Recognition of Aircraft Engine Sound Based on GMM-UBM Model

Shuai Yuan[1], Chengli Sun [2,*], Haoge Yang [2]

[1]*Science and Technology on Avionics Integration Laboratory,Shanghai, China*
[2]*Information School, Hangkong University, Nanchang, China*
[*]*Corresponding author*

Abstract—Gaussian mixture model-universal background model (GMM-UBM) is a commonly-used speaker recognition technology, and which has achieved good effect for detection speaker's sound. In this paper, we explore GMM-UBM method for use with abnormal aircraft engine sound detection. We designed a GMM-UBM based aircraft engine sound recognition system, which extracts MFCC feature parameters and trains the GMM-UBM models using maximum a posteriori (MAP) adaptive algorithm. Experimental results show the GMM-UBM based aircraft engine sound recognition system can achieve higher recognize rate in real-word aircraft engine sound test.
Keywords: Speaker Recognition; GMM-UBM; MFCC; Abnormal sound detection; MAP

1 Introduction

In daily life conditions, we often need to recognize many kinds of environmental sound. Sound recognition technology has been successfully used in the bio-medicine, agriculture, public security, astronomy and many other fields. Sound recognition also plays important role in Aircraft field. Aircraft engine sound signal contains a lot of important information, which makes the aircraft engine sound signal recognition being a very important tool for engine status diagnosis [1-2].

There are many research works are conducted in sound recognition community. Harma.M proposed an audio monitoring system using in the office environment [3]. They combine the tone and spectrum centroid features to detect the events occurred in the office. In [4], an audio classification system for detecting crime in the elevator is reported. Their system uses GMM models to classify and identify alarm sounds events. In literature [5], Ntalampiras et al. use mel- frequency cepstrum coefficient (MFCC), MPEG-7, and CB-TEO feature to detect dangerous sound events occurred in subway station. In [6], Cheng Qingyun employs GMM method to identify two kinds of abnormal sounds in office environment. In [7], Zhou proposed to construct feature set including MFCC and logarithmic domain subband energy. They use Adboost algorithm to select effective features to detect conference room sound event. In [8, 9], the researchers developed an audio monitoring system that uses MFCC and short-term energy as sound event characteristics. GMMs were used as recognition models to detect abnormal sound events in the elevator environment. Theoretically, the engine sound signal recognition are the same as the speech recognition, therefore the speech recognition technology can be explored for aircraft engine sound recognition. However, as aircraft engine sound acquisition is difficult, few aircraft engine sound processing research literatures are reported until now.

In this paper, a GMM-UBM based model is used to identify the engine sound, we use MFCC as sound features, and GMM-UBM models is trained by MAP optimization algorithm. The system can effectively recognize abnormal engine sound only require few data.

2 Engine Sound Recognition System Based on GMM-UBM

2.1 Universal Background Model

GMM-UBM is a high order GMM, it can reduce the variation of speaker utterance differences [10]. GMM is a linear weighted combination of a certain number of Gaussian probability density functions. The M-order Gaussian mixture model GMM uses the linear combination of M single Gaussian distributions to describe the distribution of frame features in the feature space:

$$p(x \mid \lambda) = \sum_{i=1}^{M} w_i b_i(x), \qquad (1)$$

where X is a D-dimensional eigenvector, w_i is the weight for each Gaussian density function which should satisfy the following condition:

$$\sum_{i=1}^{M} w_i = 1. \qquad (2)$$

The Gaussian distribution function $b_i(x)$, can be expressed as:

$$b_i(x) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}}\exp\{-\frac{1}{2}(x-\mu)'\Sigma_i^{-1}(x-\mu)\},$$

(3)

where μ is the mean vector and Σ is the covariance matrix. Thus, a complete Gaussian mixture model can be expressed as

$$\lambda = \{w_i, \mu_i, \Sigma_i\}_{i=1,2,\cdots m}.$$

(4)

GMM learning generally use the classic EM (Expectation-maximization) algorithm. According to the maximum likelihood estimation criterion, the likelihood of GMM model can be expressed as:

$$\log p(x \mid \lambda) = \sum_{t=1}^{T}\log p(x_t \mid \lambda)$$

(5)

According to EM algorithm, the weight, mean and variance of the *i*-th GMM model can be calculated by

$$w_i' = \frac{1}{n}\sum_{j}^{n}p(i \mid x_j, \lambda)$$

(6)

$$\mu_i' = \frac{\sum_{j}^{n}x_j p(i \mid x_j, \lambda)}{\sum_{j}^{n}p(i \mid x_j, \lambda)}$$

(7)

$$\sigma_i' = \frac{\sum_{j}^{n}(x_j - \mu_i')^2 p(i \mid x_j, \lambda)}{\sum_{j}^{n}p(i \mid x_j, \lambda)}.$$

(8)

The posterior probability of the i-th GMM model is

$$p(i \mid x_j, \lambda) = \frac{w_i p_i(x_t \mid \lambda_i)}{\sum_{k}^{M}w_k p_k(x_t \mid \lambda_k)}.$$

(9)

Speaker adaptation schemes should ideally be effective for small amounts of speaker-specific adaptation data and converge to a true speaker dependent estimate when a large amount of data is available. Most GMM-based speaker recognition systems are trained using maximum a posteriori (MAP) estimation [11].

For a particular Gaussian mean, with prior mean μ_0, the estimate is

$$\mu' = \frac{\tau\mu_0 + \sum_{t=1}^{T}\gamma(t)o_t}{\tau + \sum_{t=1}^{T}\gamma(t)},$$

(10)

where τ is a meta-parameter which gives the bias between the ML estimate of the mean from the data, and the prior mean o_t is the adaptation vector at time *t* from a T length set and $\gamma(t)$ is the probability of this Gaussian at time t. Usually, values of τ is often set between two and twenty.

Speaker recognition technology can be transferred into a sound recognition problem, so the sound recognition can be built based on GMM [12-19]. For a given sound feature vector $\{X_t\}$, *t*=1,2,..., *T*, assuming we have *K* classes of aircraft engine sound which is different from each other, the purpose of sound recognition is to find the class of sound *k*, whose corresponding GMM model *k* obtain the largest posterior probability P(*k*/*X*).

2.2 Engine sound recognition system architecture

Figure 1 shows the GMM-UBM based aircraft engine sound recognition system. In the training phase, after pre-process the engine sound data, we extract the MFCC feature of the engine sound. Then the UBM model is trained via EM learning algorithm. Finally, each individual GMM model is obtained using MAP of the UBM with the aircraft engine sound data. In the recognition phase, the matched likelihood is calculated used the variation between the output scores of GMM and UBM

$$S = \frac{1}{T}\sum_{t=1}^{T}\left[\log p_i(X_t \mid \lambda_{TAR}) - p_i(X_t \mid \lambda_{UBM})\right], (11)$$

where X_t is the eigenvector of a certain frame of audio to be recognized, λ_{TAR} and λ_{UBM} represent the target sound model and the UBM model, respectively.
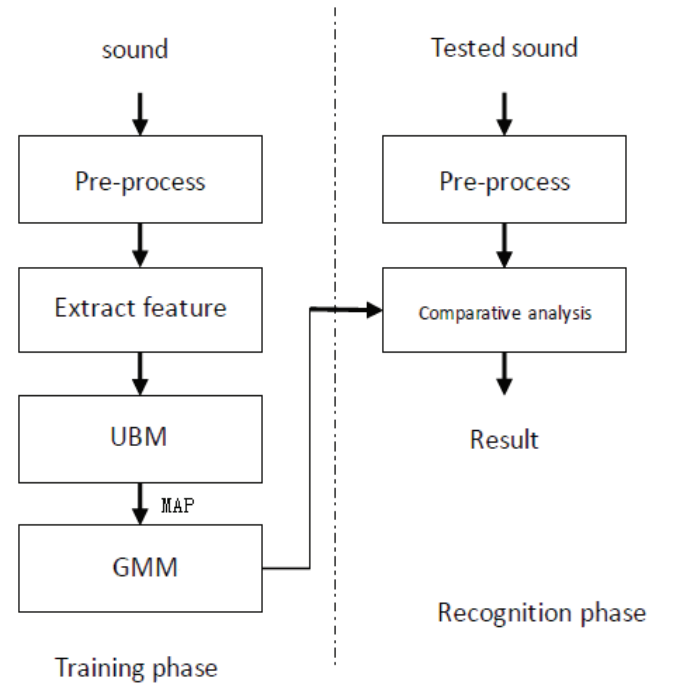

Figure 1. GMM-UBM based aircraft engine sound recognition system.

3 Experimental results and analysis

The aircraft engine sound database used for experiments contains a number of aircraft engine sounds from Civil Aviation Boeing 727 and Military Fighter F-15. In the training phase, we selected 61 engine audio for UBM training, including 12 start-up audio, 8 stop audio, 30 normal operation audio and 11 crash audio, the total length of the sound is 388s. the test sound are 6 start-up, 6 stop, 29 normal and 11 crash audio clips. MAP is made using 332s length audio. In the test phase, the test engine sound include engine start sound, stop sound, run sound and some engine crash sound. The total length of the engine sound used in the experiment is 1474s, containing of engine starting sound 347s, stop sound 205s, the normal operation

sound 629s, and crash sound 293s. All the sound data sampling frequency is 16 KHz, monaural recording, 16bit format. The sound features are MFCC coefficients. Recognition rate is used for evaluation the performance of sound recognition system.

$$\text{recognition rate} = \frac{\text{number of right}}{\text{total number of samples}} \times 100\%$$

3.1 The effect of mixture number on performance

Firstly, we compare the effect of different mixture component on the recognition performance. We use the 12 dimensional MFCC characteristic parameters. It can be seen from the experimental results that the best average recognition is achieved when the mixture number is 256. when increase the number of mixture components, the recognition rate is also increases, but when the mixture number increase to a certain value, the recognition rate no longer increases.

Table 1. Recognition performance with the increase of mixture number

Mixedness	Recognition rate					
	8	16	32	64	128	256
Start-up	0.333	0.500	0.500	0.666	0.666	0.666
Stop	0.666	0.666	0.666	0.833	0.833	0.833
Normal	0.896	0.965	0.965	1.000	1.000	0.862
Crash	0.636	0.727	1.000	1.000	1.000	1.000

3.2 The effect of different MFCC coefficients on performance

Standard MFCC coefficient only represent static coefficient feature. In this experiment, we will evaluate the combination of standard and dynamic MFCC feature. Let △MFCC be the first-order differential coefficient, and △△MFCC be the second-order differential coefficient. △MFCC and △△MFCC reflect the dynamic characteristics. It can be seen from Table 2 that the recognition rate of the△△MFCC and △MFCC can significantly improve recognition performance of the start-up sound, which is due to the dynamic characteristics of the start-up sound. In the experiments, we can see that HMFCC (the combination of static and dynamic coefficients) can improve system performance. For the normal and crash sounds, the HMFCC feature can obtain the 100% recognition rate.

Feature	Recognition rate				
	MFCC	△MFCC	△△MFCC	MFCC-△MFCC	HMFCC
Start-up	0.333	0.333	0.500	0.500	0.666
Stop	0.500	0.500	0.500	0.666	0.833
Normal	0.311	0.172	0.276	0.965	1.000
Crash	0.166	0.273	0.273	0.818	1.000

From above experiments we can see that increase the number of mixture components, recognition rate can improve as well, but when it reaches a certain value, the recognition rate began to decline. This is because the GMM model is a generative mode which relies on the amount of training data. The higher the mixture order is, the more complex the model will be. The amount of data in this experiment is relatively small. If the mixture number increases too much, it will result in excessive fitting problem and the GMM model performance become deteriorate.

4 Conclusion

In this paper, a GMM-UBM based aircraft engine sound recognition system is presented. This system can work well in little limited data case. Experimental results show that the recognition system with the combination of dynamic and static MFCC features could achieve higher recognition rate. In future research work, we will investigate incorporating more prior knowledge about the aircraft engine sound into the framework, such as the sound continuity property, sound duration property, to further improve the robustness and accuracy for the proposed system. We also interest in applying DNN (Deep Neural Networks) for aircraft engine sound recognition.

Acknowledgment

References

1. Chenlei. Sound Recognition Technology for Helicopter Using Vector Quantization [D]. Shandong Univerdity.2014.
2. Leida. Study on Prediction of Health Status of Civil Aviation Engine Based on Intelligent Learning Model[D]. Harbin Institute of Technology.2013.
3. Harma E, McKinney F, Skowronek J.Automatic Surveillance of the Acoustic activity in our Living Environment.Proc.IEEE Internation Conference on Multimedia and Expo.DigitalSignaoProcess.Netherlands,2005,152-152.
4. Radhakrishnan R,Divakaran A.Systematic Acquisition of Audio Classes for Elevator Surveillance.SPIE Image and Video Communications and Processing.San Jose,CA,USA,2005,64-17.
5. Ntalampiras S, Potamitis I, Fakotakis N. An Adaptive Framework for Acoustic Monitoring of Potential

Hazards. EURASIP Journal on Audio, Speech, and Music Processing.2009. 2: 1-15

6. Qingyun Cheng. Research on Two Kinds of Abnormal Sound Recognition in Office Environment Based on GMM[D]. Anhui University.2016.

7. Zhou X, Zhuang X, Liu M, Tang H, Hasegawa-Johnson M, HMM-based Acoustic Event Detection with AdaBoost Feature Selection. Lecture Notes in Computer Science. 2008. 4625: 345-353.

8. Jianjie Ye. Design and Implementation of Supervised and Staged Abnormal Sound Detection System[D]. South China University of Technology.2014.

9. Xiaoyun LV, Hongxia Wang. Research on Abnormal Audio Recognition Algorithm Based on MFCC and GMM [D]. Southwest Jiaotong University.2010.

10. Bingxi wang, Qudan, Pengxuan. Practical Fundamentals of Speech Recognition[M]. Beijing:National Defense Industry Press,2005.1.

11. Zhengping Zhang. Research on Continuous Adaptive Algorithm Based on GMM-UBM Speaker Model[J]. Communication Power Technology,2016.33(2).

12. Akinori Ito,Akihito Aiba.Dectetion of Abnormal Sound Using Multi-stage GMM for Surveillance Microphone.International Conference on Information Assurance and Security.2009.

13. Ntalampiras S, Potamitis I, Fakotakis N. An Adaptive Framework for Acoustic Monitoring of Potential Hazards. EURASIP Journal on Audio, Speech, and Music Processing.2009. 2: 1-15.

14. S Bou-Ghazale, J H L Hansen.A comparative study of Traditional and Newly Proposed Feature for Recognition of Speech under Stress[J].IEEE Trans on Speech and Audio Processing,2000;8(4):429-442.

15. J H L Hansen,M A Clements.Source Generator Equalization and Enhancement of Spectral Properties for Robust Speech Recognition in Noise and Stress[J]. IEEE Trans on Speech and Audio Processing,1995;3(5):407-415.

16. B D Womack,J H L Hansen.N-Channel Hidden Markove Models for Combined Stressed Speech Classification and Recognition[J].IEEE Trans on Speech and Audio Processing,1999;7 (6):668-677.

17. P. kenny, G. Boulianne, P. Ouellet,and P. Dumouchel. Joint factor analysis versus eigenchannels in speaker recognition[J]. IEEE Trans on Speech and Audio Processing.2007;15 (4):1435-1447.

18. T. Hasan, John H. L. Hansen. A study on Universal Background Model training in Speaker Verification[J]. IEEE Trans. Audio Speech Lang. Process. September 2011, Vol. 19, 7, pp. 1890 – 1899.

19. C J Leggetter, P C Woodland. Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Model[J]. Computer Speech and Language,1995