Text Dependent Speaker Recognition Study

Hui-hong XU

(College of Information Engineering ,Eastern Liaoning University, Dandong 118000, L iaoning, China)

Email :xuhuihong001@163.com

Key words: Speaker Recognition; HMM; LBG Algorithm; Text Dependent

Abstract: The speaker recognition is a sort of biometrics according to person's sound. This paper proposed a method that extracted characteristic parameter from sound signal by LPCC and MFCC. Improving LBG algorithm, training and testing the samples by continuous left-right HMM, A speaker recognition algorithm was given. Trough experiment, the result in 4 changes continuous left-right HMM is best.

1 Introduction

Speaker recognition[1] also known as voiceprint recognition which extract information from speaker voice and judge the identification, it with fingerprint, face and iris recognition belongs to the category of biometrics. speaker recognition is divided into aspects of the research content, the speaker identification and speaker to confirm. The former is to judge the voice of the differential input is belong to who and give the acceptance or refusal, the later is to determine whether the input-voice belong to the speaker.

The content of speech recognition is predetermined which is called text-dependent recognition. If no matter what words are said for speaker recognition, we called this text-independent speaker recognition.

At present, there are several methods to study the speaker recognition, speaker recognition based on the template, vector quantization, Gauss mixture model , Hidden Markov models and artificial neural network.

The paper’s content is based on the text dependent speaker recognition, and which use linear cepstrum coefficient and Mel frequency cepstrum coefficient feature parameters of speech signal extraction ,the speech signal is coded by the improved LBG codebook, and a continuous left-right Markov models are used on the training and testing samples. speaker recognition based on content is realized.

2 The System Workflow

The system work flows as shown in figure 1.the training of voice and speech recognition are preprocessed. HMM model and training model for every speech material after feature extraction and codebook generation are built. and then the speech signals of waiting for verifying are matched with every model ,the matching probability are calculated, the maximum matching probability will be the result of recognition[2].

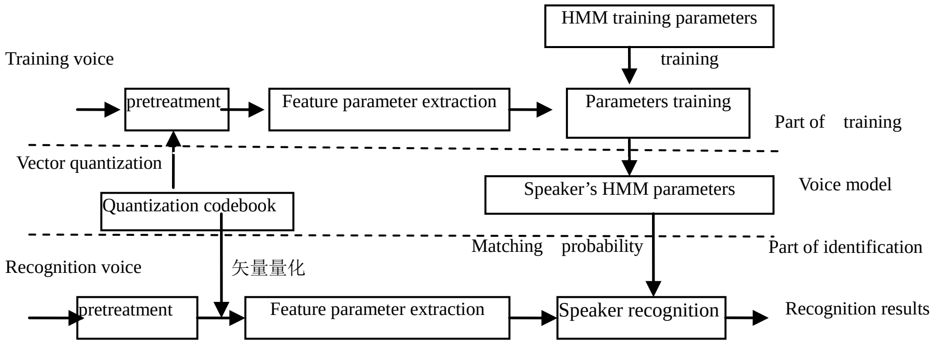

Fig.1 flow chart of text-dependent speaker recognition based on HMM technoloy

3 Feature Extraction and Improved Algorithm of Codebook Generation

Before carrying out the analysis and processing for speech signals. them must be pretreated ,including amplification voltage of signal samples, sampling, pre-emphasis, endpoint detection and noise removal etc. After pretreatment, the feature extraction and vector quantization coding for speech signals are finished.

3.1 Linear prediction cepstrum coefficient(LPCC)

Linear prediction cepstrum coefficient (LPCC) [3]is a linear prediction coefficient(LPC)which expressed in the cepstral domain. This feature is based on the speech signal as autoregressive signal assumption. and use linear analytical predictions to obtain the cepstrum coefficients. typical parameters of LPCC for voice solving flow as shown in Figure 2.

Owing to the usage of channel system function's minimum phase characteristics, LPCC can avoid the complexity of phase convolution. Advantages of LPCC are a small amount of calculation and easy to realize.

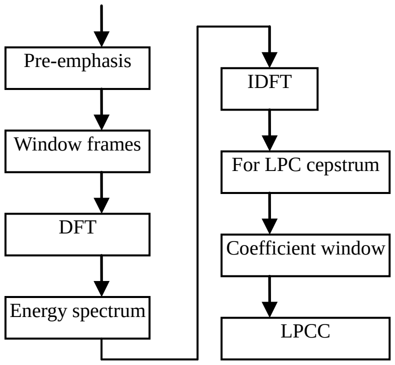

Fig.2 Lpcc soling flow chart

3.2 Mel frequency cepstrum coefficient spectrum

Mel frequency cepstrum coefficient (MFCC) spectrum[4] is transformed into a nonlinear spectrum based on Mel frequency standard, then switch to the spectral domain. By fully considering the characteristics of hearing people, and without any assumptions, the MFCC parameter has a good recognition performance and anti noise ability

The calculation process of Speech MFCC parameters is shown in Figure 3.

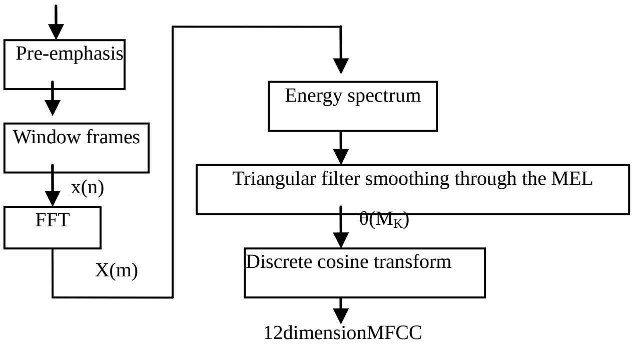

Fig.3 Mfcc calculation process

3.3 Improved Codebook Generation Algorithm

3.3.1 Codebook Generation Principle of Vector Quantization

Vector quantization (VQ)[5] is a technique for data compression, the voice signal must be vectorize during training and recognition, During training, VQ is used to cluster analysis, which means some representative samples can be selected from training samples and be made to reference codebook .During recognition ,the sample data will be replaced by the given code of codebook according to minimum distortion principle.

3.3.2 Improved LBG Algorithm

The large influence on recognition by the quality of the codebook is proved after many experiments, LBG algorithm is a kind of commonly VQ algorithm that proposed by Linde et al. and the basic idea of LBG is to choose an initial codebook C_0, the initial classification P_0 is determined according to the principle of the nearest neighbor, quantizing distortion D_0 is computed, and update the code vector C_i,,fall in the samples' center of the division P_i, so as to get a new codebook c1 and new division p1,the process is repeated until the decline of the ratio of the quantization distortion $(D_{L-1} – D_L) / D_{L-1}$ for the before and the last is less than the set value θ .

The problem of the LBG algorithm is dependent on manual to determine the initial codebook, the improper distribution of initial codebook usually cause the subset empty in the process of the use, while another partition subset can contain excessive number of samples. Therefore, in this paper, the improved LBG algorithm obtain the partition of sample by splitting method, So all of the samples are subset of the initial portion, and each partition subset will continue to split into two, the action will stop until the desired partition number m or the number of samples in each subset are smaller than the parameter θ , This will avoid manual to determine the initial codebook, but also to ensure each subset contains sample is not empty, and to eliminate the greatest amount of empty bag cavity .

4 Experiments and Analysis of Result

Total 10 volunteers participated into the experiment in noise-free environment, and total 40 wav format sound files with the same content from 5 male and 5 female speakers have been recorded, 20 files for the generation of characteristic state sequence of each person, and the other 20 files for test. The sampling frequency is $11025H_Z$, the quantification is 16bit, the identification test adopts single-channel voice, the frame length of recording data is 512 points (the number of sampling points), frame shift is 256 points, pre-emphasis is $1\text{-}0.95Z^{-1}$, with Hamming window speech extracted features by frames.

4.1 Training of Voice

The speech is modeled by the continuous left-right HMM in the experiment,everyone's sound files are input respectively and the sequences of characteristic value are generated by improved LBG and LBG algorithm, training by Baum-welch algorithm, we get A HMM(λ=(π,A, B)) for everyone,the state of the model number is six.

4.2 Recognition of Voice

Input the feature vector(Q) of voice to be measured, calculate the Q by forward-backward algorithm, and calculate matching probability $P(O|\lambda)$ for the HMM$(\lambda=(\pi,A,B))$ model of speech, when its value is greater than a given threshold, it is true, false otherwise.

4.3 Analysis of experimental results

In different parameters, different codebook algorithm combination circumstances, the continuous, left-right model of 2-transfer,3-transfer and 4-transfer are tested, the result is showed in the following table 1.

Table 1 Experimental different HMM parameters

Codebook algorithm+parameters	2-transfer(%)	3-transfer (%)	4-transfer (%)
LBG+lpcc	93.5	93.2	94.0
LBG+mfcc	94.0	92.5	94.6
LBG(improved)+lpcc	93.7	92.6	95.0
LBG(improved)+mfcc	94.5	93.5	96.3

As the conclusions are as follows:

(1) the codebook optimization method significantly improves the recognition rate, the elimination of large bag cavity space is very effective method.

(2) for different values of the left -right HMM structure, the number of state transfer has great effect on the experimental results, the experimental data can be found that 4 transfer ,left-right type is better than the 2 or 3 transfer.

(3)MFCC parameter extraction is better than LPCC under the same number of state, the same

HMM model which can be proved that the better recognition performance with MFCC. Among all the tests, the 4 transfer , left - right type, improved LBG algorithm, MFCC parameter extraction experiment mode is the best.

References

[1] Wang Yu,Mu Zhichun. A Survey On Multmodal Biometrics Technologies[J]. Computer Applications and Software,2009(12):31-32

[2] Liu Yao-he, Song Ting-xin. Speech Recognition and Control Applications Technology [M]. China Science Press, 2008.

[3] Liu Yun-bing, Zhu Yan-cheng, Peng Jin et al. Performance of Hidden Markov Model in Speaker Recognition [J]. Software Guide,2006,12(12):15-16.

[4] Yao Tian-yen. Digital Speech Processing [M].Huazhong University of Science and Technology Publishing House,2007.

[5] Li Tao, Wang Jun-pu, Wu Xiu-qing et al. The SVM Learning Strategy Based on Improved LBG Algorithm [J]. Journal of Fudan University (Natural Science), 2004, 43 (5) : 42-43.