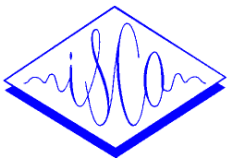

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

Two-Stage Speaker Identification System Based On VQ and NBDGMM

Si Luo, Hu Qi Xiu
siluo@263.net, xxs_dau@mail.tsinghua.edu.cn
Department of Computer Science, Tsinghua University

Abstract

In this paper, a new speaker identification system is presented. The system can be divided into two subsystems, one close-set speaker identification system and one speaker verification system. The VQ model is used in the close-set speaker identification system and a new method called NBDGMM (Normalization Based on Difference of GMM) is introduced. Experiments have been done to prove that this speaker identification system is efficient.

1.Introduction

This Speaker identification system can be divided into two stages. The first one is a close-set speaker identification system. The second one is a speaker verification system. In this paper, the performance and the correlation of VQ and Gaussian Mixture Models (GMM), two models of close-set speaker identification system, are carefully compared. The second stage of our system is a speaker verification system; score normalization is necessary for speaker verification system. In this paper, a new normalization method based on GMM is presented. This new normalization method is called Normalization Based On Difference of Gaussian Mixture Models (NBDGMM). We use the difference between the new model (modified by the unknown speech) and the original speaker model as the value to decide whether the unknown speaker should be accepted or rejected. The new method is effective and concise; it does not use any information of background speakers. Satisfying ERR of our system is achieved by using the methods descript above.

2.Two Models of Closet-Set Speaker Identification System

VQ models are widely used in the text-independent speaker identification systems. [1] VQ algorithm is used to create a speaker model for every speaker in the training set. And some distance measure is used in the recognition stage to compute the distance between the unknown speaker and all the speakers in the training set. Then the speaker whose distance is the minimum is selected out as the result corresponding to the unknown speaker.

The figure of the speaker identification system based on VQ is shown as the following,

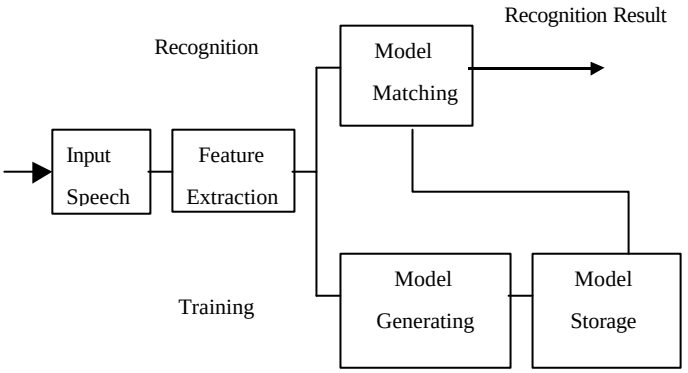

Fig 1. The Speaker Identification System Based on VQ

The following formula is used in VQ model to compute the distance between the unknown speaker and the speakers in the training set.

$$Dj = \frac{1}{T}\sum_{k=1}^{T} \min_{1 \le p \le M} d(a_k, b_p^j) \qquad (1)$$

Dj $(1 \le j \le N)$ is the distance between the unknown

speaker and the speakers in the training set. a_k is the speech feature. T is the total number of the speech data. M is the total number of the codewords in a codebook.. b_p^j is the pth codeword of speaker j.

Gaussian Mixture Models (GMM)[2] are also used in speaker identification systems. A lot of research has been done to apply this method to speaker recognition these years [3][4]. A Gaussian Mixture Model can be created to represent every speaker in the training set. Expectation-Maximization (EM) algorithm is used to compute the GMM for every speaker. And in the recognition stage, the speaker who' s GMM has the largest likelihood to the unknown speaker is selected out as the recognition result.

We have compared the recognition accuracy of the VQ and GMM models for close-set speaker identification systems. The result is shown in the following figure.

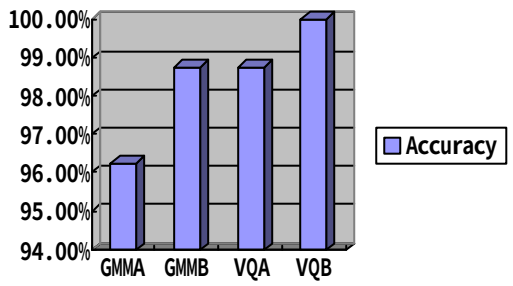

Fig2. The Accuray of VQ and GMM Models

The result is got on the same experiment database. So we can see from this experiment that VQ model can get better performance on the close-set speaker identification system.

The correlation of these two models is also evaluated. We propose a function to compute the correlation.

$$correlation = \frac{The\ times\ if\ the\ first\ choice\ is\ same}{Total\ number\ of\ \exp eriments} \quad (2)$$

In our experiments, there are 80 speakers in the training set and 40 speakers who are not in the training set. They each provide 10s speech data as recognition speech. From the experiments, we can get that the correlation between GMMA and GMMB is 80%, the correlation between VQA and VQB is 85%, the correlation between GMMB and VQB is 55%. From this, we can see that the correlation between VQ and GMM is fairly little. We all know that the less correlate between two subsystems the better for the performance of the whole system. This fact provides us the theory foundation that we can expect better performance by combining the two subsystems to a whole system.

3.The New Normalization Method for GMM

For the GMM model, we use the following formula to compute the result of speaker identification.

$$l_k = \arg\max \ p(Z/l) \quad l_k(l_k \in \{l_i / i = 1,2..,N\}) \quad (3)$$

When this method is applied in the close-set speaker identification system, we can get fairly good performance as mentioned above. But if we apply this method into the open-set speaker identification system, which means that we must bring forward a threshold to decide whether to accept or reject the unknown speaker, the result is not satisfied. From many experiments, we can see that there is much deviation among the values we get from the function 3. This fact makes much trouble for us to decide the specified threshold.

So we propose a new method for the GMM model to normalize the score of the distance measure. This normalization method is called Normalization Based on Difference of GMM (NBDGMM). We know that the training stage of GMM is the steps of Expectation and Maximization. Our method is to use the input speech of the unknown speaker and the models of the speakers in the training set to make a step of EM, then we can get a new model, the distance between the new model and the specified speaker model in the training set can be used as the result of recognition. From our experiments, we can see that we can easily set a threshold to decide whether to accept or reject the unknown speaker by this measure.

The process of the verification stage is expressed in the following figure:

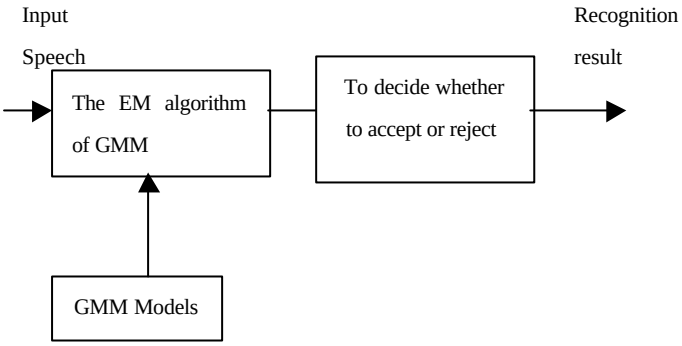

Fig3. The verification process based on NBDGMM

There are also many different ways of NBDGMM. For example, when the difference of the new model and the models in the training set is computed, we can only compute the difference of the mean, or compute the difference of the deviation, or compute the difference of the probabilities of different mixtures, or compute the combination of any of these three factors. From experiments, it is found that if we only compute the difference of the mean, the highest verification accuracy can be expected.

4.Database, Speaker Identification System and Experiments

An own-recorded database is used for our experiments. This database contains total 120 speakers, who speak Chinese. The speakers are from different parts of China, 74 of them are male, 46 of them are female. All the data are recorded from microphone. The sample rate is 8khz. We use 50s for every speaker to train the system. And we use 15s to test the system.

The speaker identification system is expressed in the following diagram:

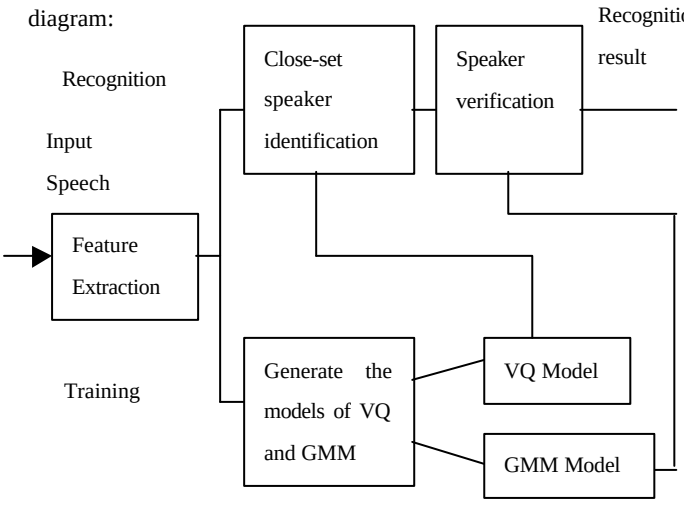

Fig4. The Two-Stage Speaker Identification System

The ERR (Equal Error Rate) is:

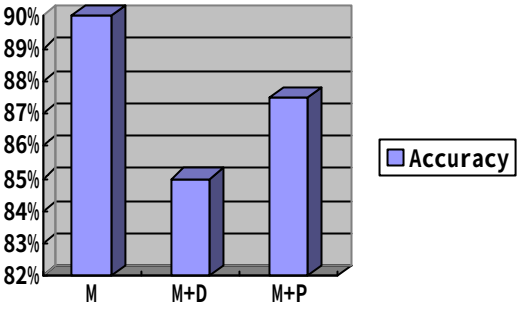

Fig5. The accuracy of our new two-stage speaker identification system

Here, M means that we only compute the difference of the mean between the new model and the speaker' s model in the training set, M+D means that we compute the difference of both the mean and the deviation, M+P means that we compute the difference of both the mean the probabilities.

5. Conclusion

In this paper, we compare the efficiency and correlation between the VQ and GMM models for the close-set speaker identification system. We also propose a new normalization method called NBDGMM for the GMM speaker verification system. Then a two-stage speaker identification system is introduced based on the fact we get from the previous work. Good performance has been shown in our experiments based on our new system.

6.References

[1]. S.Furui "Recent Advances in Speaker Recognition" First International Conference, AVBPA ' 97

[2]. Reynolds, D. A. A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification, PhD Thesis, Georgia Institute of Technology, September, 1992

[3]. C. Martin-Del-Alamo, J. Caminero-Gil, C. De-La-Torre, L. Hernandez-Gomez, Tecnologia Del Habla, Telefonica I+D, Madrid (SPAIN), Discriminative Training of GMM for Speaker Identification, ICASSP96

[4]. C. Tadj, P. Dumouchel P. Ouellet, GMM Based Speaker Identification Using Training-Time-Dependent Number of Mixtures, ICASSP 98