International Journal of Engineering & Technology, 7 (2.32) (2018) 39-43

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Emotional Classification of Acoustic Information With Optimal Feature Subset Selection Methods

Dr. Swarna Kuchibhotla[1], Mr. Niranjan M.S.R[2]

[1]Dept of CSE, K L (Deemed to be University), Vaddeswaram, India
[2]Dept of EEE, National Institute of Technology, Tiruchirapalli, India.
**Corresponding author E-mail: drkswarna@kluniversity.in*

Abstract

This paper mainly focuses on classification of various Acoustic emotional corpora with frequency domain features using feature subset selection methods. The emotional speech samples are classified into neutral, happy, fear , anger, disgust and sad states by using properties of statistics of spectral features estimated from Berlin and Spanish emotional utterances. The Sequential Forward Selection(SFS) and Sequential Floating Forward Selection(SFFS)feature subset selection algorithms are for extracting more informative features. The number of speech emotional samples available for training is smaller than that of the number of features extracted from the speech sample in both Berlin and Spanish corpora which is called curse of dimensionality. Because of this feature vector of high dimensionality the efficiency of the classifier decreases and at the same time the computational time also increases. For additional improvement in the efficiency of the classifier a subset of features which are optimal is needed and is obtained by using feature subset selection methods. This will enhances the performance of the system with high efficiency and lower computation time. The classifier used in this work is the standard K Nearest Neighbour (KNN) Classifier. Experimental evaluation proved that the performance of the classifier is enhanced with SFFS because it vanishes the nesting effect suffered by SFS. The results also showed that an optimal feature subset is a better choice for classification rather than full feature set.

Keywords: *MFCC, SFS, SFFS, KNN*

1. Introduction

Human Vocal emotions depicts a key role in the aspect of multimodal man machine communication [2][14]. Affective computing is an emerging research field, applied to the computers in such way by training it to identify and respond to human emotions in a more suitable and reasonable way [21]. The main objective of affective computing is to include the mutual understanding between the machine and the human by collecting and analysing the affective information.

A classifier is used to classify the speech samples into emotional states by using the acoustic features extracted from them. Before extracting the features from the speech samples, preprocessing is performed. Preprocessing includes filtering , framing and windowing [9] The features derived from the speech samples are mainly categorized in 2 ways one is prosodic features and the other is spectral features. Spectral features are more effective than prosodic features that's why in this paper spectral features like Mel Frequency Cepstral Coefficients are applied.[13]

After extracting the features from the speech samples some simple statistics are applied on these features because their sensitivity is less to the linguistic information[26] But the problem is the number of features extracted from speech samples are more than that of the number of speech samples which is called the curse of dimensionality. When the classifier uses all these features the efficiency of the classifier will not attain the optimal state. Therefore there is a need of an algorithm which selects subset of features and optimizes the performance of the classifier. Such algorithm is called a wrapper[8]. For instance, n is the total number of speech samples then there exists a possibility of 2^n number of feature subsets. The major task for now is to search for the best suitable feature set in a space of all available feature subsets. This feature set will now best classifies the emotional state with high efficiency and low classification error rate.

Each and every feature subset selection algorithm mainly focuses on two factors. One is search strategy and the evaluation procedure. Feature subset selection is done by using search strategy and the fitness of the features that have been selected depends on the evaluation procedure by using criterion function.

Exhaustive search, random search and sequential search are different types of search strategies [26][6][12]. The number of features are increased exponentially in exhaustive search, so even for small number of features this method is not practical. In random search algorithms, insertion or deletion of features is done randomly. Evolutionary algorithms are comes under this strategy [19][15][5]. Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) are examples under the sequential search. This search strategy is easy to implement and is having less execution time. Initially SFS starts with a feature set of 0 items, best selected feature is to be added at each iteration and SBS starts with whole feature set and deletes least performing feature from the whole feature set at each iteration[19].

Filters and Wrappers are two categories of procedures of feature selection which depends on the criterion function that has been chosen. Orthogonality, mutual information and correlation are

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

utilized to accomplish feature selection. The assessment methods for feature selection in filters are more powerful than that of wrapper approaches in computation. The fitness of the features are tested very fast with this approach but it leads to features which are not optimal, if they are dependent on the classifier[19].

Even though, the Filters are faster than wrappers, wrappers are train the classifier with selected features, so they are more discriminative for a specific classifier [7] [8] One of the best frequently used method for wrappers in choosing features is the sequential forward floating selection and the classification error is estimated with the help of validation set. (SFFS) algorithm [17] [26] In this paper the SFS and SFFS are implemented to enhance the efficiency of classification technique used or to decrease the rate of error in KNN classifier. To achieve this cross validation is employed which is used to minimize the classifier error rate [25] [3].

This paper is designed in the following way, Section 2 describes various corpora, feature extraction and classification algorithm used, section 3 describes feature subset selection methods used for dimensionality reduction, section 4 describes the experimental results and section 5 describes the conclusion.

2. Material used

The research with emotional database is dealing with acted, induced and spontaneous speech samples generally [10][27]. Lot of emotional speech corpora are designed in that some are Emo-DB(Emotional database of Berlin), DES(Emotional database of Spanish), SES (Emotional database of Spanish), English and Chinese emotional speech corpora. In this work, the acted emotional speech databases are used in specific Berlin and Spanish databases are used.

2.1 Berlin emotional speech corpora

Berlin database is publicly available , emotional speech corpora of simulated type which contains seven basic emotions boredom, anger, disgust, happiness, fear, neutral and neutral etc. Five thirty five German emotional speech corpora are there totally and are described in Table 1. These are simulated by ten German native actors (5 actors and 5 actresses) [1]

Table 1: Number wise emotional samples in Berlin Database

Emotion	No. of files
Happy	71
Neutral	79
Anger	127
Sad	62
Boredom	81
Fear	69
Disgust	46
Total	535

2.2 Spanish emotional Speech corpora

As a part of IST Interface Project (”Multi modal Analysis/Synthesis System for Human Interaction to Virtual and Augmented environments”), a speech corpora of emotions for English, French, Slovenian, and Spanish language has been created. In these 4 speech emotional corpora, Spanish emotional corpora is applied our work[4].

This database includes the 7emotional classes like joy, anger, fear, disgust, surprise, neutral and sadness. As a whole there are 184 sentences for every emotion which includes sentences, a text passage and isolated words which are depicted in Table 2. These samples are recorded professionally by a male and female speakers[11].

Table 2: Item Identifier for Spanish Database

Sentence type	Item identifier
Digits and numbers	151 to 160
Isolated words	161 to 184
Paragraph text	135 to 150
Interrogative sentences	101 to 134
Affirmative sentences	1 to 100

2.3 Methods used

In this paper, spectral features are extracted and K-Nearest Neighbour classification algorithm is used along with various feature subset selection algorithms which provide a feature set as an output as well as a scalar value or quality of the feature for all the features.

2.4 Feature extraction

The Mel Frequency Cepstral Coefficients (MFCC) are spectral representation of the speech signal which are more powerful in extracting the most relevant emotional state of the acoustic signal [13]. These spectral features denotes speech signal short term power spectrum by using a linear cosine transform of a log power spectrum on a non linear melscale of frequency. [9] The method to implement the MFCC are depicted in [18] [24] [23]. 18 MFCC Coefficients are extracted along with their first order and second order derivatives for every single frame which produces 54 spectral features as a total. Six simple statistics are used to estimate the power of these features like Mean, Variance, Minimum, Range, Skewness and Kurtosis and are normalized to have zero mean and unit variance i.e 54*6=324 features related to spectral properties are extracted for every speech sample.

2.5 Classification algorithm

In this paper a classic classification algorithm like K Nearest Neighbor (KNN) is applied. The database used is separated into two phases training phase and testing phase. The database is divided into different emotional classes like neutral, happy, anger, fear, disgust and disgust. Each test speech sample is assigned a class label that is seen more frequently among the k labelled training speech samples which are most nearer to the test speech sample in terms of the Euclidian distance. KNN is a robust pattern recognition method that has been given a sufficient training data, which can design complicated non-linear decision boundaries in the feature space [16][22]. As KNN is applicable for features of high relevancy and SVM is applicable for even non relevant features. This discussion clarrifies that KNN is the suitable algorithm for our work because the performance of the non-linear pattern classification procedure depends on the feature set of high quality.

3. Feature Subset Selection Algorithms

Selection of features is the procedure for searching a group of x features from a given set of X measurements i.e $x < X$ without decreasing the efficiency of the classifier. As discussed earlier the optimal search methods are not suitable for problems having high dimension [17] and ours is a high dimensional problem so we have to use suboptimal search methods like SFS, SBS, plus-l-minus-r, SFFS and SBFS. In subset feature selection, every feature is allocated a value to represent its fitness. In this paper we execute SFFS and SFS algorithms and compare the outcomes over two emotional speech databases.

3.1 Sequential forward Selection(SFS)

A well known and frequently applicable method for selection of features is sequential forward selection and is proposed by whitney in 1971 [28][16]. It begins with a set of 0 items, and updating it iteratively by adding most relevant features by using criterion function. The subset of features are evaluated by KNN classifier

while the remaining features are used for training. The value of un weighted average recall (UAR) maximum is taken as the criterion function over a range of k values k 5;10; :::150. UAR is a class oriented exact classification rate averaged over the existing classes. The nesting effect is the drawback of SFS. A plus-l-minus-r method is developed to reduce the nesting effect of feature subsets by stearns in 1976 [20]. But the disadvantage of this procedure is there is no method to estimate the values of l and r to reach the best feature set. Rather than keeping these values fixed let them give a chance to float i.e in order to reach the optimal solution these values are changes accordingly[17]. These methods are labelled depending on the direction of the search of floating methods. The search method in SFBS is in the backward direction where as the search method in the SFFS is in the opposite direction.

3.2 Sequential forward floating selection (SFFS)

SFFS uses basic SFS procedure for the initial creation of new feature set which further deletes the least significant features in the newly created set which then provides the better updated set. The SFFS is mainly implemented in 3 steps. It initially begins with a set of null values Y=0.

Step 1 Insertion : Collection of most relevant features with respect to Y is done by using basic SFS procedure and update the Y. If it reaches d number of features then stop else goto next step.

Step 2 Conditional deletion: The most irrelevant feature k is found in Y. It it is recently added then keep it in Y and go to previous step, else delete the feature k. The Y is now better updated than the previous step.

Step 3 Conditional deletion continuity: The process of finding the most irrelevant features in Y is continued. If its deletion will (a) leave the Y with two features at least and (b) the criterion function value of the best feature set is less than the value of J(Y) then delete that feature and repeat the same step. When these two conditions are about to cease then return to step1.

4. Experiment Evaluation

The described search methods have been evaluated by KNN classifier with the speech emotional corpora of Berlin and Spanish and the outcomes are compared efficiently. The feature subset selection algorithms used here are sequential forward selection (SFS) and Sequential floating forward selection (SFFS). Databases is distributed into training phase and testing phase again for feature selection the training set is divided into training and development sets. In SFS, first the classifier is trained with training data and development set is used to determine whether the feature is loaded into the best feature set basing on the criterion function. The criterion function used in this paper is un weighted average recall. After getting the optimal feature subset the test speech samples are classified using the KNN classifier.

The experiments are also conducted to detect the classifier efficiency without selection of features, i.e using fullset of features. Which is low due to high dimensional problem i.e number features exceeds the number of instances of the database. The experiments with SFS outperformed than with conventional SFS with the help of K Nearest Neighbour classifier. With these selection of features procedures the classifier efficiency is enhanced by the reduction in the length of the feature vector in the feature space as well as the reduction in the computational time of the classification technique.

4.1 Analysis of results with KNN Classifier

The experiments are conducted with speech samples of Berlin and Spanish with KNN classifier with full feature set along with the feature subset selected from SFS and SFFS and the percentage accuracie are shown in Table 3. The outcome of the classifier with full feature set is 65% for Belin and 60.56% for Spanish databases. There exists a 10% enhancement in the outcome of the KNN classification technique with SFS and is further improved with a percentage of 20% with SFFS than with SFS.

Table 3: Emotion recognition percentage accuracy of KNN classifier by using different feature subset selection algorithms

KNN Classifier	Berlin	Spanish
Allfeatures	65%	60.56%
SFS	75%	73.33 %
SFFS	81.11 %	77.67%

The graphical representation of these three methods is depicted in Fig. 1. The horizontal axis denotes, the feature selection method used and the vertical axis denotes the performance of the classification technique used. The Blue line denotes the enhancement in the accuracy of the Berlin database and Red line denotes the enhancement in the accuracy of the Spanish database. The performance of the KNN classifier with Berlin database is slightly higher than that of the Spanish database.

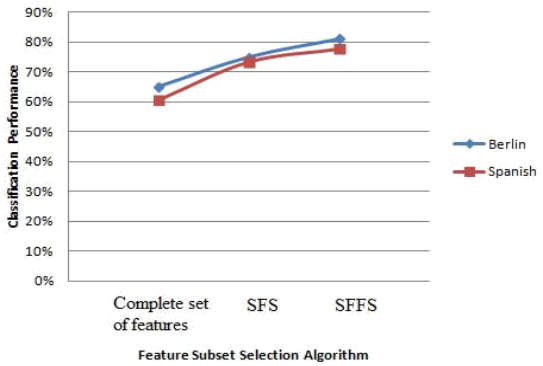

Figure 1: Emotion detection accuracy performance comparison of Berlin and Spanish corpora using various feature selection algorithms.

The Fig.2 shows the most clear graphical representation of performance comparison of each feature selection method individually with both the databases. The Blue bar represents the Berlin database and Red bar represents the Spanish database. The performance of the classifier is between 60% - 65 % with full feature set and the performance with SFS is between 65%- 75 %. The overall performance is more with SFFS which reaches betweeen 75% - 85% approximately on both databases.

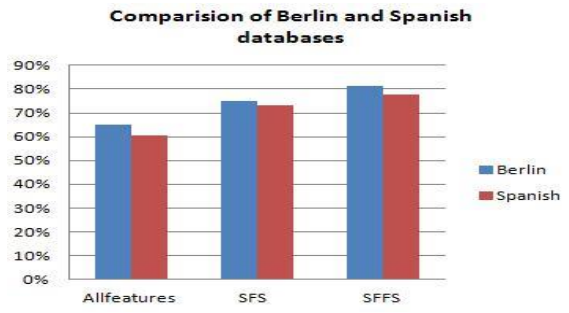

Figure 2: Performance of Emotion detection accuracy comparison of All features with features extracted by SFFS and SFS for both the emotional speech corpora.

4.2 Analysis of results with confusion matrices for Berlin database

The experimental results are analysed more clearly for each feature selection algorithm with the help of confusion matrices for each database individually. The Table 4 shows the confusion matrices with complete feature set, SFS and SFFS for Berlin database. The emotions happy, anger and sad are best classified with full feature set. The SFS also includes the emotions neutral and fear as best classifed along with the emo-tions classified with full feature set. But totally all emotions are best classified with SFFS. The Table 5 shows the best feature subset which is selected from Sequential forward selection (SFS) and Sequential floating forward selection (SFFS). We obtain best feature group of 21 with SFS and a feature set of 18 best features wiht SFFS. All these are the features selected from 324 Mel frequency Cepstral Coefficients. The full set include this complete 324 features.

Berlin	H	N	A	S	F	D
H	90	-	10	-	-	-
N	-	53	17	-	13	17
A	7	7	73	3	3	7
S	33	-	30	37	-	-
F	3	14	-	-	80	3
D	3	10	10	-	20	57

(a)

SFS	H	N	A	S	F	D
H	77	3	13	7	-	-
N	-	84	-	3	10	3
A	10	-	67	10	-	13
S	10	-	10	77	3	-
F	-	17	-	7	73	3
D	3	13	14	-	3	67

(b)

SFFS	H	N	A	S	F	D
H	90	3	-	7	-	-
N	-	87	3	-	7	3
A	10	-	83	7	-	-
S	13	7	10	70	-	-
F	-	10	-	7	80	3
D	-	10	10	-	3	77

(c)

Table 4: Emotion classification performance of Berlin Speech samples in percentage using (a) using all features (b) (c) the most important features with SFS and SFFS . Here D, F, S, A, N and H are disgust, Fear, Sad, Angry, Neutral and Happy respectively.

Berlin	Best Feature Combination
SFS(21 features)	1 5 2 90 101 239 52 289 162 179 184 243 191 298 244 152 15 132 135 109 111
SFFS(18 features)	1 5 2 90 101 239 289 179 184 244 36 13 246 34 140 113 114 223

Table 5: The most favourable feature subgroup extracted with SFFS and SFS for Berlin database

4.3 Analysis of results with confusion matrices for Spanish database

The Table 6 shows the confusion matrices for full feature set, SFS and SFFS for Spanish database. The emotions happy and anger are classified good with full feature set. Except the emotion disgust the remaining emotions are better classified with SFS. But all the emotions happy, anger, fear, neutral, disgust and sad are best classified with SFFS.

The Table 7 represents the combination of best features with SFFS and SFS for Spanish database. We obtain a feature set of 11 features with SFS and 8 features with SFFS. This feature combinations will give the best results rather than with full feature set.

Spanish	H	N	A	S	F	D
H	87	3	7	-	-	3
N	16	47	17	-	13	7
A	4	10	70	-	3	13
S	37	3	3	57	-	-
F	7	13	23	-	57	-
D	20	10	13	-	10	47

(a)

SFS	H	N	A	S	F	D
H	70	-	10	10	5	5
N	3	70	10	3	-	13
A	10	7	73	-	3	7
S	7	3	7	83	-	-
F	4	3	3	-	90	-
D	17	10	10	10	-	53

(b)

SFFS	H	N	A	S	F	D
H	83	3	-	7	-	7
N	3	71	10	3	-	13
A	-	10	73	-	10	7
S	3	-	7	83	7	-
F	3	10	-	-	80	7
D	10	-	-	13	-	77

(c)

Table 6: Emotion classification performance of Spanish Speech samples in percentage using (a) using all features (b) subdivision of features with SFS (c) subdivision of features with SFFS Here D, F, S, A, N, H represents Disgust, Fear, Sad, Anger, Neutral and Happy respectively.

Spanish	Best Feature Combination
SFS(11 features)	69 10 29 13 247 15 32 16 134 36
SFFS(8 features)	1 235 5 184 300 4 28 17

Table 7: Best subset of features extracted with SFFS and SFS for Spanish database

4.4 Evaluation of each individual emotion with feature subset selection algorithms

The tabular representation performance of each emotion with each feature selection procedure is depicted in Table 8. The graphical representation of comparison of feature selection algorithms within the database is shown in Fig 3. The blue line represents the full set, red line represents the SFS and the green line represents the SFFS.

Berlin	H	N	A	S	F	D
Set (324 features)	90	53	73	37	80	57
SFS(21 features)	77	84	67	77	73	67
SFFS(18 features)	90	87	83	70	80	77

(a)

Spanish	H	N	A	S	F	D
Set (324 features)	87	47	70	57	57	47
SFS (11 features)	70	70	73	83	90	53
SFFS (8 features)	83	71	73	83	80	77

(b)

Table 8: Emotion detection efficiency for various emotions with various subset selection algorithms using Berlin and Spanish speech corpora.

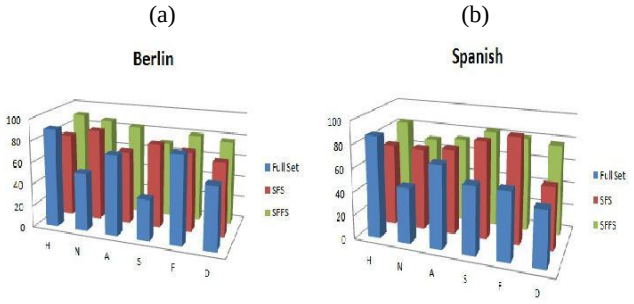

Figure 3: Emotion detection performance accuracies comparison for speech corpora (a) Berlin (b) Spanish

5. Conclusion

Classification of emotional speech samples from Berlin and Spanish databases was done with KNN classifier by using feature selection. Various feature selection algorithms like SFS and SFFS were presented. These procedures are utilized to recognize the most related features for the classification of speech samples into different emotional states. The results are widely analysed and compared against these two feature selection algorithms along with two databases. The results demonstrate the following things. 1.huge reduction of high dimensionality in feature space is achieved which indicates the improvement in computational savings. 2. Classification performance of KNN Classifier is improved using feature selection methods. Sequential forward floating selection accuracy is best when compared with the accuracy of sequential forward selection in both the databases and also the results are slightly more for Berlin database than for Spanish database. Experimental results showed an improvement of 20% with feature selection algorithms when compared with complete set of features.

References

[1] Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Wal-ter F Sendlmeier, and Benjamin Weiss. A database of german emotional speech. In Interspeech, pages 1517– 1520, 2005.

[2] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapat-soulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. Signal Processing Magazine, IEEE, 18(1):32–80, 2001.

[3] Bradley Efron and Robert J Tibshirani. An introduction to the bootstrap, volume 57. CRC press, 1994.

[4] Vladimir Hozjan, Zdravko Kacic, Asuncion Moreno, Antonio Bonafonte, and Albino Nogueiras. Interface databases: Design and collection of a multilingual emo-tional speech database. In LREC, 2002.

[5] Alejandro Jaimes and Nicu Sebe. Multimodal human– computer interaction: A survey. Computer vision and image understanding, 108(1):116–134, 2007.

[6] Anil Jain and Douglas Zongker. Feature selection: Evalu-ation, application, and small sample performance. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(2):153–158, 1997.

[7] George H John, Ron Kohavi, Karl Pfleger, et al. Irrele-vant features and the subset selection problem. In ICML, volume 94, pages 121–129, 1994.

[8] Ron Kohavi and George H John. Wrappers for feature subset selection. Artificial intelligence, 97(1):273–324, 1997.

[9] Swarna Kuchibhotla, HD Vankayalapati, RS Vaddi, and KR Anne. A comparative analysis of classifiers in emo-tion recognition through acoustic features. International Journal of Speech Technology, pages 1–8, 2014.

[10] Swarna Kuchibhotla, BS Yalamanchili, HD Vankayala-pati, and KR Anne. Speech emotion recognition using regularized discriminant analysis. In Proceedings of the International Conference on Frontiers of Intelligent Com-puting: Theory and Applications (FICTA) 2013, pages 363–369. Springer, 2014.

[11] Swarna Kuchibhotlaa, Hima Deepthi Vankayalapati, BhanuSree Yalamanchili, and Koteswara Rao Anne. Analysis and evaluation of discriminant analysis tech-niques for multiclass classification of human vocal emo-tions. In Advances in Intelligent Informatics, pages 325– 333. Springer, 2015.

[12] Huan Liu and Lei Yu. Toward integrating feature selection algorithms for classification and clustering.Knowledge and Data Engineering, IEEE Transactions on, 17(4):491–502, 2005.

[13] Iker Luengo, Eva Navas, and Inmaculada Hernaez´. Fea-ture analysis and evaluation for automatic emotion iden-tification in speech. Multimedia, IEEE Transactions on, 12(6):490–501, 2010.

[14] Iain R Murray and John L Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. The Journal of the Acoustical Society of America, 93:1097, 1993.

[15] Maja Pantic and Leon JM Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE, 91(9):1370–1390, 2003.

[16] Jouni Pohjalainen, Okko Ras¨anen,¨ and Serdar Kadio-glu. Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. Computer Speech & Language, 2013.

[17] Pavel Pudil, Jana Novovicov˘a,´ and Josef Kittler. Floating search methods in feature selection. Pattern recognition letters, 15(11):1119–1125, 1994.

[18] Nobuo Sato and Yasunari Obuchi. Emotion recognition using mel-frequency cepstral coefficients. Information and Media Technologies, 2(3):835–848, 2007.

[19] Mohammad Hossein Sedaaghi, Dimitrios Ververidis, and Constantine Kotropoulos. Improving speech emotion recognition using adaptive genetic algorithms. In Proc.European Signal Processing Conference (EUSIPCO),Polland, 2007.

[20] Petr Somol, Pavel Pudil, J Novovicov˘a,´ and Pavel Paclık. Adaptive floating search methods in feature selection, Pattern recognition letters, 20(11):1157–1163, 1999.

[21] Jianhua Tao and Tieniu Tan. Affective computing: A review. In Affective computing and intelligent interaction, pages 981–995. Springer, 2005.

[22] Sergios Theodoridis and Konstantinos Koutroumbas. Pat-tern recognition. IEEE TRANSACTIONS ON NEURAL NETWORKS, 19(2):376, 2008.

[23] HD Vankayalapati, KR Anne, and K Kyamakya. Ex-traction of visual and acoustic features of the driver for monitoring driver ergonomics applied to extended driver assistance systems. In Data and Mobility, pages 83–94. Springer, 2010.

[24] Siddha VR Kyamakya K. Vankayalapati HD, Anne KR. Driver emotion detection from the acoustic features of the driver for real-time assessment of driving ergonomics process. International Society for Advanced Science and Technology (ISAST) Transactions on Computers and Intelligent Systems journal, 3(1):65–73, 2011.

[25] Dimitrios Ververidis and Constantine Kotropoulos. Fast sequential floating forward selection applied to emotional speech features estimated on des and susas data collec-tions. In Proc. XIV European Signal Processing Conf, 2006.

[26] Dimitrios Ververidis and Constantine Kotropoulos. Fast and accurate sequential floating forward feature selection with the bayes classifier applied to speech emotion recognition. Signal Processing, 88(12):2956–2970, 2008.

[27] Thurid Vogt, Elisabeth Andre,´ and Johannes Wagner. Automatic recognition of emotions from speech: a re-view of the literature and recommendations for practical realisation. In Affect and emotion in human-computer interaction, pages 75–91. Springer, 2008.

[28] A Wayne Whitney. A direct method of nonparametric measurement selection. Computers, IEEE Transactions on, 100(9):1100–1103, 1971.