Australian Journal of Basic and Applied Sciences, 4(9): 4357-4364, 2010
ISSN 1991-8178

Modified Mfcc for Speaker Recognition

[1]Alireza Salahshour Vaskas [2]Ahmad Esfandiyari [3]Shahaboddin Shamshirband

[1]Sama Organization (Affiliated with Islamic Azad University) Sari Branch
[2]Islamic Azad University, Sari Branch
[3]Young Researchers Club, Islamic Azad University Chalous Branch

Abstract: in this research is described a method for feature extraction from the signal that is based on the MFCC analysis. The researchers discover that the major fraction of the speech information is in the low frequency and the information in high frequency is very rubbish. So they make a method that emphasizes the low frequency of a signal. In this method the emphasis is done using MEL-frequency filter bank that the filter bank is wrought from some band pass filter and we convert the filters in this filter bank to low pass filter and compare the MODMFCC with MFCC.

Key words: Cestrum, MFCC, vocal tract

INTRODUCTION

In everyday life, it is a common experience for people to be able to identify speakers by their voices.

In speech technology, many attempts have been made aiming at modeling such human ability for a number of applications, such as in security access control systems, or in specific investigation fields like computational forensics.

For speaker recognition the feature extraction is an important section and many type of feature extraction is used in many studies such as LPCC that is referred to the vocal tract as an all pole filter and the LPC coefficients are the all pole filter coefficients and finally the LPC technique is combined with cepstrum technique. Another feature extraction technique is Mel frequency cepstral coefficients that are multiplying short term Fourier transform magnitude at the Mel frequency filter bank.

In this text the MFCC technique is used for feature extraction.

The speaker recognition consists of following five steps:

- framing and windowing the speech signal
- preprocessing
- main analysis
- modeling & matching

After framing & windowing & preprocessing we analyze the signal of each frame with MFCC analysis. The MFCC is a main analysis in speech and speaker recognition that is based on emphasis the low frequency of a signal using multiplying the signal in MEL frequency filter bank.

Finally Speaker models are constructed from the features extracted from the speech Signal. Then compute a match score that is a measure of the similarity between the input feature vectors and some model.
But we change the filter bank in MFCC analysis in this paper and modify it.

Mfcc Analysis:

Mel-frequency cepstrum coefficients (MFCC) are well known features used to describe speech signal. They are based on the known evidence that the information carried by low-frequency components of the speech signal is phonetically more important for humans than carried by high-frequency components. Technique of computing MFCC is based on the short-term analysis, and thus from each frame a MFCC vector is computed.

MFCC is similar to the cepstrum calculation except that one special step is inserted, namely the frequency axis is warped according to the Mel-scale. Summing up, the process of extracting MFCC from continuous speech is illustrated in Figure 2.1.

Corresponding Author: Alireza Selahshoor 1electronic and Computer Department, Sama Organization (Affiliated with Islamic Azad University) Sari Branch.
E-mail: Ali.selahshoor@gmail.com

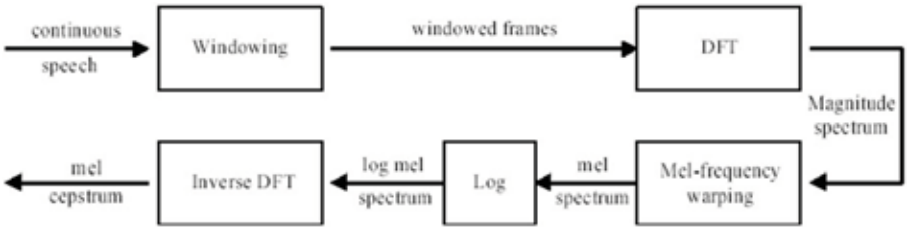

Fig. 2.1: the steps of MFCC analysis

Mfcc Process:

As described above, to place more emphasize on the low frequencies one special step before inverse DFT in calculation of cepstrum is inserted, namely Mel-scaling. A "Mel" is a unit of special measure or scale of perceived pitch of a tone. It does not correspond linearly to the normal frequency; indeed it is approximately linear below 1 kHz and logarithmic above. This approach is based on the psychophysical studies of human perception of the frequency content of sounds. One useful way to create Mel-spectrum is to use a filter bank, one filter for each desired Mel-frequency component. Every filter in this bank has triangular band pass frequency response. Such filters compute the average spectrum around each center frequency with increasing bandwidths, as displayed in Figure 2.2.

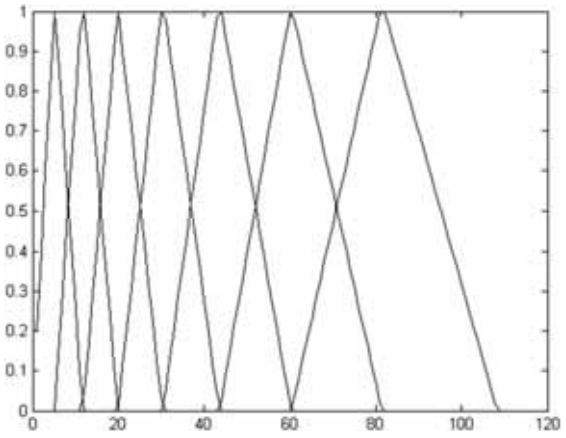

Fig. 2.2: MEL frequency filter bank

Mel Filter Bank and Multiplying in Signal:

MFCC were designed as a set of discrete cosine transform decorrelated parameters, which were computed through a transformation of the logarithmically compressed filter-output energies. These energies were derived through a perceptually spaced bank of some equal height triangular filters that are applied on the Discrete Fourier Transform (DFT)-Ed speech signal. In brief, given N-point DFT of the discrete input signal X (n).

$$X(k) = \sum_{n=0}^{N-1} x(n).e^{\left(\frac{-j2\pi nk}{N}\right)}, \quad k = 0,1,...,N-1 \tag{1}$$

A filter bank with *M* equal height triangular filters is constructed.

Each of these *M* equal height filters is defined as:

$$H_i(k) = \begin{cases} 0 & for \quad k < f_{b_{i-1}} \\ \dfrac{k - f_{b_{i-1}}}{f_{b_i} - f_{b_{i-1}}} & for \quad f_{b_{i-1}} \le k \le f_{b_i} \\ \dfrac{f_{b_{i+1}} - k}{f_{b_{i+1}} - f_{b_i}} & for \quad f_{b_i} \le k \le f_{b_{i+1}} \\ 0 & for \quad k > f_{b_{i+1}} \end{cases} \tag{2}$$

Where i stands for the i-th filter, $f\,b\,i$ are the boundary points of the filters, and k=1,2,..., N corresponds to the k -th coefficient of the N -point DFT (Todor Ganchev *et al*, 2005). The boundary points $f\,b\,i$ are expressed in terms of position, which depends on the sampling frequency $F\,s$ and the number of points N in the DFT:

$$fb_{(i)} = \left(\frac{N}{F_s}\right).\hat{f}^{-1}_{mel}\left(\hat{f}_{mel}(f_{low}) + i.\frac{\hat{f}_{mel}(f_{high}) - \hat{f}_{mel}(f_{low})}{M+1}\right) \tag{3}$$

Here, the function $f\hat{}mel(.)$ States the MEL transformation:

$$f_{mel} = 1127.\ln\left(1 + \frac{f_{lin}}{700}\right) \tag{4}$$

$f\,low$ and $f\,high$ are respectively the low and high boundary frequencies for the entire filter bank, M is the number of filters, and f^{-1}_{mel} is the inverse to MEL transformation, formulated as:

$$\hat{f}^{-1}_{mel} = 700e^{\frac{\hat{f}_{mel}}{1127}-1} \tag{5}$$

Here, and everywhere next, the sampling frequency $F\,s,$ and the frequencies $f\,low$, $f\,high$, and $f\,lin$, are in Hz, and the $f\hat{}mel$ is in mels. Equation (7) guarantees that the boundary points of the filters are uniformly spaced in the Mel scale [8].

Having the filter bank constructed, the MFCC parameters are computed, as:

$$C_j = \sum_{i=1}^{M} X_i.\cos\left(j.(i-\frac{1}{2}).\frac{\pi}{M}\right) \quad with \quad j = 1,2,...,J \tag{6}$$

Where M is the number of filters in the filter bank, J is the number of cepstral coefficients which are computed (usually $J<M$), and $X\,i$ is formulated as the "log-energy output of the i-th filter". Here, the "log-energy output of the i-th filter" is understood as:

$$X_i = \log_{10}\left(\sum_{k=0}^{N-1} |X(k)| . H_i(k)\right) \quad i = 1,2,\ldots,M \tag{7}$$

How Do the Mel Filter Bank Emphasize on Low Frequencies:

Fig.2.2 is shown a MEL frequency filter bank. In this filter bank many filters is in low frequency and a few filters is in high frequencies.

Each filter extracts one coefficient from signal by (7) so using MEL filter bank many coefficients are extracted from low frequencies and a few coefficients are extracted from high frequencies [8].

Conversion the Mfcc Filters to Low Pass Filters with Linear Function:

As described in previous section, to place emphasizes on the low frequencies one special step namely Mel filter bank is used.

In this section for more emphasis on low frequency the MEL filter bank is modified.

The filters in MEL filter bank are triangular band pass filters.

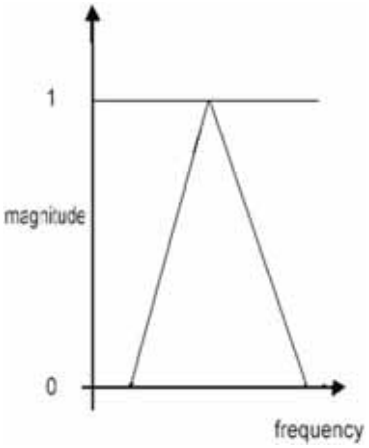

Fig.3.1: A filter in MEL filter bank

Our idea is very simple, for more emphasis on low frequency we convert the band pass filters to low pass filters.

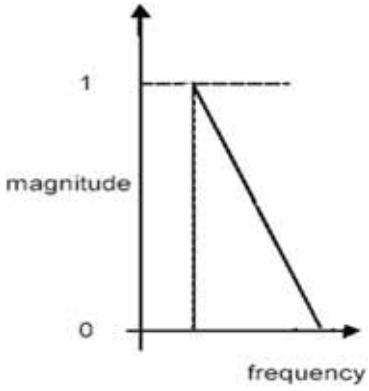

Fig.3.2: low pass filter for MEL filter bank

So the filter bank emphasizes the low frequency in two stages, once the MEL filter bank extracts many coefficients from low frequencies and a few coefficients from high frequencies, second the filters in filter bank

is low pas filters and each coefficient depends on low frequency in the each filter domain. The modified filter bank is shown in fig.3.3

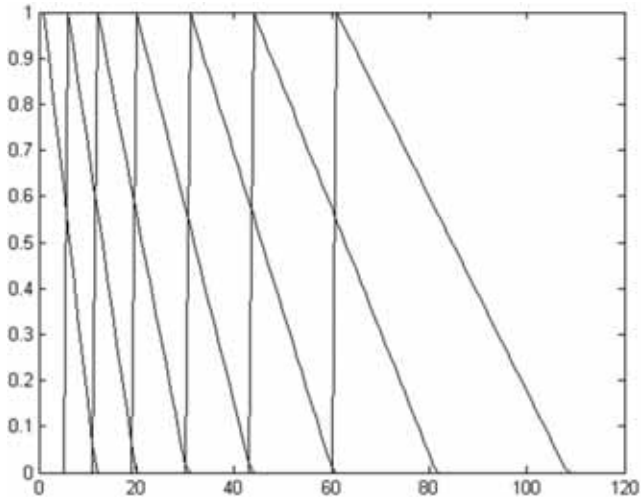

Fig. 3.3: the modified MEL filter bank with linear function

The Modified Mel Filter Bank with Linear Function:

A filter bank with *M* equal height low pass filters is constructed.

Each of these *M* equal height filters is defined as:

$$H_i(k) = \begin{cases} 0 & for \quad k < f_{\delta i-1} \\ \dfrac{f_{\delta(i+1)} - k}{f_{\delta(i+1)} - f_{\delta(i-1)}} & for \quad f_{\delta i-1} \le k \le f_{\delta i+1} \\ 0 & for \quad k < f_{\delta i+1} \end{cases} \tag{8}$$

Where *i* stands for the *i*-th filter, *f b i* are the boundary points of the filters, and *k*=1,2,..., *N* corresponds to the *k* -th coefficient of the *N* -point DFT. The boundary points *f b i* are computed from (3) and are expressed in terms of position, which depends on the sampling frequency *F s* and the number of points *N* in the DFT. Here, the function *f ˆmel (.)* is the MEL transformation formulated as (4).

f low and *f high* are respectively the low and high boundary frequencies for the entire filter bank, *M* is the number of filters, and f^{1}_{mel} is the inverse to MEL transformation, formulated as (5).

Here, and everywhere next, the sampling frequency *F s,* and The frequencies *f low*, *f high*, and *f lin*, are in Hz, and the *f ˆmel* is in MELs. Equation (7) guarantees that the boundary points of the filters are uniformly spaced in the Mel scale. Having the filter bank constructed, the MFCC parameters are computed, as (6).

Where *M* is the number of filters in the filter bank, *J* is the number of cepstral coefficients which are computed (usually *J <M)*, and *X i* is formulated as the "log-energy output of the i-th filter". Here, the "log-energy output of the i-th filter" is understood as (7).

Conversion the Mfcc Filters to Low Pass Filters with Exponential Function:

In previous section we change the filters in mel filter bank to low pass filters as shown in fig..3.3, in this section we make the low pass filter with exponential function as shown in fig.4.1

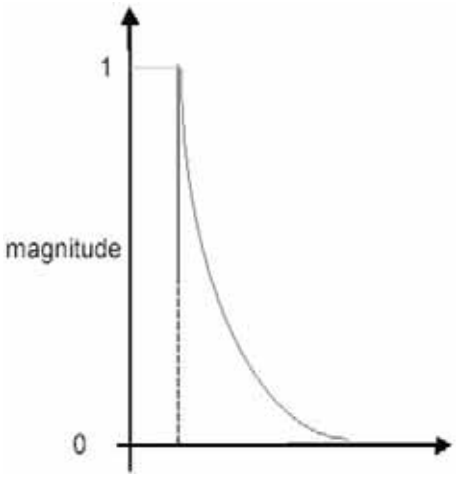

Fig. 4.1: the low pass filter with exponential function

The Modified Mel Filter Bank with Linear Function:

A filter bank with *M* equal height triangular filters is constructed.

Each of these *M* equal height filters is defined as:

$$H(k)=\begin{cases} 0 & for \quad k<f_{bi-1} \\ e^{\frac{-S}{f_{bi+1}-f_{bi-1}}(k-f_{bi-1})} & for \quad f_{bi-1}k<f_{bi+1} \\ 0 & for \quad k<f_{bi+1} \end{cases} \tag{8}$$

Where *i* stands for the *i*-th filter, *f b i* are the boundary points of the filters, and *k*=1,2,..., *N* corresponds to the *k* -th coefficient of the *N* -point DFT. The boundary points *f b i* are computed from (3) and are expressed in terms of position, which depends on the sampling frequency *F s* and the number of points *N* in the DFT. Here, the function *f ˆmel (.)* is the MEL transformation formulated as (4).

f low and *f high* are respectively the low and high boundary frequencies for the entire filter bank, *M* is the number of filters, and f^{1}_{mel} *October 11, 2010* is the inverse to MEL transformation, formulated as (5).

Here, and everywhere next, the sampling frequency *F s,* and The frequencies *f low*, *f high*, and *f lin*, are in Hz, and the *f ˆmel* is in MELs. Equation (7) guarantees that the boundary points of the filters are uniformly spaced in the Mel scale.

Having the filter bank constructed, the MFCC parameters are computed, as (6).

Where *M* is the number of filters in the filter bank, *J* is the number of cepstral coefficients which are computed (usually *J <M)*, and *X i* is formulated as the "log-energy output of the i-th filter". Here, the "log-energy output of the i-th filter" is understood as (7).

Figure 4.2 is shown the modified MEL filter bank with exponential function.

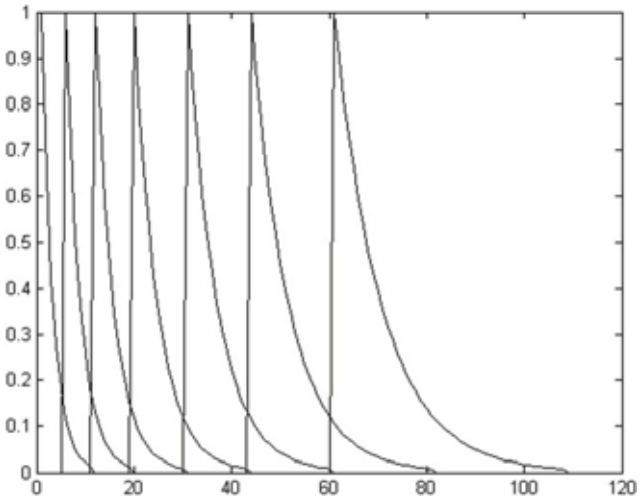

Fig. 4.2: the modified MEL filter bank with exponential function

Experimental Setup:

Data Base:

The data base in this project contained from 10 sentences from 20 various speakers from TIMIT that 7 sentences are used for training and 3 sentences are used for testing the network.

The Neural Network:

The used neural network in this project is three layers back propagation network that first layer has 7 neurons, hidden layer has 17 neurons and output layer has 20 neurons that is equal to number of speakers.This network has 20 inputs.

The used function in first layer and hidden layer is tansig that is between -1 and 1. The used function in output layer is logsig that is between 0 and 1.

Apply each frame to input of network and get a number between 0 and 1 in each rows of output of network that each number is probability of each speaker. The speaker with higher probability is winner in the frame.

Training:

First apply feature vector of each frames at input of network and the network trained using the desired output. The selected algorithm for training this network is LM.

Results:

The simulation results illustrate in following table:

Table.1: The results

Number of speakers	feature	Recognition%
20	MFCC	85%
20	MODMFCC (linear function)	90%
20	MODMFCC (exponential function)	**95%**

The recognition percent in this table is low because the simulation is text independent. In text independent, the data base in training and testing process is various.

Usually, in the text independent simulation the recognition percent is lower than text dependent simulation.

Finally, we can see in the table that the recognition percent of the modified MFCC with exponential function is better than other analysis.

Summary and Concluding Remarks:

In this paper is described MFCC that is a main analysis for speaker recognition and is explained how do the Mel filter bank emphasize on low frequencies and for more emphasis on low frequencies is used from low pass filters instead of band pass filters in Mel filter bank with the exponential and linear functions.

Finally is found the modified MFCC with exponential function is best.

REFERENCES

Richard Petersens Plads, 2002. "Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music".

Politecnico di Bari, 2008. " Frame Length Selection in Speaker Verification Task" .

Tomi Kinnunen, 2003. "Spectral Features for Automatic Text-Independent Speaker Recognition".

Adjoudj, Réda and Boukelif Aoued, 2005. "Artificial Neural Network & Mel-Frequency Cepstrum Coefficients-Based Speaker Recognition".

Muzhir, Shaban Al-Ani., Thabit, Sultan Mohammed and M. Aljebory, 2007. "Speaker Identification: A Hybrid Approach Using Neural Networks and Wavelet Transform".

Adjoudj, Réda and Boukelif Aoued, 2005. "Artificial Neural Network & Mel-Frequency Cepstrum Coefficients-Based Speaker Recognition".

Rajesh, M., Hegde Hema, 2007. "Significance of the Modified Group Delay Feature in Speech Recognition".

Todor, Ganchev., Nikos Fakotakis and George Kokkinakis, 2005. "Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task".