International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.11 (2015), pp.33-42
http://dx.doi.org/10.14257/ijmue.2015.10.11.04

An Algorithm Study for Speech Emotion Recognition Based Speech Feature Analysis

Ji Zhengbiao, Zhou Feng and Zhu Ming

Yancheng Institute of Technology, College of Information Engineering, Yancheng, 224051, China
zhengbiaoj@126.com

Abstract

Based on the studies of the emotional feature, a new method based on the emotional feature analysis for emotion recognition in speech signal characteristics is proposed. Collected from 10 speakers with joy, anger, surprise and sad emotion statement, a total of 3000 sentences, 10 emotion characteristics extracted from these speech data. According to the experiments in speech emotion corpus using different classifiers, the proposed method can both achieve relatively better performance than the state-of-the-art dimensionality reduction methods and 92.5% of the average emotional recognition rate is obtained. The future work will be focused on searching for more effective emotional characteristic parameter and recognition method.

Keywords: *Speech signal, Emotional feature analysis, Emotion recognition*

1 Introduction

Speech signal processing is always a hot topic in the world [1-3]. Speech not only contains the text symbol information, but also contains the information such as people's feelings and emotions[4]. Therefore, emotion characteristics of artificial processing and recognition of speech signal is of great significance[5]. Although, the research of emotional feature speech recognition from signal is a hot subject in recent years, some effective identification methods have been proposed. Huang Chengwei, *etc.* study the cross-database speech emotion recognition based on online learning. Experimental results show that by introducing the online learning module speech emotion recognition system can be better adapted to new data [6]. Jin yun, *etc.* proposed a two-layer Multiple Kernel Learning (MKL) scheme for speaker-independent speech emotion recognition. Results prove the effectiveness of the feature selection method [7]. XU Xinzhou proposed an algorithm based on graph learning and graph embedding framework, Speaker-Penalty Graph Learning (SPGL) to solve the problems caused by different speakers. The proposed method with linear and kernelized mapping forms can both achieve relatively better performance than the state-of-the-art dimensionality reduction methods [8].

Researches of the emotion feature analysis and recognition in speech signals is proposed in this paper. Aimed at containing joy, anger, surprise and sadness four kinds of emotion speech signals, characteristics of time structure, amplitude structure, pitch frequency and resonance peak structure is analyzed. Based on the analysis above, 10 characteristic parameters for emotion recognition are extracted. We put forward a new emotion feature recognition method based on principal element analysis. With 10 emotion characteristics extracted, for 1000 sentences collected from 10 speakers of emotional statements, 92.5% of the average emotion recognition rate has been got.

2. The Selection of Speech Data for Emotion Analysis

In this paper, the selection of language sentence for experiment analysis mainly comes

ISSN: 1975-0080 IJMUE

from two aspects followed. First, statements selected must not contain a particular aspect of emotional tendency; secondly, statements selected must contain high emotional freedom, for the same statement can exert all kinds of emotions. Moreover, to the length of the statement, composition of consonants and auxiliary components, all differences between male and female should be considered. According to principles above, 60 sentences for sentiment analysis are selected [9]. In this paper, the emotion type is roughly divided into joy, anger, surprise and sadness, and all the common emotions are classified as much as possible into this category, which is considered as reasonable classification for computer sentiment analysis research. In order to obtain the original speech data, 60 statements from 10 male speakers with joy, anger, surprise and sadness is pronounced once again. At the same time, speakers are told to pronounce each sentence once again calmly as much as possible without emotion. Through the process above 3000 language sentences are collected for experiment. In the classification experiments, 2000 sentences are taken for training and 1000 sentences for recognition. To test the effectiveness of the speech data collected for emotion experiment, an audition experiment is carried out. 5 speakers differing from the 10 above are required sitting in front of computer terminals and given collected statements with various emotions randomly. Then the speakers judge the emotion type of voices by subjective evaluation. After repeated listening and comparing, meaningful test in math (Mcnemar test) [10] is implemented. The unobvious emotion characteristics of sentence are deleted and redone.

3. Selection and Extraction of Emotional Characteristics

In general, emotional features of speech are expressed by voice rhythm changes [11]. For example, when a person is angry, the rate of speech is more quickly and the volume will be large, while higher tones will be got. In addition, the emotion information in speech signals might be influenced by statements vocabulary contents [12]. So that, in order to make the results of the analysis eliminate the impact, generally, through the analysis of relative relationship between emotional speeches and calm voices without emotion, the relative characteristics of the structure, characteristics and distribution rule are find out for processing and recognition of different emotional speech signals [13]. 16 bit A/D conversion of input signals are implemented according to the 12 KHZ, and then the sampling signals are added with windows of 23.22 ms (256 points) and Hamming with 10 ms moving windows. For in order to take use of emotion information in voice signals as much as possible, 10 emotion characteristics are selected as parameters for emotion recognition as follow: pronunciation duration, average pitch frequency, maximum pitch frequency, average rate of pitch frequency change, average amplitude energy, amplitude energy dynamic range, average frequency of resonance peak, average rate of resonance peak frequency change, the average value of resonance peak value point average slope in the regression straight line and average resonance peak value [14].

3.1 Durations of Statements Pronunciation

Durations from start to finish of each emotion statement are calculated. These times include other silent parts, which are contributed to emotion. In recognition durations and corresponding calm duration ratios of emotion statements are considered as recognition feature parameters.

3.2 Pitch Frequency

Using the Cepstrum method pitch frequency is calculated by frame, while the pitch frequency curve of median filtering and linear smoothing conduct are carried out [15]. The trajectory curve maximum of pitch frequency from emotion signal, the curve of average pitch frequency and average rate of changes are extracted. Moreover, the average rates of change from pitch frequency are difference average absolute value of the point to each frame from speech signal pitch frequencies. In recognition, the emotional statements of pitch frequency, maximum and average corresponding calm statements of pitch frequency, the maximum value of the average difference; emotion statements of pitch frequency change rate and the ratios between the corresponding calm statements for identification with the variation of the pitch frequency are considered as characteristic parameters.

3.3 Amplitude

In this paper, the amplitude of the average energy and dynamic range characteristics are analyzed and compared [16]. In order to avoid the influence of the silent and noise in pronunciation, we only consider the average of the absolute value exceeds a certain threshold from the amplitude short-term energy. Recognition when the amplitude of the emotion statement average energy, dynamic range, corresponding calm statement amplitude of the average energy and difference of dynamic range are considered as characteristic parameters for recognition.

3.4 Formant

Formant is an important parameter of channel characteristics. We use the linear prediction method (LPC) and 14 order prediction coefficients, then frequency response curve of sound channel is used to estimate predict coefficients. With Peak Picking method is introduced to calculate the formant frequency. This paper analyzes the first resonance peak frequency, the first resonance peak frequency of the average change rate, the first four formant value points with average slope of the regression straight line and former 4 average values of formants. Emotional statements 1 of each frame before the average frequency of the formant, four formant value points before the average slope of the regression straight line and average value of four formant and corresponding calm statement are the difference between these parameters and the first formant frequency change rate and the ratio of the corresponding calm statements are selected as recognition using characteristic parameters.

4. Emotion Recognition Experiments and the Results

Nine meaning characteristic parameters variables are extracted from speech signal the nine. For there is a correlation between these variables, it is hoped to eliminate the correlation between the variables by an orthogonal transformation. After the orthogonal transformation covariance between the new variable is equal to 0, analysis of data is simplified. On the other hand, we have nine parameters extracted from speech signal with different angles which contain the emotional information from the speech signals. For designing a comprehensive index, it generally means emotional information contained in the speech signals. Principal component analysis is a kind of statistical method which is more than the original index into one of the few independent comprehensive indexes. It is shown in Figure 1.

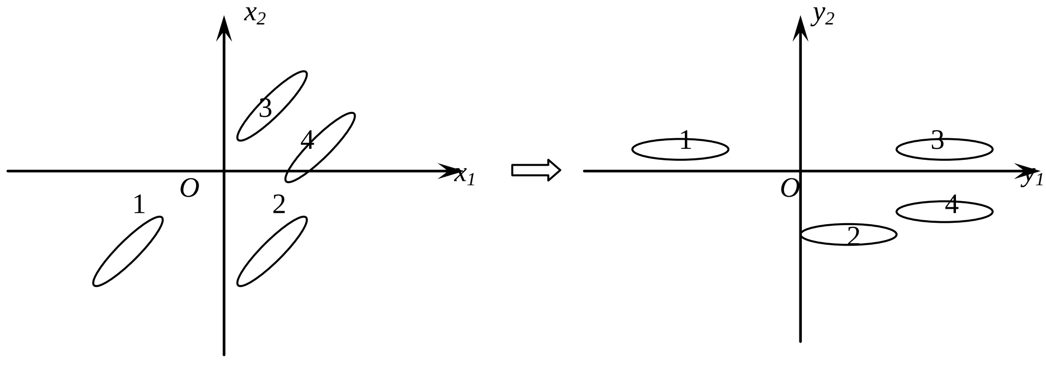

Figure 1. Transformation Diagram in Principal Component Analysis

The original feature vector is set as $\vec{X}$, made up of P characteristics. $\vec{Y}$ is come from the orthogonal transformation of $\vec{X}$, that is $\vec{Y} = U\vec{X}$ or each component $y_1, y_2, \cdots, y_p$ is a linear combination of each component from $\vec{X}$, namely

$$\begin{aligned} y_1 &= u_{11}x_1 + u_{12}x_2 + \cdots + u_{1p}x_p \\ y_2 &= u_{21}x_1 + u_{22}x_2 + \cdots + u_{2p}x_p \\ &\cdots \quad \cdots \\ y_p &= u_{p1}x_1 + u_{p2}x_2 + \cdots + u_{pp}x_p \end{aligned} \tag{1}$$

Including u_{ij} meet

$$\sum_{j=1}^{p} u_{ij}^2 = 1 \quad \text{and} \quad \sum_{j=1}^{p} u_{ij}u_{kj} = 0 \qquad (i \neq k) \tag{2}$$

According to the requirements of $\vec{Y}$ the j component y_j and the k components y_k should be independent, and if $j < k$, the variance y_j is greater than or equal to the variance of y_k, so the relationship between orthogonal transformation matrix $\vec{U}$ and diagonal matrix $D(\vec{Y})$ is

$$\vec{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1p} \\ u_{21} & u_{22} & \cdots & u_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ u_{p1} & u_{p2} & \cdots & u_{pp} \end{bmatrix} \qquad D(\vec{Y}) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} = \Lambda$$

$$D(\vec{Y}) = D(\vec{U}\vec{X}) = \vec{U}D(\vec{X})\vec{U}' = \Lambda \quad \text{or} \quad D(\vec{X})\vec{U}' = \vec{U}'\Lambda \tag{3}$$

Here, $\lambda_1, \lambda_2, \cdots, \lambda_p$ are the variance of $y_1, y_2, \cdots, y_p$, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the characters of $D(\vec{X})$. The j column vector of $\vec{U}'$ or the j column vector of $\vec{U}$ is characteristic unit vector response of λ_j.

The orthogonal transformation of $\vec{U}$ to meet the requirements of the $\vec{Y}$ of each component are independent with each other are found, and makes y_1 is with the largest variance, y_2 holds the largest variance of all independent random variables.

In fact we cannot get $D(x)$, can only get its' sample variance S_x. Hence, in practical, we set off from the sample covariance matrix S_x, solving an orthogonal transformation $\vec{U}$ and transform S_x to a diagonal matrix, $\vec{U}$ meet $\vec{U}S_x\vec{U}' = \Lambda$. Starting from the sample covariance matrix S_x and the principal component is called the sample principal component. The coordinates of each sample points x_i plug type 4-1 can get a principal component of each sample point coordinates. In order to eliminate the difficulties of sample point coordinates caused by different units and make the requirements of mathematical expectation is zero. Before calculation, the original data should be standardized, as shown in type 4. p is dimension of the sample in the type above, n is unit number of sample, $\bar{x}$ represents the average of the sample.

$$\tilde{X} = \begin{bmatrix} s_1^{-1} & 0 & \cdots & 0 \\ 0 & s_2^{-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & s_p^{-1} \end{bmatrix} \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_1 & \cdots & x_{1n} - \bar{x}_1 \\ x_{21} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{2n} - \bar{x}_2 \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} - \bar{x}_p & x_{p2} - \bar{x}_p & \cdots & x_{pn} - \bar{x}_p \end{bmatrix} \tag{4}$$

Here, $s_a = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ai} - \bar{x}^2\right)}$ is the standard deviation of a lines in original data matrix.

Because λ_j is variance of sample points on the j principal components, it represents dispersion degree of sample points in the direction of the principal component. If the λ_j value is small, the dispersion degree between samples is small on the y_i direction. The effect of principal component analyzing the sample data is not big, which can be neglected. Therefore, we define η_j for the contribution rate of the j principal component, and $\lambda_1 + \lambda_2 + \cdots + \lambda_p$ are the trace of the matrix $\Lambda = \vec{U} S_x \vec{U}'$.

$$\eta_j = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \qquad j = 1, 2, \cdots, p \tag{5}$$

Because λ_j is ranked according to size, so as long as the accumulation contribution rate $\sum_{j=1}^{m} \eta_j$ of first m（$m \le p$）principal component is close to 1, the rest p-m principal component can be abandoned. The number of sample principal component coordinate is only m, so as to achieve the purpose of approximating the sample points in low dimensional space. General experience has shown that cumulative contribution rate is greater than 0.8 or 0.85.

The correlation coefficient $r\left(y_k, x_j\right)$ of principal component y_k and the original component x_j is called the load of the j factor on the k principal component, which reflects the relationship between the original variables and the main component. The geometric interpretation of the relationship is the length of a projection on the principal axis of the unit length in original coordinate. The sample valuations for this load can be calculated as

$$r\left(y_k, x_j\right) = \sqrt{\lambda_k}\,\vec{U}_{kj} \Big/ \sqrt{S_{jj}} \tag{6}$$

Load factor has the following two properties

$$\sum_{j=1}^{p} S_{jj}\, r^2\left(y_k, x_j\right) = \lambda_k \qquad \left(k = 1,2 \cdots p\right) \tag{7}$$

$$\sum_{k=1}^{p} r^2\left(y_k, x_j\right) = 1 \qquad \left(j = 1,2 \cdots p\right) \tag{8}$$

In practical use, we will start from the standardized data matrix, its sample variance matrix is equal to the original data correlation matrix $\vec{R}$, so here is $\vec{R}\vec{U}' = \vec{U}\Lambda$, diagonal component $\lambda_1, \lambda_2 \cdots \lambda_p$ of Λ is characteristic of $\vec{R}$.

Because diagonal components of the correlation matrix q main are 1, accordingly

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p = r_{11} + r_{22} + \cdots + r_{pp} = P \tag{9}$$

Contribution rate formula and factor load formula transform as, respectively

$$\eta_j = \frac{\lambda_j}{P} \qquad \text{and} \qquad r\left(y_k, x_j\right) = \sqrt{\lambda_k}\, U_{kj} \tag{10}$$

Table 1. Table of Emotional Speech Characteristic Values

characteristic value	2.8540	1.5325	1.4327	0.8558	0.7348	0.5752	0.5212	0.3828	0.1081
contribution rate	0.317	0.170	0.160	0.095	0.082	0.063	0.058	0.043	0.012
cumulative contribution rate	0.317	0.487	0.647	0.742	0.824	0.887	0.945	0.988	1.000

As shown in Table 1, the cumulative contribution rate of the first five principal components is 82.4%. If the other principal components are rejected, the loss of information is only 17.6%. Therefore, we decided to choose the first five principal components.

Table 2. Table of Load Factors

variable of principal component	y_1	y_2	y_3	y_4	y_5
T	-0.5594	0.0016	0.4062	-0.1095	0.06325
F_0	0.7956	0.3383	0.0762	-0.3811	0.1119
$F_{0\ range}$	0.8523	0.3240	0.0992	-0.2472	0.1450
$F_{0\ rate}$	0.6818	0.1382	-0.3704	0.1359	0.0660
A	0.3419	-0.3194	0.7698	-0.0592	-0.0415
A_{range}	0.4937	-0.3846	0.5578	0.2646	-0.2235
F_1	0.5121	-0.3052	-0.1966	0.5302	0.4747
$F_{1\ range}$	0.0288	0.7108	0.2158	0.5000	-0.1270
$F_{1\ rate}$	0.3051	-0.6677	-0.3588	-0.1193	-0.0612

4.1. Recognition Method

Using the Gaussian mixture model (Figure 2) for recognition (input vectors have been implemented the principal component analysis). Mixed Gaussian distribution model is a single-state model, which holds multiple Gaussian distribution function. As an example, to a feature vector, set its cumulative probability can be made of the following (11) shown where the 3 mixed Gaussian distribution model is obtained.

$$P_k = \lambda_1 f_1(\vec{y}) + \lambda_2 f_2(\vec{y}) + \lambda_3 f_3(\vec{y}) \tag{11}$$

In(11), $f_i(\bullet)(i = 1,2,3)$ is a Gaussian distribution function and $\lambda_1, \lambda_2, \lambda_3$ is the weight, meeting $\lambda_1 + \lambda_2 + \lambda_3 = 1$. In the training phase, using PCA first, the training in each *p* vector $\vec{X}_i = \{x_{i1}, x_{i2}, \cdots, x_{ip}\}$ transform into vector $\vec{Y}_i = \{y_{i1}, y_{i2}, \cdots, y_{im}\}\ (m \le p)$ composed of the main components and then centroids are calculated through vector quantization method [17].

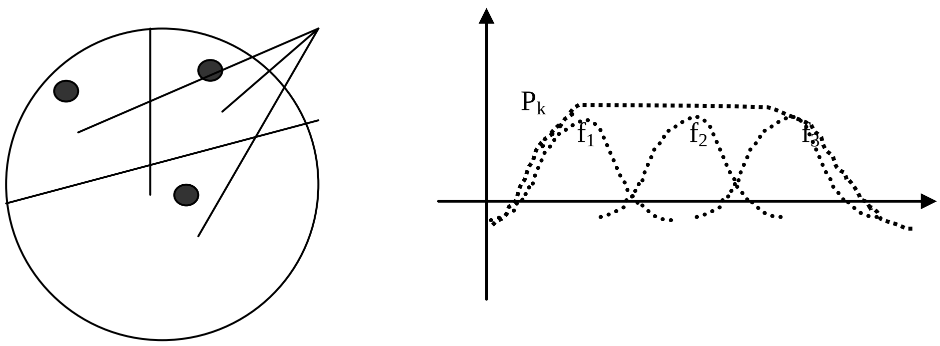

Figure 2. Schematic Diagram of GMM Model

Emotional categories of training characteristic vector of code, and calculate the corresponding variance $\vec{\sigma}_n = (\sigma_{n1}, \sigma_{n2}, \cdots, \sigma_{nm})^t \ (n = 1,2,\cdots)$ of the vector $\vec{\mu}_n = (\mu_{n1}, \mu_{n2}, \cdots, \mu_{nm})^t \ (n = 1,2,\cdots)$ to the yard. So every yard vector and the corresponding variance compose a Gaussian distribution function [18]. In this paper, the covariance matrix of Gaussian distribution function is as shown in the following (4-11). In recognition, to the speech emotion characteristics of one main component $\vec{y}$, using (4-10) and getting the probability to other values. Then the greatest probability of the emotional category is the recognition result.

$$\Sigma_n = \begin{bmatrix} \sigma^2{}_{n1} & 0 & \dots & 0 \\ 0 & \sigma^2{}_{n2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & \cdots & \sigma^2{}_{nm} \end{bmatrix} \quad (n = 1, 2) \tag{12}$$

Supposed we set up a q Gaussian mixture model with the points on m dimensions main coordinates, as bellow:

$$P_q = \sum_{i=1}^{q} \lambda_i \cdot \frac{1}{\sqrt{2\pi}} \exp\left(- \frac{\left(\vec{Y}_i - \vec{\mu}_i\right)}{2\vec{\sigma}_i^2} \right) \tag{13}$$

In practical application, through step by step classification method, the coordinate’s points of the sample points on the principal coordinates are classified into *s* class. Then, classification is modified in accordance with some optimal principle, until the classification is reasonable. In our experiment, the minimum distance method is proposed.

Now if there are n p-dimension samples in all, need to be divided into s class, the following steps would be employed:

a) Preliminary classification

First artificially selecting of s condensation points, every point is a vector of p-dimension and each one is a single category. And then to calculate the distance between each sample and condensation point, while each sample is classified to the category the nearest condensation points belonging to. Thus all the samples are differed into the s classes. Then these m classes are regarded as $G_1^0, G_2^0, \cdots, G_s^0$. In order to speed up the calculation and obtain a reasonable classification, the selection of the condensation points is according to the ‘minimum maximum principle’:

Firstly the most remote two sample points $\vec{x}_{i1}, \vec{x}_{i2}$ of all the samples are selected as the first two condensation points, namely:

$$d\left(\vec{x}_{i1}, \vec{x}_{i2}\right) = \max \left\{ d\left(\vec{x}_i, \vec{x}_j\right) \ : i \neq j \right\} \tag{14}$$

And then the third condensation point $\vec{x}_{i3}$ is selected, making the smaller distance between $\vec{x}_{i3}$ with $\vec{x}_{i1}$ and $\vec{x}_{i2}$ is and the biggest one of the all points with $\vec{x}_{i1}$ and $\vec{x}_{i2}$. Then $\vec{x}_{i4}$ is selected as the same principle.

If *k* condensation points are already selected as $\vec{x}_{i1}, \vec{x}_{i2}, \cdots, \vec{x}_{ik}$, the condensation point q as number of k+1 is selected according to the following principles

$$\begin{aligned} &\min \left\{ d\left(\vec{x}_{ik+1}, \vec{x}_{ir}\right) \quad , \quad r = 1, 2, \cdots, k \right\} \\ &= \max \left\{ \min \left\{ d\left(\vec{x}_j, \vec{x}_{ir}\right) \quad : r = 1, 2, \cdots, k \right\} : j \leq n \right\} \end{aligned} \tag{15}$$

b) Classification modification

The centers of gravity of all categories as $\bar{x}_1^0, \bar{x}_2^0, \cdots, \bar{x}_s^0$ are calculated, which are considered as s new condensation points. Then the distances between each sample and new condensation points are calculated. Moreover, each sample are classified into classes determined by the nearest condensation points. Then the classification $G_1^1, G_2^1, \cdots, G_s^1$ is obtained after being modified for the first time.

c) Repeat 2 until sample classification is no longer changed

4.2. Recognition Results

Using the emotion recognition method (method1), 1000 test emotion statements are conducted into emotion recognition experiments. The results are shown in Table 3. In order to compare, results of classification recognition according to the Mahalnobis distance scale (method 2) are also listed in the Table 3.

Table 3. Emotion Recognition Results [%]

Emotion type	happy	angry	surprise	sad	average
Recognition method1	91	92	89	98	92．50
Recognition method 2	84	85	83	95	86．75

5. Conclusions

This paper is present from four aspects such as time structure of emotion sentences *etc.* Emotion speech signals with joy, anger, surprise and sad 4 kinds of emotions are analyzed and compared, for finding out the distribution regularities in characteristic of different emotional signals. Then the recognition method of speech emotion category based on the main element analysis method is proposed. Through the recognition experiment on 1000 emotion sentences, which results show that using this recognition method can obtain basically close to the normal performance of human recognition effect. The future work will be focused on searching for more effective emotional characteristic parameter and recognition method. Moreover, further analysis and recognition experiments in broader scope will be proposed.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (Grant No. 61273266, 61201326) and the Natural Science Foundation of Shandong Province (Grant No. ZR2014FQ016). The authors would like to thank the reviewers for their valuable suggestions and comments.

References

[1] W. Qingyun, J. Xinchun and R. Liang, "Compressed Sampling and Source Localization of Miniature Microphone Array", IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, vol. 97, no. 9, **(2014)**, pp. 1902-1906.

[2] Liang R. Y., J. Xi and L. Zhao, "Experimental study and improvement of frequency lowering algorithm in Chinese digital hearing aids", Acta Physica Sinica, vol. 61, no. 13, **(2012)**, pp. 1-13.

[3] Liang R., J. Xi and J. Zhou, "An improved method to enhance high-frequency speech intelligibility in noise", Applied Acoustics, vol. 74, no. 1, **(2013)**, pp. 71-78.

[4] Schuller B., B. Vlasenko and F. Eyben, "Cross-Corpus Acoustic Emotion Recognition: Variances and Strategies", IEEE Transactions on Affective Computing, vol. 1, no. 2, **(2010)**, pp. 119-131.

[5] Ververidis D. and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods", Speech Communication, vol. 48, no. 9, **(2006)**, pp. 1162-1181.

[6] Huang C. W., R. Y. Liang and Q. Y. Wang, "Practical Speech Emotion Recognition Based on Online Learning: From Acted Data to Elicited Data", Mathematical Problems in Engineering, 2013, 2013, **(2013)**, pp. 1-9.

[7] Jin Y., P. Song and W. Zheng, "Speaker-independent speech emotion recognition based on two-layer multiple kernel learning", IEICE Transactions on Information and Systems, E96-D, no. 10, **(2013)**, pp. 2286-2289.

[8] Xu X. Z., C. W. Huang and C. Wu, "Graph Learning Based Speaker Independent Speech Emotion Recognition", Advances in Electrical and Computer Engineering, vol. 14, no. 2, **(2014)**, pp. 17-22.

[9] Meng H. and N. Bianchi-Berthouze, "Affective state level recognition in naturalistic facial and vocal expressions", Cybernetics, IEEE Transactions on, vol. 44, no. 3, **(2014)**, pp. 315-328.

[10] Barker J. and X. Shao, "Energetic and informational masking effects in an audiovisual speech recognition system", Audio, Speech, and Language Processing, IEEE Transactions on, vol. 17, no. 3, **(2009)**, pp. 446-458.

[11] El Ayadi M., M. S. Kamel and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases", Pattern Recognition, vol. 44, no. 3, **(2011)**, pp. 572-587.

[12] Savva N., A. Scarinzi and N. B. Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience", Computational Intelligence and AI in Games, IEEE Transactions on, vol. 4, no. 3, **(2012)**, pp. 199-212.

[13] Boril H. and J. H. Hansen, "Unsupervised equalization of Lombard effect for speech recognition in noisy adverse environments", Audio, Speech, and Language Processing, IEEE Transactions on, vol. 18, no. 6, **(2010)**, pp. 1379-1393.

[14] Zhang C., Y. Liu and Y. Xia, "Reliable Accent-Specific Unit Generation With Discriminative Dynamic Gaussian Mixture Selection for Multi-Accent Chinese Speech Recognition", Audio, Speech, and Language Processing, IEEE Transactions on, vol. 21, no. 10, **(2013)**, pp. 2073-2084.

[15] Lu L., K. Chin and A. Ghoshal, "Joint uncertainty decoding for noise robust subspace Gaussian mixture models", Audio, Speech, and Language Processing, IEEE Transactions on, vol. 21, no. 9, **(2013)**, pp. 1791-1804.

[16] Andrzej W. and W. Edyta, "Integrating Bayesian networks into fuzzy hypothesis testing problem - Case based presentation", 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services, Healthcom 2013, October 9, 2013 - October 12, 2013[C].Lisbon, Portugal:IEEE Computer Society, **(2013)**.

[17] Liu H., W. Wang and K. Li, "A novel speech coding algorithm for cochlear implants", 2012 5th International Conference on Biomedical Engineering and Informatics, BMEI 2012, October 16, 2012 - October 18, 2012[C].Chongqing, China: IEEE Computer Society, **(2012)**.

[18] Rieger S. A., R. Muraleedharan and R. P. Ramachandran, "Speech based emotion recognition using spectral feature extraction and an ensemble of kNN classifiers", 9th International Symposium on Chinese Spoken Language Processing, ISCSLP 2014, September 12, 2014 - September 14, 2014[C].Singapore, Singapore:Institute of Electrical and Electronics Engineers Inc., **(2014)**.

Author

Biao Jizhengbiao, He was born in 1981year. He received the M.S. degree from Hohai University in 2007 year. He is a professor at Yancheng Institute of Technology. His research interests include instrumentation and measurement technology.