Research of Speaker Recognition Based on Combination of LPCC and MFCC

Yuan Yujin
Electronic Information Engineering
Training and Experimental Center
Handan College
Handan, China
yiyi_yu1980@yahoo.com.cn

Zhao Peihua
Electronic Information Engineering
Training and Experimental Center
Handan College
Handan, China
dani88021@yahoo.com.cn

Zhou Qun
Department of Information
Engineering
Handan College
Handan, China
zhouqun888@yahoo.com.cn

***Abstract*—Spearker recognition used widely in our lives is an important branch of authenticating automatically a speaker's identity based on human biological feature . Linear Prediction Cepstrum Coefficient (LPCC) and Mel Frequency Cepstrum Coefficient (MFCC) are used as the features for text-independent speaker recognition in this system. And the experiments compare the recognition rate of LPCC , MFCC or the combination of LPCC and MFCC through using Vector Quantization (VQ) and Dynamic Time Warping (DTW) to recognize a speaker's identity. It proves that the combination of LPCC and MFCC has a higher recognition rate.**

Keywords- speaker recognition; LPCC; MFCC; VQ; DTW

I. Introduction

The goal of speaker recognition is to automatically authenticate a speaker's identity by his/her voice among a population. It is widely applied to many fields from confidential data access to audio indexing in multimedia. Extracting speaker's personal audio traits is the key to speaker recognition. Linear Prediction Cepstrum Coefficient (LPCC) reflects the difference of the biological structure of human vocal track, and Mel Frequency Cepstrum Coefficient (MFCC) is based on the human ears' non-linear frequency characteristic[1]. This paper presents a speaker recognition system based on the Vector Quantization (VQ) and Dynamic Time Warping (DTW), which uses the combination of LPCC and MFCC as features, and compares the recognition rate of speaker recognition which used LPCC, MFCC, or the combination of LPCC and MFCC as features through a series of experiments.

II. Extracting Features

A. LPCC and Iits Differential Cepstrum Coefficient

LPCC as a very important feature reflects the differences of the biological structure of human vocal track. Computing method by LPCC is a recursion from LPC Parameter to LPC cepstrum according to All-pole model. Its recursion is as follows:

$$\begin{cases} c_1 = a_1 \\ c_n = a_n + \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k}, 1 < n \le p \\ c_n = \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k}, n > p \end{cases} \quad (1)$$

Where $a_1, \ldots, a_p$ is p-order LPC feature vector, c_n , n=1, …, p, p is the first p values of the cepstrum.

The definition of the differential cepstrum coefficient of LPCC[2] (*LPCC) is as follows:

$$c_m(t) = \sum_{i=-k}^{k} i \cdot c_m(t+i) / \sum_{i=-k}^{k} i^2 \quad (2)$$

Where $c_m(t)$ and $c_m(t+i)$ both express a frame of speech parameter, k which is a constant often chooses 2.

B. MFCC and Its Differential Cepstrum Coefficient

The MFCC which is different from other frequency cepstrum focuses on the human ears' non-linear frequency characteristic, and the size of Mel frequency corresponds to the relation of actual frequency's logarithmic distribution on the whole and accords with the human ears' characteristic . The idiographic relationship between Mel frequency and actual frequency is as follows:

$$Mel(f) = 25951g(1 + f / 700) \quad (3)$$

Where the dimension of actual frequency f is Hz.

The flow of calculateing MFCC parameter is as follows:

978-1-4244-6585-9/10/$26.00 ©2010 IEEE

1) Ascertain the number of used serial points of each frame speech. N equals to 256 in this system. First pre-emphasize each frame $s(n)$ and transform pre-emphasized $s(n)$ through Discrete Fast Fourier Transform (DFFT) , then gain discrete power spectrum $S(n)$ through the square of MOD .

2) Acquire power value after calculating $S(n)$ through M filters $H_m(n)$,namely calculate the sum of the product of $S(n)$ and $H_m(n)$ on the point of each discrete frequency to gain M parameter p_m , where m=0, 1, ..., M-1.

3) Calculate the natural logarithm of p_m to gain L_m , where m=0, 1, ... , M-1.

4) Calculate the discrete cosine transform of L_0 , L_1 ,..., L_{m-1} to gain D_m ,where m=0, 1, ..., M-1.

Ignore D_0 which represents the ingredient of direct current[3], and make D_1 , D_2 , ..., D_k as MFCC parameter . At last , calculate MFCC with one-order differencing to gain a group of new MFCC differential coefficient as a group of feature vector.

The calculative formula of differncing parameter of MFCC (*MFCC) is as follows:

$$d(n) = \frac{1}{\sqrt{\sum_{i=-k}^{k} i^2}} \sum_{i=-k}^{k} i \cdot c(n+i) \tag{4}$$

Where c and d both represent a frame of speech parameter, k which is a constant often chooses 2, here differencing parameter is called as the linear combination of the first 2 frames and the second 2 frames parameter of current frame[4].

III. Method of Speaker Recognition

A. Traditional Method of Speaker Recognition

Vector Quantization (VQ) is an important method of digital signals processing. Regard each speaker's speech which waits to be recognized as a signal source which is represented by a codebook formed by clustering the feature vector extracted from the trained speech series of this speaker. Training and recognizing are two steps of VQ. Training is namely establishing N codebooks of N speakers and these codebooks are not superposing each other in the feature space. In the step of recognizing, firstly extract a group of vectors from speech waiting to be recognized, then use N codebooks founded in the system to quantize these vectors with VQ to gain O={ o_1 , o_2 ,..., o_l}, namely judge that group of vectors in accordance to the codebook in feature space. Assume that the number of code words of these N codebooks is M .

Dynamic Time Warping (DTW) is an early and classical arithmetic, which based on the thought of dynamic programming and can resolve the matching problem of the difference of speech's length. This paper stores LPCC and MFCC distilled by speech processing according to frames. Referenced template and test template was compared with DTW arithmetic for template matching, calculate the distance between referenced template and test template by DTW, and the template of the minimum distance is the best matching result[5]. The gist of using DTW arithmetic is cepstrum anamorphic measure .

Fig.1 shows the traditional approach of speaker recognition that only extracting LPCC and MFCC with VQ and DTW .

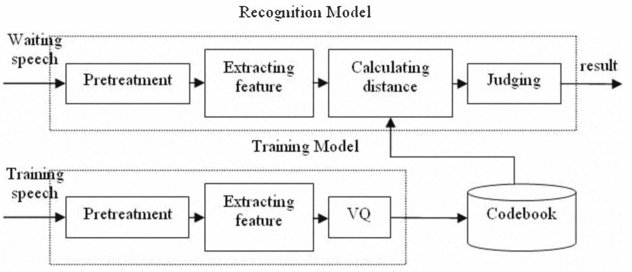

Figure 1. Traditional speaker recognition

B. Speaker Recognition with Combination of Features

Mel frequency cepstrum reflects the human ears' non-linear frequency characteristic, Linear Prediction Cepstrum reflects the differences of biological structure of human vocal track, their one-order differential coefficients (*MFCC , *LPCC) both describe their own dynamic characteristics. When use MFCC parameter to recognize, the system tends to judge this speaker as a legal speaker of the system if a speaker embezzles the password. So use LPCC which reflcts the difference of the biological structure of human vocal track and its one-order differencing as a feature to enhance the security of the system. First the template matching by DTW arithmetic is applied to two combined feature vectors in the process of recognition. Then set the distance threshold p of template matching in order to reduce miscarriage of justice. If the distance of template matching d

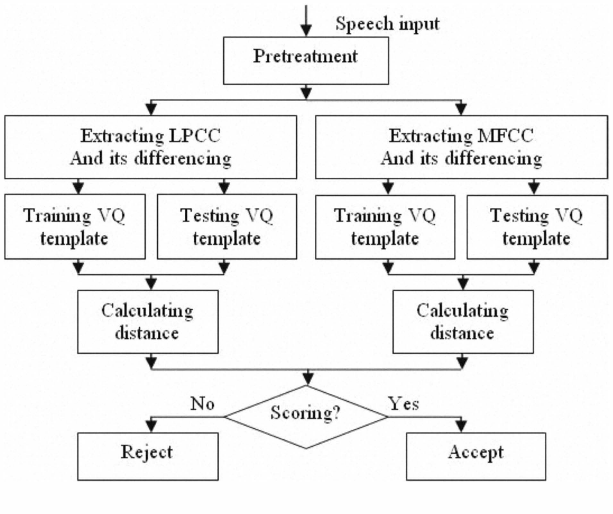

Figure 2. Speaker recognition based on LPCC and MFCC

is larger than p, the speaker is considered as an illegal speaker, even if he was a legal speaker. Compare the script code i, j of the ninimum distance. If i equal to j, the speaker is judged as a legal speaker, even if he was an illegal speaker.

IV. Result of The Experiment

A. *Experiment Facilities*

In normal laboratory conditions, 40 people's continuous digit speech library is recorded with the length of speed string between 1 to 10, and continuous digit speech (i.e. 0-9) occurrences of each string length are approximately the same. Each length of continuous digit speech string for each one was recorded 30 times, the top 10 times for model training, the end 20 times for testing.

This experiment used Matlab 7.0 as the development environment, and did three kinds of experiments towards different feature vector with DTW arithmetic. Accuracy[6] is used as the standard of evaluating the recognition performance of the system in this paper.

Experiment one : Train the template and test the system to form the whole system of speaker recognition with 12-order LPCC coefficient and its one-order differencing *LPCC .

Experiment two : Train the template and test the system to form the whole system of speaker recognition with 16-order MFCC coefficient and its one-order differencing *MFCC .

Experiment three : Use the combination of two kinds of feature vector of experiment one and experiment two ,and adjust the result of recognition with DTW arithmetic .

B. *Results ofThe Experiments*

Table 1 shows the results of the experiments. The effect of using solely LPCC and *LPCC or MFCC and *MFCC is not as that of using the combination of LPCC, MFCC, *LPCC and *MFCC. LPCC makes up for MFCC's failure in describing the characteristics of vocal track, moreover *LPCC and *MFCC reflect the dynamic characteristic of speech and vocal track, so the combination of these feature vectors better reflect the individual characteristic of a speaker .

TABLE I. Result of The Speaker Recognition Experiment

	LPCC, *LPCC	**MFCC, *MFCC**	**Combination**
Different speakers, different contents	95.52%	96.30%	97.12%
Different speakers, the same partly contents	87.65%	89.82%	91.43%
The same speakers, different contents	79.74%	81.48%	82.54%

V. Conclusions

The paper used VQ and DTW arithmetic to recognize a speaker's indentify through extracting the combination of LPCC, MFCC, *LPCC and *MFCC, and compare strenghs and weaknesses of using LPCC, MFCC, *LPCC,*MFCC and their combination as speech features. The experiments showed the combination of LPCC, MFCC, *LPCC and *MFCC improved the performance in aspects of the recognition rate and recognition time .

References

[1] Zhiyou Ma ,"Further Extraction for Speaker Recognition",IEEE International Conference on Systems,Man and Cybernetics,4153-4158,2003.

[2] Campbell J.P,"Speaker Recognition :A Tutorial",Proc.of the IEEE,vol.85,no.9,pp.1437-1462,sep.1997.

[3] Fakotakis N,Sirigos J,"A high performance text-independent speaker identification system based on vowel spotting ang neural nets",proceedings of the IEEE,"Speech and Signal processing ".Atlanta,GA,USA,1996.

[4] Furui S,"Recent advances in speaker indentification",Pattern Recognition Letters,vol.18,no.9,pp.859-872,1997.

[5] Deng Haojiang,Wang shoujie,Xing Cangju,Liu Qian,"Research of Text-Independent Speaker Recognition Using Clustering Statistic ",Jurnal of Circuits and Systems,2001.