Speech Emotion Recognition Based on Fuzzy K-NN Algorithm with Fractionally Spaced Blind Equalization

Yuan Tao[a], Deng Chunhong[b], Shi Wangyang[c]

Information Engineering Department, Anhui Technical College of Mechanical and Electrical Engineering, Wuhu 241002, China

[a]yuantao_1988@yeah.net, [b]ahjddch@126.com, [c]shiwangyang1688@163.com

Keywords: Fuzzy KNN, speech emotion, fractional spaced equalizer, k nearest neighbor algorithm.

Abstract. Due to the noise and multi-channel room acoustic environment, practical speech emotion recognition remains an unsolved challenge. In this paper, we first study the fractional spaced blind equalizer for speech preprocessing. The noise interference is effectively removed and more detailed emotional features are reserved. Second, the fuzzy k nearest neighbor algorithm is used to classify speech. Finally, the proposed algorithm is compared with traditional speech emotion recognition algorithms. Experimental results show that the fractionally spaced equalization is effective for practical speech emotion recognition.

Introduction

Language is an important part of civilization and speech is the most convenient way for communication. Speech signal conveys a lot of useful information. Modern linguistic study focuses on the language information and achieves many results. However there are a lot of parallel linguistic information are over looked in traditional speech recognition, such as speaker emotional state, gender and age. This type of speaker information is very important for human-computer interaction, automatic recognition and virtual reality.

Speaker emotion recognition is based on the emotional feature analysis. The most important speech emotion features include speech quality feature, spectral feature and prosodic feature [1,2]. We cannot achieve good results using only one of these types of features. Many researchers try to fuse these emotional features to get optimal feature set. Most of the current researches are focused on the ideal lab environment. The received signal is ideal. However, in practical applications, noise and multi-path acoustic environment are the major challenges in practical speech emotion recognition. In this paper, we first enhance the emotional features using fractionally spaced equalizer for actual acoustic channels. Second, we use fuzzy KNN classifier to model speech emotions and finally the results are verified in experiment using a local database [3].

Speech Emotional Feature Preprocessing Based on Fractionally Spaced Equalization

Fractionally spaced sampling algorithm is based on the baud space method. Speech sampling rate is larger than $1/T$ baud rate in the oversampling step. Related research shows that fractionally spaced sampling can be treated as multi-channel system model. More speech details can be reserved [4,5]. System layout is shown in Fig.1.

$s(k)$ is the speech signal sequence with period T; $\boldsymbol{c}^{(i)}(k)(i=0,1\cdots P-1)$ is the sub-channel impulse response; P is the fractional sampling factor; The impulse response of i-th sub-channel is $\boldsymbol{c}^{(i)}(k)=c[(k+1)P-i-1]$; $\boldsymbol{n}^{(i)}(k)$ is the additive noise; $\boldsymbol{y}^{(i)}(k)$ is the input signal [6,7,8] satisfies

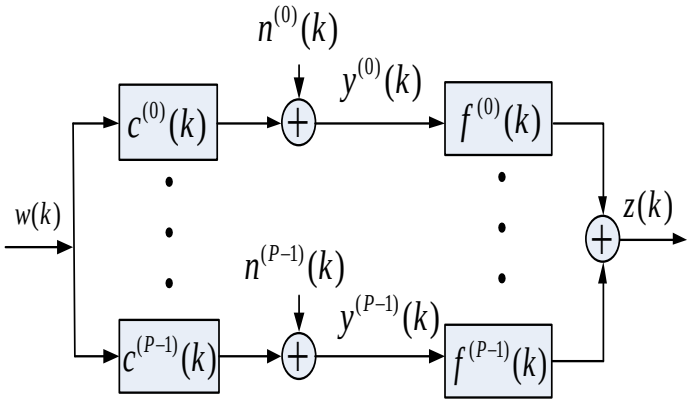

Fig.1 Structure layout of fractionally spaced preprocessor.

$$\boldsymbol{y}^{(i)}(k)=\sum_{j=0}^{N_c-1} s(j)\cdot c^{(i)}(j)+n^{(i)}(k) \tag{1}$$

where, N_c 为 is length of impulse response of the baud spaced channel

$\boldsymbol{f}^{(i)}(k)$ is the weight vector of the equalizer

$$\boldsymbol{f}^{(i)}(k+1)=\boldsymbol{f}^{(i)}(k)+\mu z^{(i)}(k)e(k)\boldsymbol{y}^{(i)*}(k)\ (i=0,\cdots P-1) \tag{2}$$

Where modulus of the signal is $R_2=\mathrm{E}\{|s(k)|^4\}/\mathrm{E}\{|s(k)|^2\}$, μ is step size, $e(n)=R_2-|z(k)|^2$ is the error.

The system output is

$$z(k)=\sum_{i=0}^{P-1}\boldsymbol{f}^{(i)}(k)*\boldsymbol{y}^{(P-i-1)}(k)=\sum_{i=0}^{P-1}\boldsymbol{f}^{(i)}(k)*[\boldsymbol{s}(k)*\boldsymbol{c}^{(P-i-1)}(k)+\boldsymbol{n}^{(P-i-1)}(k)] \tag{3}$$

Better noise robustness can be achieved for the processed speech signal by fractional spaced equalizer and it is more suitable for practical speech emotion recognition. The traditional speech emotion recognition can be improved, and the enhanced signal is suitable for fuzzy KNN classification.

Speech Emotion Recognition Based on Fuzzy KNN

Speech Emotion Feature Extraction. In this paper the speech feature used are short-term energy, short-term amplitude, zero cross rate and pitch frequency. The first three features are extracted in the time domain. The pitch frequency is estimated in frequency domain. The feature vector consists of the parameters of each frame and it is used as input in fuzzy KNN algorithm.

Fuzzy KNN Algorithm. In this paper we use fuzzy KNN for speech emotion recognition. The membership function is calculated for each sample based on the contribution of emotion features. First, the emotion dispersion is calculated according to each emotion type. The higher dispersion indicated higher uncertainty.

The calculation of emotion feature contribution is shown in the following steps

(1) For emotion type C, first the averaged feature value is calculated from training sample set X under C types of emotions, according to the fractional blind equalization. It is denoted as $\boldsymbol{P}_{ij}$ ($i=1,2,...C$, $j=1,2,...N$, N is emotion feature number). Then each utterance feature $\boldsymbol{P}_{ijn}$ (n is the emotion sample index) is normalized $B_{ijn}=P_{ijn}/\sum_{i=1}^{c}P_{ij}$

(2) Calculate the feature dispersion for certain emotion $\varphi_{ij}=\sqrt{\sum_{k=1}^{n}B_{ijk}}$.

(3) Calculate the contribution of each feature $\omega_i = \sum_{l=1}^{j} \varphi_{il}$

(4)The contribution u_{ij} of feature parameter φ_{ij} is $u_{ij} = \varphi_{ij} / \omega_i$

(5)The feature contribution and the Euclidean distance are summed $u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} d(x, X_j)}{\sum_{j=1}^{k} d(x, X_j)}$.

The contribution of each feature is taken as the sum weight and it is simple enough for KNN classification. The inner relation between features is modeled and the performance of emotion classification is improved.

Experimental Results

Emotional Speech Database. In this experiment, the Chinese speech emotion database is collected locally. The sampling rate is 16 KHz, and the quantization bit is 16bit. The utterances include 12 students (6 males and 6 females, age from 20 to 25) under five different emotion states (Sadness, Happiness, Neutrality, Fear and Anger). Text content consists of 25 sentences. The verified emotion utterance number is 790. In order to compare the recognition rates using different methods, we separate the dataset into two groups. One group is processed by fractionally spaced equalization. The other is not enhanced. Due to the personality difference all the emotional features are normalized. The normalized features are used for training and testing [3].

Extracting Emotion. Features In the training stage, we select 30 sentences for feature extraction. We can see that the pitch frequencies of neutrality and sadness are close to each other but their shot-term energy features are far away. For anger and happiness, the pitch frequencies are different, and anger has higher pitch frequency. Their shot-term energy features are close to each other. Results show that the combination of feature contribution and Euclidean distance is suitable for speech emotion analysis.

Experimental Results. Order to verify the effectiveness of the proposed fractionally spaced blind equalization and fuzzy KNN emotion recognition (FSE-FNN) we compared the recognition results with the traditional fuzzy KNN (FNN). In each test, the FNN and FSE-FNN are compared. In the first test k=5 and in the second test, k=11. The recognition results are shown in Fig. 2 and Fig. 3.

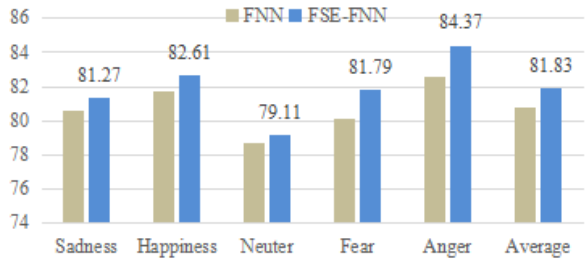

Fig.2 Recognition rate of two algorithms, k=5

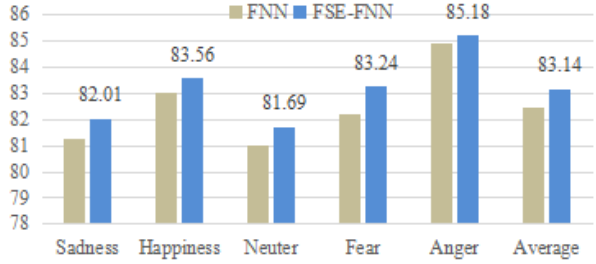

Fig. 3 Recognition rate of two algorithms, k=11

From the experimental results in Fig.3 and Fig.4 we can see that

(1) Under different k values, recognition rate is higher when k=11. The improvement is about 1.31%. When K increases, more neighbor samples are presented and error rate is reduced. However, the computational cost is also higher.

(2) With same k value, FSE-FNN performs better than FNN. The proposed fractionally spaced equalization brings better result for speech emotion recognition. The noise is removed in the preprocessing step. It is suitable for KNN emotion classification.

(3) With the same k value and the same classification algorithm, the anger emotion has the best result. The feature of anger is clearer and speaking rate is higher when people express anger.

Compared with neutrality and sadness, the feature is not very clear, and the recognition is more difficult.

Conclusions

In this paper, we study the traditional fuzzy KNN classifier in emotional speech recognition and we propose a novel preprocessing step using fractionally spaced equalization to cope with room acoustic channels. It has better result in our Chinese emotion database and the speech signal is enhanced for KNN classification. It has good practical value and may be applied to various real world applications, especially for room environments with multi-channel interference such as echo and reverberation.

Acknowledgements

This work was financially supported by the 2015 Institute of young teachers to support the development of teaching and research projects (2015yjzr026), the Research and Implementation on the key technologies of Video Conference System Based on Mobile Terminal, Anhui provincial network technology practice base project (2013sjjd070); Anhui Provincial Communication Technology professional and comprehensive reform pilot (2015zy148); Anhui Province university discipline of top-notch talent academic funding for key projects (gxbjZD2016098).

Reference

[1]Wenjing Han, Haifeng Li, Huabing Ruan, Lin Ma. Review on speech emotion recognition progress [J]. Journal of Software, 2014, 01 37-50.

[2]Chengwei Huang, Yun Jin, Qingyun Wang, Li Zhao, Cairong Zou. Multimodal emotion recognition based on speech and ECG signals [J]. Journal of Southeast University (Nature Science Edition), 2010,05 895-900.

[3] Jie Li, Ping Zhou. Study on the progress of speech emotion feature analysis [J]. Sensors and Micro Systems, 2012,02 4-7.

[4]Li Yan. GPS positioning correction algorithm based on fractional spaced equalization technology [J]. Computer Simulation, 2014, 11 438-441.

[5]Qi-rong MAO, Xin-yuPAN, Yong-zhao ZHAN, Xiang-jun SHEN. [J]. Frontiers of Information Technology & Electronic Engineering, 2015, 04 272-283.

[6] Qingjie Liu, Jinshu Chen, Yong Shen. An Arbitrary Equationally Spaced Predistortion Structure [J].Telecommunication Engineering, 2015, 06 665-670.

[7] Yecai Guo, Huapeng Wu, Hui Wang, Miaoqing Zhang. DNA Genetic Bat Algorithm Based Fractionally Spaced Multi-modulus Algorithm [J]. Acta Armamentarii, 2015, 08 1502-1507.

[8] Xiaolin Zhou, Weixiang Lin, Xin Zhao. A MLSE based new fractionally spaced equalization scheme [J]. Journal of Shanghai Normal University(Natural Sciences) ,2015 , 05 528-532.