Gammatone Filterbank Analyses and Comparisons between GFCC and MFCC

Lai Yen Ting, Liang Kai Wen, Chang Pao Chi
National Central University, Taiwan
ytlai.vaplab@gmail.com

***Abstract*—In this paper, we propose a time domain Gammatone filtering, analysis of characteristic of filters and comparison between Mel frequency cepstral coefficient (MFCC) and Gammatone frequency coefficient cepstra (GFCC). Because the filterbank contains many filters, we also apply some methods in order to reduce the computing time of filtering. In the comparison between MFCC and GFCC, we design a simple speaker recognition test with different noise rate to compare the difference of each other. Finally we find that GFCC has better robustness in noisy condition.**

I. Introduction

Human speech is a common signals that appear in our lives. We can directly distinguish the sound difference between several people or even in noisy environment. Also we can memorize the sound from a person and identify him next time. Our hearing system can deal with complex and various tasks of sound so we would like to simulate the properties of human hearing system and apply them on machines.

Gammatone filterbank can be used to simulate the cochlea as a function of time. In our project we want to test the characteristic of filters with several basic analysis. Furthermore, we make a comparison between GFCC and MFCC. A speaker recognition test with different noise rate is conducted to compare the noise robustness. From testing result, GFCC show that they have more robustness in noisy condition.

II. Experiments

A. Time-domain Gammatone filtering

According to the work done by Qi et al. [1], time-domain Gammatone filtering gives a faster implementation than frequency-domain based methods. Fig. 1 shows the processing of time-domain Gammatone filtering. The filter z-transform takes the following form:

$$\hat{G}(z) = \frac{3a}{1 - 4mz^{-1} + 6m^2z^{-2} - 4m^3z^{-3} + m^4z^{-4}}$$

The order is 4 because previous works have shown that 4 order filter has an excellent fit to auditory filter shapes of the cochlea. The coefficient *m* in function is $e^{-2\pi b/fs}$ and *a* is set to1.

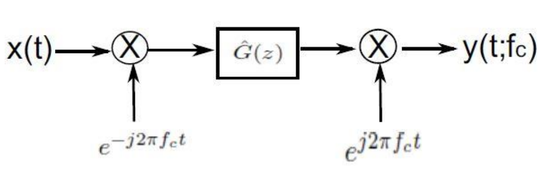

Figure. 1 Time domain Gammatone filtering.

B. Gammatone filterbank

In our project, we form different size of filterbanks to conduct different experiments. The bandwidth b is 1.019 ERB (equivalent rectangular bandwidth) and can be represented as follows:

$$b = 1.019 * 24.7 * (4.37 * f_c/1000 + 1).$$

A set of Gammatone filters with different center frequency f_c form a filterbank. Each center frequency of a filter has same ERB to its neighbor.

C. Gammatone frequency coefficient cepstra

We construct GFCC as following steps.

1. Pass the input signal through a Gammatone filterbank with 32 filters by time-domain Gammatone filtering.
2. Sample the filtered result with framing rate 100Hz and average-based framing method is applied.
3. Take the cubic root of each bin.
4. Do the discrete cosine transform (DCT), and get first 13 coefficients.
5. Compute the delta components and delta-delta components and get total 39 coefficients.

D. Mel-scale Frequency Cepstral Coefficients

We construct MFCC as following steps.

1. Pre-emphasis signal by a high-pass filter.
2. Frame the signal with framing rate 100Hz, 50% overlapping.
3. Apply hamming window to each frame.
4. Perform Fast Fourier transform (FFT) to get magnitude spectrum.
5. Wrap the magnitude spectrum into mel-spectrum using 29 triangular bandpass filters where center frequencies of the filters are equally distributed on the mel scale.
6. Do the discrete cosine transform (DCT), and get first 13 coefficients.
7. Compute the delta components and delta-delta components and get total 39 coefficients.

E. Speaker recognition test

We form a simple speaker recognition test to analyze the noise robustness between GFCC and MFCC [1][2][3]. Gaussian mixture model (GMM) is used to be the recognition model. We construct GMM for each speaker in the dataset. GFCC and MFCC are used as features of speech and model input separately. For training, we extract the features matrix with each frame features per row as the input, and set mixture number as 3. For testing, we extract the feature matrix of testing sample and calculate the power density function (pdf) per frame. Later, we sum the log probability of each frame pdf and denote the result as a correlation comparing coefficient. We calculate the correlation comparing coefficients with each GMM and choose the maximum one as the best match of testing sample.

III. RESULTS

All of our experiments are conducted on Matlab. Some of the code is open source refer to Ma et al [4].

A. *Pole-zero plot of a Gammatone filter*

Fig. 2 is the pole-zero plot of a Gammatone filter with central frequency 50Hz. The filter has a zero at 0 and 4 same poles at 0.98.

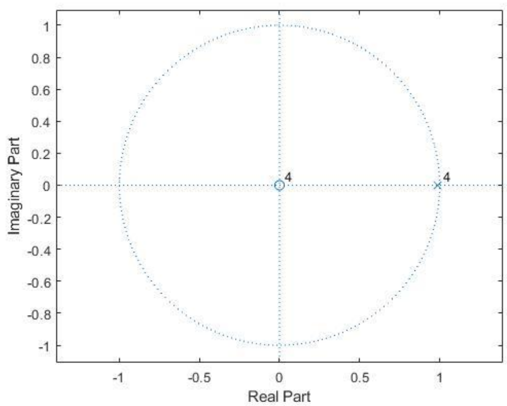

Figure. 2 Pole-zero plot of a Gammatone filter.

B. *Frequency responses of a Gammatone filterbank*

Fig. 3 is the frequency response of a Gammatone filterbank with 32 filters. The central frequencies of the filters are equally spaced between 50 Hz and 8k Hz on the ERB-rate scale.

C. *Output signal after filtering*

If a filterbank has k filters, output of it will be k band signals. The signal of each band carry the part of original signal corresponded to the related central frequency. We reconstructed the output signal by sum all of them. Fig. 4 shows the comparison between original signal and output signal.

D. *Comparison between GFCC and MFCC*

As we describe in section II, part E, a speaker recognition test is conducted to analyze the noise robustness between GFCC and MFCC. We form a dataset [5] which has 50 speakers and each has 2 short utterances, one for training, another one for testing. The white Gaussian noise are add to all utterances in testing set with different SNR per round. The testing result is shown in Fig. 4. We obvious that GFCC has more robustness in noisy condition, although the MFCC performs better than GFCC at high SNR part. We think the reason is the setting of GMMs are not the best tune for GFCC.

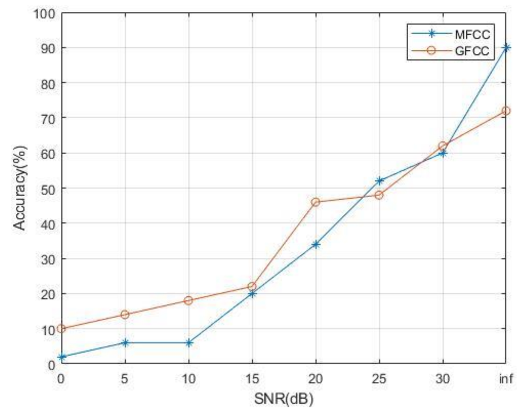

Figure. 4 Accuracy of speaker recognition test with different noise rate.

IV. REFERENCE

[1] J. Qi, D. Wang, Y. Jiang and R. Liu, "Auditory features based on Gammatone filters for robust speech recognition," *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, Beijing, 2013, pp. 305-308.

[2] Y. Shao, Z. Jin, D.L. Wang, and S. Srinivasan, “An auditory-based feature for robust speech recognition,” *in ICASSP*, 2009, pp. 4625–4628.

[3] X. Zhao and D. Wang, "Analyzing noise robustness of MFCC and GFCC features in speaker identification," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, pp. 7204-7208.

[4] Ning Ma, "An Efficient Implementation of Gammatone Filters," retrieved from: http://staffwww.dcs.shef.ac.uk/people/N.Ma/resources/gammatone/.

[5] Weinberger, S., “Speech accent archive”, George Mason University, 2013.

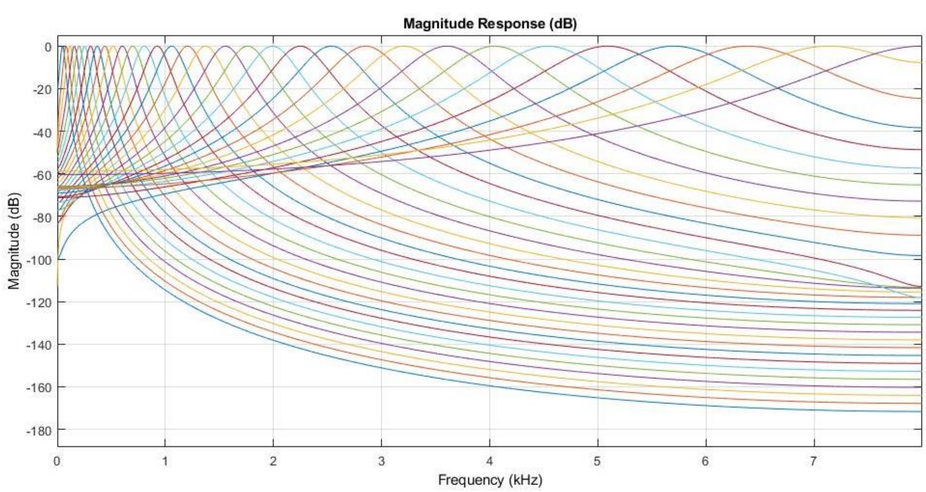

Figure. 3 Frequency response of a Gammatone filterbank with 32 filters.