is the most important sound produced by our auditory environment and speech is a complex and abstract use of voice [1]. Genuine emotional eruptions produce physiological changes which in turn affect speech production [2]. A prevalent view states that identifying these emotional states from speech as accurately as possible is a challenging task and has been an area of research for several decades. Many researchers are exploring this area due to its promising applications such as in man-machine interaction and in health and psychological related applications [3, 4].

Speech Emotion Recognition (SER) scheme aims to assign a label from defined emotion classes for the emotional state of an individual from his or her speech. The two main activities in SER are the extraction of an appropriate set of discriminant features and the development of an efficient classification algorithm [4]. In view of this, many approaches have been developed to extract relevant features from speech signals. The goal of SER is to make the human-computer interaction as natural as possible [5].

Speech is produced by a source signal generated in the throat, which is filtered by vocal tract cavities (convolution of time-varying vocal tract system and vocal source) [6, 7]. These source and filter components are to be separated from the speech signal in order to characterize and model these components independently. The techniques for modeling and parameterizing the vocal tract system are well-established, and a majority of feature extraction schemes rely on this. However, relatively little effort has been put in the case of the vocal source. This could be due to the popularity of the vocal tract related features and the complexity in characterizing the source signal [7]. However, since extracting and combining features from the vocal tract system does not bring in much significant improvement in speech processing tasks, the focus has therefore moved towards parameterizing the excitation source [8].

The source related features can be computed directly from the speech signal or can be extracted from the Linear Prediction (LP) residual signal [8], which can be processed in time, frequency, cepstral or time-frequency domain. However, processing the LP residual in the time domain has the advantage over the others, as the artifacts of digital signal processing in the other domains will be negligible [9].

There are several studies in literature that have demonstrated that source related features from LP residual signal contain speech, speaker, language and emotion-related information and they have been used in various speech processing tasks. This is presented in Table 1.

Table 1
Source Related Features in Speech Processing Tasks

Author	Source Related Features	Tasks
Drugman et al. [8]	10 Excitation base features (EBF)	Speech Recognition
Drugman et al. [10]	Source-related features	Voice Activity Detection
C. Hanilci and F. Ertas [11]	LPRC	Speaker Verification
Yegnanarayana et al. [12]	Glottal closure instants	Speech Enhancement
Nurminen et al. [14]	F0, voicing, energy, and harmonic amplitudes.	Speech Synthesis
Gangamohan et al. [15]	F0, SoE, Energy Ratio	Discrimination of Anger and Happy
Al-Talabani et al. [18]	MFCC, LPCC and WOCOR	Emotion Recognition

Concerning SER studies in general, vocal tract related features along with their different combinations are commonly examined in the literature. Numerous techniques have been developed and used to extract appropriate vocal tract features, related to emotions from speech over the years. The two most popular vocal tract feature extraction methods are Mel Frequency Cepstral Coefficients (MFCC) and Linear Prediction Cepstral Cepstral Coefficients (LPCC) [7]. These features are combined with prosodic features related to fundamental frequency (F0), energy and speaking rates to form feature vectors [8, 11, 13, 15, 16, 17].

However, it is worth noting that very few studies explored the significance of source related features in emotion recognition. Among the studies, Kadiri et al. [5], proposed sub-segmental features related to excitation source information (F0, a strength of excitation and energy of excitation) to develop an emotion recognition system. The study shows that there is useful emotion-related information in the excitation source features. In [13], Rao et al. used excitation source information around the GCI region, emotion-specific information from epoch parameters, GVV signal and GVV parameters for characterizing sad, anger, happiness and neutral emotions present in speech using Gaussian mixture models (GMM). The study reveals that about 42% to 63% of average emotion recognition performance is obtained when using different excitation source features. Finally, in [18] a set of features include LPCC and MFCC extracted from LP-residual samples and Wavelet Octave Coefficient of Residual (WOCOR), is proposed by Al-Talabani et al. in their study as vocal source related features. The proposed set of features is used in Support Vector Machine (SVM) and Artificial Neural Network (ANN) and tested on Kurdish, Berlin and the Aibo databases. The experiments demonstrate that the fusion of the proposed vocal source features with the common LPCC and MFCC can achieve better recognition accuracies.

Evidence from previous studies indicates that vocal source features contain emotion discriminant information. To strengthen the argument, the present study is proposed i) to ascertain that the proposed vocal source related features carry discriminant information on emotion; ii) to assess the relative performance of source and filter related features in emotion recognition and, iii) to optimize the performance of the extracted features by using a combination of both. The outcome of this study can be used as a basis for furthering exploration on the emotion-specific information present in the residual of a speech signal in detail.

This paper is organized as follows: Section II describes system methodology. Section III provides details of feature extraction, database and classifier used. The experimental setup is described in section IV. Results and discussion are presented in section V. Finally, section VI provides a summary of the study and scope of future work.

II. METHODOLOGY

Motivated by the source-filter model [6], this study proposed an SER system based on the joint use of vocal tract related features (VTF) and vocal source related features (VSF). Figure 1 shows the overview of the system methodology.

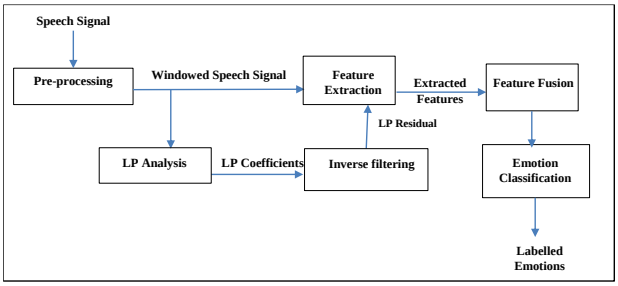

Figure 1: Overview of The System Design

The speech signals were pre-processed in which the process of pre-emphasis, framing and windowing would take place. In this study, the pre-process was done using Hamming Window on signal frames of length 20ms and overlapping of 10ms [18]. While the VTF was extracted from the windowed speech signal, the VSF was extracted from Linear Prediction (LP) residual signal. This LP residual signal was obtained by inverse filtering of the speech signal using its autoregressive parameters computed by the Linear Prediction Analysis.

The Linear Predictive model assumes a speech sample at any given instant, can be approximated as a linear combination of the p past samples, or

$$\hat{s}(n) = \sum_{i=1}^{p} a_i s(n-i) \quad (1)$$

Here ŝ(n) is the prediction of s(n), s(n-i) is the i-th step previous sample, ai is the i-th LP coefficient and p is the number of LP coefficients. The difference between the actual and predicted sample is defined as the prediction error or residual, which is given by:

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i) \quad (2)$$

The linear prediction coefficients $\{a_i\}$ are usually determined by minimizing the mean squared error over an analysis frame. The coefficients can be obtained by solving the set of p normal equations using the autocorrelation function given by:

$$\sum_{i=1}^{p} a_i R(n-i) = -R(n), \quad (3)$$

$$R(i) = \sum_{n=0}^{N-(p-1)} s(n)s(n-i), \quad (4)$$

in which R(i) is autocorrelation function [20].

We can view the computation of the error as a filtering process. The residual signal (e (n)), shown above is obtained by passing the speech signal through the inverse filter A (z), is given by:

$$E(z) = X(z)A(z) \quad \text{where} \quad A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i} \quad (5)$$

Setting correct predicative analysis order is important when estimating LP residual parameters. For the low order of prediction, the residual signal will still have significant information about the vocal tract system. However, if the analysis order is increased, the discriminative power of a residual signal is reduced. Experiments conducted in previous studies show that LP order in the range of 8 to 16

seems to be appropriate for a speech signal sampled at 8 kHz. A spectral envelope can be sufficiently fitted with this range of order and the LP residual mostly contains the vocal source information [19, 20].

III. Feature Extraction, Emotional Speech Databases and Classification

Feature extraction is an essential component in SER systems. It involves the extraction of the parameters which can best reflect the feature of emotion from the speech signals. Numerous features are extracted and used in SER. Some of the widely explored features are Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficients (LPCC), formants, energy, fundamental frequency and zero crossing rate [21]. For the purpose of this study, we proposed a set of vocal tract related features (VTF) and a set of vocal source related features (VSF)

A. Vocal Tract Related Features (VTF)

Five types of VTF were considered in this study: MFCC, LPCC, Zero Crossing Rate (ZCR), Pitch (F0) and Energy. These features have been used widely in SER systems and have been demonstrated to be useful indicators of emotions [20, 21, 22].

1) MFCCs: Most speech recognition systems are based on MFCCs. Their design imitates the non-linear characteristics of the human auditory system. Fast Fourier Transform (FFT) algorithm is ideally used for converting each frame of samples from the time domain into the frequency domain. For the purpose of this study, the usual 12 MFCC coefficients were used [21].

2) LPCCs: LPCCs are another spectral representation of speech signal which typically has a crucial impact on speech quality. They are estimated by using Linear Prediction analysis according to the speech source-filter model. In this experiment, we used 12 LPCC [21].

3) Energy: The intensity of a voice can be physically detected through the pressure of sounds. It can simply be computed by summing the square of the amplitude of the signal within the time window [20].

4) ZCR: ZCR is a duration-related feature that represents the number of times the speech signals are crossing the zero points. It is calculated as the weighted average of the number of times the speech signal changes sign within the time window [20].

5) Pitch: Pitch: Pitch or fundamental frequency (F0) is a useful feature for emotion recognition as different emotions exhibit varying vibration rates of the vocal. In this study, the pitch related features were calculated by using autocorrelation algorithm [21, 22, 23].

B. Vocal Source Related Features (VSF)

For the purpose of this study, a set of vocal source related features, which comprises of MFCC of vocal source (MFCCoVS), LPCC of vocal source (LPCCoVS), ZCR of vocal source (ZCRoVS), Energy of vocal source (EoVS) and F0 of vocal source (F0oVS) were taken into consideration. This VSF were computed from residual signals as shown in the methodology section above. These features were scaled appropriately to ensure that their components have at least similar variances [5, 7, 9, 13, 17].

C. Emotional Speech Databases

In general, SER researches are conducted using two types of speech corpora; acted speech and spontaneous speech. Studies with spontaneous speech seem more realistic. However, these databases do not contain all emotions and the low quality of the speech signal can be a problem. Besides that, legal and privacy issues also become the other factors that influence the SER studies to concentrate more on acted speech databases. Some of the commonly used emotional speech databases in SER studies are Berlin Emotional Database (EMO-DB), Danish Emotional Database (DES), and Speech Under Simulated and Actual Stress (SUSAS) [24].

In this study, the Berlin emotional speech database [25] was considered. This is one of the most exploited databases for SER studies. It consists of 535 utterances by 10 professional actors (5 male and 5 female) expressing 10 sentences in 7 emotions, namely anger, happiness, neutral, fear, sad, disgust and boredom. For this study, we considered anger, happiness, neutral, fear, sad and boredom emotion categories.

D. Classification

Classification of emotions is performed using well-known classifiers such as Support Vector Machine (SVM), Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Artificial Neural Networks (ANN) and k- Nearest Neighbor (kNN) [24]. In this study, we opted for a k-Nearest Neighbor (kNN) classifier. kNN classifier is a simple classifier that is often used in emotion recognition [24]. In general, the nearest-neighbor algorithm models the properties of any particular input x, to the class that appears most frequently in the k closest neighborhood of x in the training dataset. In order to apply the kNN algorithm, a distance metric D(x1; x2) is needed to identify the nearest neighbors of the input x and the number of the nearest neighbors, k, should be selected. In this study, the distance was calculated by using the Euclidean distance function and k was set to be equal to 5. There are two main schemes or decision rules in kNN algorithm, that is, similarity voting scheme and majority voting scheme. In our experiment, we used the majority voting scheme for classifying the unlabeled data [26].

IV. Experimental Setup

Our experiments were aimed to analyze the performance of VSF in identifying emotions. To perform experiments, samples of speech data from Berlin emotional database were taken as inputs. These samples were partitioned into training sets (67% of the data) and testing sets (33% of the data). Pre-processing was performed on these samples, followed by feature extractions. During the feature extractions, VTF was extracted directly from the original speech signal and VSF was extracted from LP-residual signal. Parameters were calculated from each feature and saved as feature vectors. Feature vectors were concatenated and kNN classifier was trained. During training, the training sets were cross-folded using ten-fold and fed to the system. In this study, we performed the leave-one-text-out cross-validation method [24]. The performance of speech emotion identification was measured based on the following equation:

$$\frac{Total\ Number\ of\ Correctly\ Recognized\ Emotion}{Total\ Number\ of\ Emotion} \times 100\%$$

The experiments were conducted using three sets of features; Set-1 represents only VTF, Set-2 represents only VSF and Set-3 represents the combination of VTF and VSF. The combination of VTF and VSF were done as feature fusion or early fusion in which VTF vectors were augmented with VSF and fed into the classifier.

V. Results and Discussion

Results obtained from the experiments using different feature sets are presented in this section. Table 2 shows the emotion identification rates using VTF. The performance in classification of anger is the highest (80%), followed by sad (78%) and happy (74%). Fear shows the lowest recognition rate (65%). The recognition rate for boredom is 68% and neutral is 70%. A further observation of the results shows that 16% of anger is misclassified as happy and 18% of boredom and 17% of fear are misclassified as neutral. However, the overall recognition rate is 72.5%.

Table 2
Emotion Recognition Rate (%) using VTF

Emotion	*Anger*	*Boredom*	*Fear*	*Happy*	*Neutral*	*Sad*
Anger	80	0	3	16	1	0
Boredom	1	68	8	0	18	5
Fear	8	5	65	6	17	4
Happy	15	2	6	74	2	1
Neutral	3	12	10	0	70	5
Sad	2	10	3	1	6	78

Analysis of the Table 2 reveals that anger and sad have more efficient identification rates. One possible reason for this could be because anger has highest values in mean and variance of pitch and mean of energy. On the other hand, sadness has decreased the mean value of pitch and low value in energy and speaking rate [27]. A slightly lower accuracy rate is shown in happy, neutral and boredom. This could be due to the fact that happiness is always misclassified as anger and neutral is misclassified as boredom [15, 23]. The fear is mixed with all other five emotions and shows lowest recognition rate.

Table 3 shows the emotion classification rate using VSF. In this experiment, happy shows the highest recognition rate (63%) followed by sad (60%) and boredom (57%). Anger shows the lowest recognition rate (51%). The recognition rate for fear is 55% and neutral is 52%. The average performance of VSF is 56.33%. The classification rates of emotions using ESF are lower than the classification rates using VTF as of the characteristic of LP residual which is noisy in nature [9].

Table 3
Emotion Recognition Rate (%) using VSF

Emotion	*Anger*	*Boredom*	*Fear*	*Happy*	*Neutral*	*Sad*
Anger	51	14	8	13	10	4
Boredom	6	57	11	5	12	9
Fear	8	9	55	8	10	10
Happy	11	6	6	63	6	8
Neutral	7	14	10	5	52	12
Sad	4	11	9	5	11	60

Table 4 shows emotion classification accuracies when combining VTF and VSF. The performance in classification of anger is 83%, boredom is 72%, fear is 68%, happy is 77%, neutral is 76% and sad is 82%. From the results shown, it is noted that there is a reduction in misclassification between anger and happy as compared to the results presented in Table 2. However, not much difference is observed in misclassification between boredom and neutral and fear and neutral. One of the possible reasons could be that anger and happy might contain some discriminative characteristics in VSF.

Table 4
Emotion Recognition Rate (%) using combination of VTF and VSF

Emotion	*Anger*	*Boredom*	*Fear*	*Happy*	*Neutral*	*Sad*
Anger	83	3	4	6	2	2
Boredom	3	72	5	1	15	4
Fear	5	7	68	2	14	4
Happy	10	0	8	77	0	2
Neutral	3	9	7	4	76	5
Sad	0	6	5	2	5	82

Emotion recognition performance comparison between VTF and VSF is depicted in Table 5. The result shows improved classification performance for each emotion. The overall recognition performance has improved to 76%. This can be explained by the remark that LP-residue still contains expedient information that is not modeled by the filter [19].

Table 5
Performance comparisons VTF, VSF and Their Combination

Emotion	*VTF*	*VSF*	*VTF + VSF*
Anger	80	51	83
Boredom	68	57	72
Fear	65	55	68
Happy	74	63	77
Neutral	70	52	76
Sad	78	60	82

There are two studies in the literature that can be considered reasonably close to this study. In [13], Rao et al. reported that the combination of excitation source features and spectral features has improved the emotion recognition performance up to 84%. Al-Talabani et al. in [18] have reported an improved accuracy of 88.4% when fusing spectral and prosodic features with excitation source feature at classification level. Even though these two studies lead to the same conclusion, however, they differ in terms of features and classifiers used.

VI. Conclusion and Future Works

This study was conducted to investigate the effectiveness of vocal source related features extracted from LP- residual signal in identifying emotion from speech. To accomplish this, two sets of features, a feature set based on vocal tract (VTF) and a feature set based on vocal source (VSF) were used. The analysis was performed using Berlin emotional

database and emotion classification was done using kNN classifier. Anger, boredom, fear, happy, neutral and sad were the six categories of emotions considered in this study. Experimental results showed that the VSF carries discriminant information on emotion with an average recognition rate of 56.33%. Comparative evaluation of VTF and VSF demonstrated that the performance of VSF is relatively low compared to VTF (which showed an average recognition rate of 72.5%). However, combining ESF and VTF improved the recognition rate to 76%. Furthermore, the detailed analysis of results showed that combining VSF with VTF could reduce misclassification of emotions to a certain extent.

From the study, it is evident that VSF carries emotion-specific information and combining them with VTF improves the classification rate. This can be explained as vocal source features from LP residue still contain emotion-specific information and they are complementary to vocal tract features [13, 19].

Our future work will focus on spontaneous databases and explore the discriminant power of different source features in emotion recognition. Another direction of the future research will be to investigate the use of vocal source features in reducing the misclassification of emotions.

ACKNOWLEDGMENT

This paper presents a work that is supported by the Universiti Sains Islam Malaysia (USIM), under the Competitive Grant (USIM-T): PPP/UTG-0144/FST/30/11414)

REFERENCES

[1] B. Pascal, F. Shirley and B. Catherine, “Thinking the voice: neural correlates of voice perception”, 2004 TRENDS in Cognitive Sciences, Vol.8 No.3 March 2004.

[2] E. Anna, M.E. Antonietta and V. Carl, "Needs and challenges in human-computer interaction for processing social-emotional information", Pattern Recognition Letters, Vol 66, pp. 42-51, 2015.

[3] M. D. Zbancioc and M. Feraru, “Using the Lyapunov Exponent from Cepstral Coefficients for Automatic Emotion Recognition", 2014 International Conference and Exposition on Electrical and Power Engineering (EPE 2014), Iasi, Romania 16-18 October 2014.

[4] Z. Xiao, E. Dellandrea, W. Dou and L. Chen, “Hierarchical Classification of Emotional Speech”, IEEE Transactions on Multimedia, 2007.

[5] S. R. Kadiri, P. Gangamohan, S. V. Gangashetty and B. Yegnanarayana, “Analysis of Excitation Source Features of Speech for Emotion Recognition”, INTERSPEECH 2015 (ISCA), Dresden, Germany, September 6-10, 2015.

[6] L. R. Rabiner and B. H. Juang, 1993. Fundamentals of Speech Recognition. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[7] T. Drugman, P. Alku, A. Alwan and B. Yegnanarayana, “Glottal Source Processing: from Analysis to Applications”, Computer Speech and Language, March 11, 2014.

[8] T. Drugman, Y. Stylianou, L. Chen, X. Chen and M. J. F. Gales, “Robust Excitation-Based Features For Automatic Speech Recognition”, 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4664 – 4668, April 2015.

[9] D. Pati and S. R. M. Prasanna, “Subsegmental, segmental and suprasegmental processing of linear prediction residual for speaker information”, Int J Speech Technol (2011), Dec 2010.

[10] T. Drugman, Y. Stylianou, Y. Kida and M. Akamine, “Voice Activity Detection: Merging Source and Filter-based Information”, IEEE Signal Processing Letters (Volume: 23, Issue: 2, Feb. 2016), pp. 252 – 256, Feb 2016.

[11] C. Hanilci, and F. Ertas, “Impact of Voice Excitation Features on Speaker Verification”, ELECO 2011 7th International Conference on Electrical and Electronics Engineering, Bursa, Turkey, December 2011.

[12] B. Yegnanarayana, S. R.M.Prasanna and K. S. Rao, “Speech enhancement using excitation source information”, Acoustics, Speech, and Signal Processing (ICASSP), 2002, pp. 541-544 May 2001.

[13] K. S. Rao, and S. G. Koolagudi, "Characterization and recognition of emotions from speech using excitation source information", 2012, International Journal of speech technology, 2013 – Springer, Volume 16, Issue 2, pp. 181-201, June 2013.

[14] 14. J. Nurminen, H. Silén, E. Helander and M. Gabbouj, “Evaluation of Detailed Modeling of the LP Residual in Statistical Speech Synthesis”, 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013), pp. 313 – 316, May 2013

[15] P. Gangamohan, S. R. Kadiri, S. V. Gangashetty and B. Yegnanarayana, “Excitation Source Features for Discrimination of Anger and Happy Emotions”, INTERSPEECH 2014, ISCA, Singapore, September 2014.

[16] L. Mary, "Multilevel implicit features for language and speaker recognition", Ph.D. thesis, Dept. of Computer Science and Engineering, Indian Institute of Technology, Madras, Chennai, India, June 2006.

[17] S. G. Koolagudi and K. S. Rao, “Emotion recognition from speech using source, system and prosodic features”, International Journal of Speech Technology, 15(2), pp. 265–289, 2012.

[18] A. Al-Talabani, H. Sellahewa and S. Jassim, "Excitation Source and Low-Level Descriptor Feature Fusion for Emotion Recognition using SVM and ANN", 5th Computer Science and Electronic Engineering Conference, 2013.

[19] Marcos Faundez-Zanuy, “On the Usefulness of Linear and Nonlinear Prediction Residual Signals for Speaker Recognition”, Advances in Nonlinear Speech Processing, vol. 4885, pp. 95-104. 2008.

[20] S. R. M. Prasanna, C. S. Gupta and B. Yegnanarayana, “Extraction of speaker-specific excitation information from linear prediction residual of speech”. Speech Communication, vol. 48, pp.1243-1261. 2006.

[21] T. Seehapoch and S. Wongthanavasu, “Speech Emotion Recognition Using Support Vector Machines”, 2013 5th International Conference on Knowledge and Smart Technology (KST), 2013.

[22] V. B. Kobayashi and V. B. Calag, “Detection of affective states from speech signals using ensembles of classifiers”, Intelligent Signal Processing Conference 2013 (ISP 2013), IET, Dec 2013.

[23] P. Vasuki, “Speech Emotion Recognition Using Adaptive Ensemble of Class Specific Classifiers”, Research Journal of Applied Sciences, Engineering and Technology 9(12), pp.1105-1114, 2015.

[24] J. Rybka and A. Janicki, “Comparison of speaker dependent and speaker independent emotion recognition”, Int. J. Appl. Math. Comput. Sci., 2013, vol. 23, No. 4, pp.797–808.

[25] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, “A Database of German Emotional Speech,” in INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4-8, 2005. ISCA, 2005, pp. 1517–1520.

[26] Behrouz Ahmadi-Nedushan, "An optimized instance-based learning algorithm for estimation of compressive strength of concrete", Engineering Applications of Artificial Intelligence, August 2012.

[27] 27. P. Shen, Z. Changjun, X. Chen, “Automatic Speech Emotion Recognition Using Support Vector Machine”, International Conference on Electronic & Mechanical Engineering and Information Technology, vol. 2, pp. 621 – 625, August 2011