Scientific Research

Identification of Noisy Utterance Speech Signal using GA-Based Optimized 2D-MFCC Method and a Bispectrum Analysis

Benyamin Kusumoputro[1], Agus Buono[2], Lina[3]

[1]Department of Electrical Engineering, Universitas Indonesia, Jakarta, Indonesia; [2]Department of Computer Science, Bogor Agricultural University, Bogor, Indonesia; [3]Department of Computer Science, Tarumanagara University, Jakarta, Indonesia..
Email: kusumo@ee.ui.ac.id, pudesha@yahoo.co.id, lina@untar.ac.id

ABSTRACT

One-dimensional Mel-Frequency Cepstrum Coefficients (1D-MFCC) in conjunction with a power spectrum analysis method is usually used as a feature extraction in a speaker identification system. However, as this one dimensional feature extraction subsystem shows low recognition rate for identifying an utterance speech signal under harsh noise conditions, we have developed a speaker identification system based on two-dimensional Bispectrum data that was theoretically more robust to the addition of Gaussian noise. As the processing sequence of ID-MFCC method could not be directly used for processing the two-dimensional Bispectrum data, in this paper we proposed a 2D-MFCC method as an extension of the 1D-MFCC method and the optimization of the 2D filter design using Genetic Algorithms. By using the 2D-MFCC method with the Bispectrum analysis method as the feature extraction technique, we then used Hidden Markov Model as the pattern classifier. In this paper, we have experimentally shows our developed methods for identifying an utterance speech signal buried with various levels of noise. Experimental result shows that the 2D-MFCC method without GA optimization has a comparable high recognition rate with that of 1D-MFCC method for utterance signal without noise addition. However, when the utterance signal is buried with Gaussian noises, the developed 2D-MFCC shows higher recognition capability, especially, when the 2D-MFCC optimized by Genetics Algorithms is utilized.

1. Introduction

Research on automatic speech and voice identification system has attracted much interest in the last few years, motivated by the growth of its applications in many areas such as in diagnosis of a rotor crack [1], classification of unknown radar targets [2], medical disease [3], and for personal and gender identification for security system [4,5]. Speaker based personal identification is the process of determining a registered speaker when an utterance speech signal is provided. In this machine-based speech identification, a gallery of speeches is firstly enrolled to the system and coded for subsequent searching. When an unidentified speech is fetched to the system, a thoroughly comparison with the each coded speech in the gallery, and the identification is then accomplished when a suitable match occurs.

Speaker identification system can be divided into two subsystems, i.e., a feature extraction subsystem and a classifier subsystem. The main function of a feature extraction subsystem is to transform the input utterance speech signal into a set of features, while a classifier subsystem have to identify and classify the speaker by comparing the extracted-features from his/her speech signal input with the ones from a set of known speakers database.

Conventional feature extraction subsystem usually used *Mel*-Frequency Cepstrum Coefficients (MFCC) and power spectrum analysis methods [6]. Power spectrum analysis method, however, shows low recognition rate for classifying the utterance speech signal under harsh noise condition [7]. To solve this problem, higher order signal analysis, i.e. bispectrum analysis method is utilized, since the bispectrum value is theoretically robust to Gaussian noise [8], which can be empirically proved by researchers such as in [7,9]. As the utterance speech in bispectrum data is represented as a pattern in 2D decision space, bispectrum analysis required two-dimensional filter design, and for that purpose, we have developed 2D-MFCC filter design method that will be explained here.

The remainder of this paper is organized as follows. In Section 2, we formulate the development of 2D-MFCC filter development. Section 3 presents the optimization of 2D-MFCC filter development by using Genetic Algorithms. Section 4 shows the experimental setup and results to demonstrate the effectiveness of the proposed method. Finally, Section 5 is dedicated to a summary of this study and suggestions for future research directions.

2. Speaker Identification System

The focus of this paper is to develop a feature extraction subsystem that could increased the recognition rate of the classifier subsystem (HMM method), to classify an utterance speech buried in a harsh noise condition. In the developed method, the feature extracting subsystem is composed of a 2D-MFCC filter design to extract the 2D information contained in the Bispectrum data. The Bispectrum data is represented as a 2D vector with MxM elements in a 2D frequency space of $f1$ and $f2$, respectively. In this section we will present a brief review of 1D-MFCC filter construction and the developed of 2D-MFCC filter construction for representing the Bispectrum data.

We developed further the feature extraction subsystem by using a Genetic Algorithm (GA) method. GA is used to optimize the filter characteristics in such that the difference between the feature vector of a speech signal without noise addition and the feature vector of a speech signal with Gaussian noise addition will be as small as possible. By reducing the difference between these two signals from the same speaker, the possibility of the speaker to be recognized correctly will be higher. As the learning method of the classifier subsystem is important aspect for increasing the recognition rate as in the soft computing methods, in this research, a Hidden Markov Model (HMM) trained by Baum Welch Algorithm is utilized [10].

In the learning phase, samples of the speaker's speech for a certain phrase of a word is inputted to the speech database, and by using these samples, the classifier system is trained to develop the reference models for those determined-speakers. In the application phase, the input utterance speech signal is compared with each of the models that has already been stored (as a reference model) on the database, and the classifier decided the winning speaker by determining the highest recognition rate for all the reference models.

2.1. Power Spectrum Analysis with 1D-MFCC Method

Suppose each tone of an utterance speech signal with an actual frequency f (Hz) is represented in Mel-frequency scale, following the relationship of:

$$\hat{f}_{mel} = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \tag{1}$$

when the actual frequency f is higher than 1000 Hz, and linear when the actual frequency is lower than 100 Hz.

The 1D-MFCC filter design method provides a triangular filter with height of 1 at its middle point, and 0 at their left and right parts for filtering the 1D Mel-frequency data. As can be seen in the Fig.2, 1D-MFCC filter can be depicted as three vertex points: $(f_{i-1},0)$, $(f_i,1)$, and $(f_{i+1},0)$ for the ith filter, with $i = 1,..., M$. It is clearly seen that determining the center point of each filter and the distance between the two adjacent center points of the filter is essential [7,9,10].

The Mel-frequency spectrum coefficients MFS_i is calculated as the sum of the filtered 1D Mel-frequency $X(j)$ that can be expressed as:

$$MFS_i = \log\left(\sum_{f=0}^{N-1} abs(X(j)) * H_i(f)\right) \tag{2}$$

where $i=1,...,M$, M the number of the triangular filter, N the number of FFT coefficients. The $abs(X(j))$ is the magnitude of jth of the FFT process of the input utterance signal, and $H_i(f)$ is the height of ith triangular at point f. The $MFCC_k$ value is then calculated by using Discrete Cosine Transform to transform the Mel-frequency spectrum coefficients back into its time domain through:

$$MFCC_k = \sum_{i=1}^{M} MFS_i * \cos\left(\frac{k * (i - 0.5) * \pi}{20}\right) \tag{3}$$

where $k=1,...,K$ the number of coefficients.

2.2. Bispectrum Analysis with 2D-MFCC Method

Bispectrum analysis of an utterance speech signal can be explained as follows. If $\{X(k)\}$, $k=0,\pm1,...,\pm2$ is a real random process, then the cumulants of order 3 is $c_3^x(\tau1,\tau2)$:

$$C_3^X(\tau1,\tau2) = \sum_{p=1}^{3}(-1)^{p-1}(p-1)!E\left(\prod_{i \in S_1} X_k\right)$$
$$* E\left(\prod_{i \in S_2} X_{k+\tau1}\right)E\left(\prod_{i \in S_3} X_{k+\tau2}\right) \tag{4}$$

where the summation extends over all partitions $(s_1,s_2,...,s_p)$, $p=1,2,3$, of the set of integers $(1,2,3)$. Bispectrum, referred to as cummulant spectra, is a Fourier transform of cummulant sequence, and formulated as:

$$C_3^x(\omega_1,\omega_2) = \sum_{\tau_1=-\infty}^{+\infty}\sum_{\tau_2=-\infty}^{+\infty} c_3^x(\tau_1,\tau_2)$$
$$* \exp\{-j(\omega_1\tau_1,\omega_2\tau_2)\} \tag{5}$$

In the case of stationary process, the cummulant order 3 can be formulated as:

$$c_3^x(\tau_1, \tau_2) = E\left\{x(t)x(t+\tau_1)x(t+\tau_2)\right\} \qquad (6)$$

Basically, there are two approaches to predict the Bispectrum, i.e. a parametric approach and a conventional approach. The conventional approach consists of the following three classes, i.e. an indirect technique, a direct technique and a complex demodulates method. Because of its simplicity, in this research, the Bispectrum data is predicted using the conventional indirect method, in which the detail of this algorithm is presented elsewhere [11].

Since Bispectrum data is represented in two dimensions of frequency $f1$ and $f2$, a 2D-MFCC filter, instead of a 1D-MFCC filter, should be used to process the data. To develop 2D-MFCC filter, we firstly construct a 1D-MFCC filter in each of $f1$ and $f2$ dimension, with in the first dimension $f1$ as $f1_i$; $i=1,....,M$ and in the second dimension $f2$ as $f2_j$; $j=1,..., N$, with $M=N$. We then combined the two separate 1D-MFCC $H_i(f1_m)$ and the other 1D-MFCC $H_j(f2_n)$ into an integrated 2D-MFCC $H_{ij}(f1_m,f2_n)$ as a pyramid shape, that can be depicted in Figure 1a. The base of this pyramid shape is a square with its corner positions are $(f1_{i-1},f2_{j-1})$, $(f1_{i+1},f2_{j-1})$, $(f1_{i-1},f2_{j+1})$ and $(f1_{i+1},f2_{j+1})$, respectively, as can be seen in Figure 1b. The connected lines between the center of the square shape and each of the corner points determined as the line a, line b, line c and line d, respectively, which its line equation can be written as:

(a)   (b)

(c)   (d)

Figure 1. The construction of 2D-MFCC filter and its calculation for Bispectrum data $B(f1_m,f2_n)$.

$$line\ a:\ f2 = \left(\frac{f2_{j-1} - f2_j}{f1_{i-1} - f1_i}\right)(f1 - f1_i) + f2_j \qquad (7)$$

$$line\ b:\ f2 = \left(\frac{f2_{j-1} - f2_j}{f1_{i+1} - f1_i}\right)(f1 - f1_i) + f2_j \qquad (8)$$

$$line\ c:\ f2 = \left(\frac{f2_{j+1} - f2_j}{f1_{i+1} - f1_i}\right)(f1 - f1_i) + f2_j \qquad (9)$$

$$line\ d:\ f2 = \left(\frac{f2_{j+1} - f2_j}{f1_{i-1} - f1_i}\right)(f1 - f1_i) + f2_j \qquad (10)$$

Using these lines, the square shape of the pyramid filter can be divided into four quadrants as can be seen in Figure 1c. Suppose we have a Bispectrum data $B(f1_m,f2_n)$ in the two dimension frequency space such as depicted at Figure 1d. The height of the 2d filter is calculated by firstly determined the quadrant of the data and then calculate the $H_{i,j}(f1_m,f2_n)$, using algorithm written below.

1. If $B(f2_n) > f2_{j-1}$, and

$$B(f2_n) < \left(\frac{f2_{j-1} - f2_j}{f1_{i-1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

$$B(f2_n) < \left(\frac{f2_{j-1} - f2_j}{f1_{i+1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

Then $B(f1_m,f2_n) \in$ quadrant I;

$$H_{i,j}(f1_m,f2_n) = \frac{B(f2_n) - f2_{j-1}}{f2_j - f2_{j-1}} \qquad (11)$$

2. If $B(f2_n) < f2_{j+1}$, and

$$B(f2_n) > \left(\frac{f2_{j+1} - f2_j}{f1_{i+1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

$$B(f2_n) > \left(\frac{f2_{j+1} - f2_j}{f1_{i-1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

Then $B(f1_m,f2_n) \in$ quadrant II;

$$H_{i,j}(f1_m,f2_n) = \frac{f2_{j+1} - B(f2_n)}{f2_{j+1} - f2_j} \qquad (12)$$

3. If $B(f1_m) > f1_{i-1}$, and

$$B(f2_n) > \left(\frac{f2_{j-1} - f2_j}{f1_{i-1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

$$B(f2_n) < \left(\frac{f2_{j+1} - f2_j}{f1_{i-1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

Then $B(f1_m,f2_n) \in$ quadrant III;

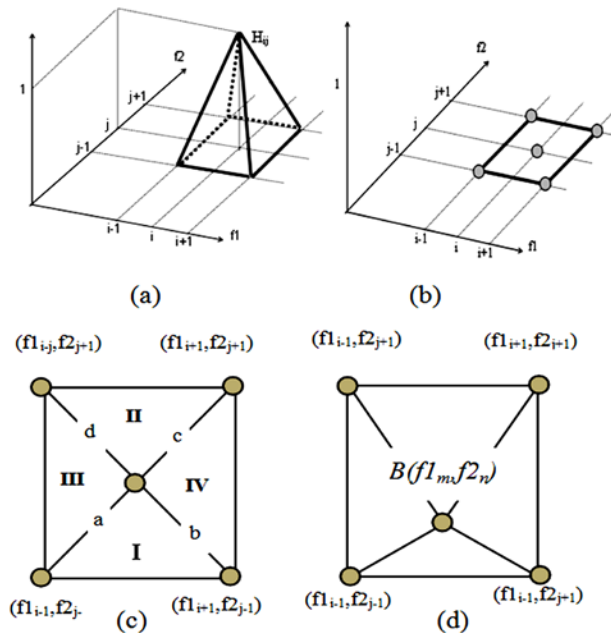$$H_{i,j}(f1_m,f2_n) = \frac{B(f1_m) - f1_{i-1}}{f1_i - f1_{i-1}} \qquad (13)$$

4. If $B(f1_m) < f1_{i+1}$, and

$$B(f2_n) > \left(\frac{f2_{j-1} - f2_j}{f1_{i+1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

$$B(f2_n) < \left(\frac{f2_{j+1} - f2_j}{f1_{i+1} - f1_i}\right)(B(f1_m) - f1_i) + f2_j$$

Then $B(f1_m, f2_n) \in$ quadrant IV;

$$H_{i,j}(f1_m, f2_n) = \frac{f1_{i+1} - B(f1_m)}{f1_{i+1} - f1_i} \tag{14}$$

Using the same calculation such as in the 1D-MFCC method (see Eq. (2)), the *Mel*-frequency Bispectrum coefficients $MFS_{(i,j)}$ is calculated through:

$$MFS(i,j) = \log\left[\sum_{m=1}^{128}\sum_{n=1}^{128} B(f1_m, f2_n) \right.$$
$$\left. * H_{i,j}(f1_m, f2_n)\right] \tag{15}$$

for the 2D-filter height $H_{i,j}(f1_m, f2_n)$, with $m=1,...,M$, $n=1,...,N$, and $M=N=128$. The $MFCC_{(i,j)}$ for 2D-MFCC is then calculated through 2D-cosine transform, as:

$$MFCC_k = \sum_i^N \sum_j^N MFS(i,j) * \cos\left(\frac{k(i-0.5)\pi}{N}\right)$$
$$* \cos\left(\frac{k(j-0.5)\pi}{N}\right) \tag{16}$$

where $k=1,2,3,...,K$ the number of the coefficient.

3. Optimization of 2D-MFCC Filter using Genetic Algorithms

The 2D-MFCC method is developed for calculating the *Mel*-Bispectrum coefficients $MFCC_k$ (in Eq. 21) of two-dimensions Bispectrum *Mel*-frequency $B(f1_m, f2_n)$, by calculating the 2D-filter height $H_{i,j}(f1_m, f2_n)$. Since the center position of each filter is very essential in determining this 2D-filter height $H_{i,j}(f1_m, f2_n)$, optimizing the position of the filter's center is necessary for reducing the total error. Thus the goal of the optimization process, i.e. Genetic Algorithms [17], is to minimize the difference between the *Mel*-Bispectrum coefficients of a speech signal buried with a Gaussian noise and that without a Gaussian noise by designing the optimized 2D-filter height $H_{i,j}(f1_m, f2_n)$.

The chromosome representation is constructed as follows. Suppose M is the maximum number of triangular filters on each frequency dimension $f1$ and $f2$, respectively, and F is the maximum frequency for both each dimensions. Suppose the distance between each of the center position of those filters, as $x_1, x_2, x_3, ..., x_{M+1}$ such that $x_1+x_2+x_3+...+x_{M+1}=F$, where x_i is the distance between i^{th} filter center with the next $(i+1)^{th}$ filter center, with $i=2,3,4,...,M$. For representing the optimized set of

these filters that will be used in the 2D-MFCC, the distance between two filters center is coded into 7 binary digits. Then the chromosome that represents a set of filters could be coded by binary digit with a length of $7*(M+1)$ digits, i.e., the first seven digits for x_1, the second seven digits for x_2, and so on. A simple illustration of the chromosome representation process is explained here. Suppose we have four triangular filters on one-dimension frequency domain, with its center position are *2.5, 4.5, 6.5* and *8.0*, respectively, with the maximum frequency F is *10*. The distance between each of filter's center are $x_1=2.5$, $x_2=4.5-2.5=2$, $x_3=6.5-4.5=2$, $x_4=8-6.5=1.5$, and $x_5=10-8=2$. The chromosome then consists of 5 locus, i.e. x_1, x_2, x_3, x_4, and x_5, in which each locus is coded by binary digit with length of 7 to be $7*5=35$ digits.

The Fitness Function is calculated so that the determined set of filters produced the *Mel*-Bispectrum coefficient $MFCC_k$, that have with very similar characteristics between the input speech signal added with Gaussian and that without a Gaussian noise addition. This fitness function can be mathematically formulated as follow:

$$fitness(i) = \frac{d(B_1, B_3) * d(B_2, B_4)}{d(B_1, B_2) * d(B_3, B_4)} \tag{17}$$

where B_1 is the Bispectrum data $B(f1_m f2_n)$ of the signal without noise addition, B_2 is the Bispectrum data $B(f1_m f2_n)$ of the signal added with 20dB Gaussian noise, B_3 is the $B_2 - B_1$, B_4 is the Bispectrum data $B(f1_m f2_n)$ of 20dB Gaussian noise, and $d(B_k, B_l)$ the distance between a feature vector of Bispectrum data B_k and a feature vector of Bispectrum data B_l.

A conventional roulette wheel is then used to select the winning chromosome in population. Chance for any chromosome to be selected is proportional to their fitness value. The Crossover technique is then used to alter two chromosomes into their offspring, and in this research, an arithmetic crossover technique is utilized. Suppose two parents $X=(x_1, x_2, x_3, ..., x_{N+1})$ and $Y=(y_1, y_2, y_3, ..., y_{N+1})$ and by using an arithmetic crossover technique, their offspring are $X=(x_1, x_2, x_3, ..., x_{N+1})$ and $Y=(y_1, y_2, y_3, ..., y_{N+1})$ with $a \in (0,1)$.

Mutation is a process of transforming any chromosome to its offspring through a changing of the internal gene. The mutation is started by selecting a certain chromosome to be mutated, followed by randomly generated two integer numbers p and q, $p,q \in [0,N+1]$ with N the number of the filter used. The mutation process is done by inverting the order of locus between those selected points.

Figure 2 shows a sample of the comparison of a 1D-MFCC filter-bank design by using the conventional method and a GA-based optimization method. It is clearly shows

that different pattern of the filter-bank are achieved, which is lead to better performance on its application.

4. Experiment Setup and Results

Several experiments were conducted to evaluate the performance of the developed system. The utterance speech signals were recorded as WAV files, conducted by ten Indonesian people, within the ages range of 12 – 28 years old. They were asks to say 'pudesha' with normaly tones and intuation, but allowed to lengthening their pronountiation. Each speaker uttered 80 times and digitized by sampling rate of 11 kHz within duration of 1.28 second, and each frame that consists of 512 samples per frame is read frame by frame with an overlaps of 256 samples between the adjacent frames. Training/testing paradigm is taken to be 50%: 50%, in which 400 utterance speeches are used as the training set, while the other 400 utterance speeches are taken as the testing data set.

The bispectrum analysis of each frame is conducted by using conventional indirect method as explained in [8]. We calculated the filtered bispectrum of each frame at frequency $B(f1_m, f2_n)$, and converted into $MFCC_k$ coefficients such as in Eq. 3. Number of coefficients K is determined to be 13, and as the consequence, the bispectrum value of each frame could be written as a feature vector that consists of 13 elements. For a balance comparison, this value is also used for the other feature extraction methods, including the conventional 1D-MFCC method.

Hidden Markov Model is used as the classifier in all of the experiments conducted here, and three different methods of feature extraction subsystems, i.e. the conventional 1D-MFCC method, 2D-MFCC method and 2D-MFCC-GA method are examined and compared. **Figure 7** shows
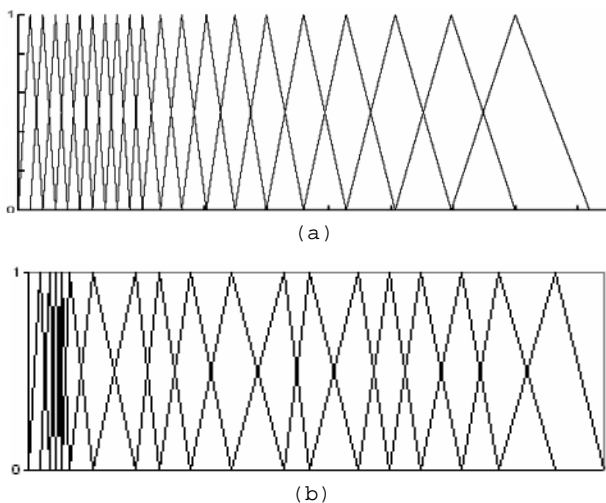

(a)

(b)

Figure 2. (a) Conventional one-dimensional MFCC filter-bank, (b) MFCC filterbank optimized by using GA.
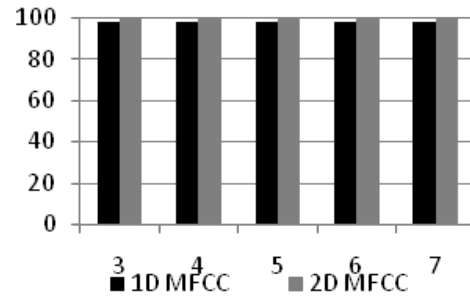

Figure 3. Comparison of recognition rate between 1D-MFCC with 2D-MFCC for a speech signal without noise addition.

a comparison of recognition rate between 1D-MFCC with 2D-MFCC for uttered speech signal without an addition of a Gaussian noise. Noted that in these experiments, we have used numerous hidden units in the HMM classifier for comparison. Experimental results depicted in **Figure 3** show that when the three different methods are used to classify an utterance speech signal without Gaussian noise addition, the recognition rates were very high, i.e. 98.4%, 99.4% and 99.0% for 1D-MFCC, 2D-MFCC and 2D-MFCC-GA, respectively. These comparable results show that the 2D-MFCC method is not necessary be used, when it is used to classify an utterance speech signal without noise addition.

This result also confirmed that the 1D-MFCC method, which was usually used in the conventional system, works appropriate enough to classify speakers when there were no noise disturbances. It is also clearly seen from this figure that the different number of hidden unit used in the HMM classifier has no influence to the recognition rate of the system. In the next experiments afterward, we have determined to use a three hidden unit HMM, for convenience.

When a Gaussian noise of 20dB is added to the utterance speech signal, however, the recognition rate of the 1D-MFCC method is dropped significantly. In order to increase the recognition rate of the systems, we have analyzed the $MFCC_k$ values for both methods as a function of each coefficient $k = 1,...,K$, for both methods; such as depicted in **Figure 4**. Clearly seen from these figures, that the first coefficient of both methods is very sensitive to the addition of the Gaussian noise, suggested that omitting this coefficient on calculating the MFCC values increases the recognition rate of both methods.

In the next experiments, we have removed the first coefficient of the MFCC methods, and by using the utterance speech signal without Gaussian noise as the inputted signal, the experimental results are depicted in **Figure 5**. As can be seen here, it is very clear that the

three feature extraction subsystems without using the first coefficient have shown a higher recognition rates, especially when the 2D-MFCC with GA is utilized.

It is confirmed that removing the first coefficient do not affected the recognition rate of the feature extraction subsystems for the utterance speech signal without Gaussian noise addition (OSS: original speech signal without Gaussian noise addition). When the utterance speech signal is buried with a 20dB Gaussian noise (OSS+20dB: original signal with 20dB Gaussian noise addition), however, the maximum recognition rates are 54.4% for the 1D-MFCC method, 70.5% for 2D-MFCC and 88.5% for 2D-MFCC-GA, respectively.
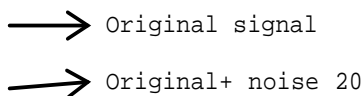
(a) 1D-MFCC

→ Original+ noise 20

↘ Original signal

(b) 2D-MFCC

→ Original signal
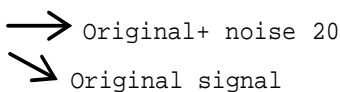
→ Original+ noise 20

Figure 4. Recognition rate comparison of the three methods by removing the 1st coefficient of the MFCC method.
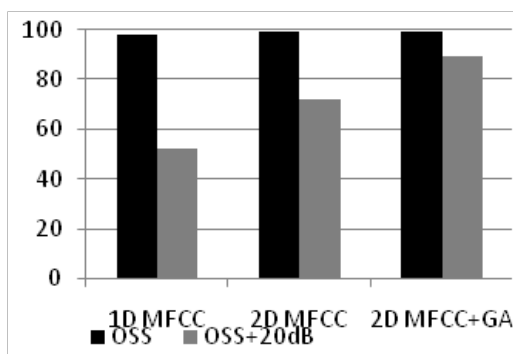
Figure 5. Comparison of the $MFCC_k$ coefficients for the original signal and its addition with Gaussian noise 20 dB.
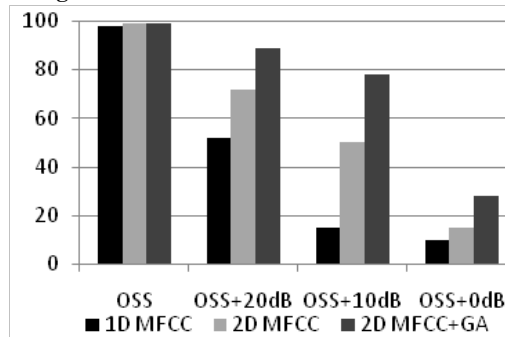
Figure 6. Recognition rate of utterance speech signal with Gaussian noise addition of 20dB using 1D-MFCC, 2D-MFCC and 2D-MFCC-GA methods with K=12.

The next experiment was conducted by buried the utterance speech signal in more harsh noise conditions, i.e. 10 dB and 0 dB, respectively. A complete comparison of the recognition rate for the 2D-MFCC and the 2D-MFCC-GA using an utterance speech signal with an addition of Gaussian noise of 20 dB, 10 dB and 0 dB, respectively, is depicted in **Figure 6**. As shown in this figure, when the Gaussian noise intensity is increasing, the recognition rate is decreased accordingly. It can also be seen that, for all of the Gaussian noise intensity level, the use of GA for optimization of 2D-MFCC for Bispectrum signal as this feature extraction subsystem always performs better than that of 2D-MFCC without GA.

5. Conclusions

We have developed the 2D-MFCC feature extraction method for processing the Bispectrum data from utterance speech signal. In this paper, we have developed an optimization of the filter design through GA method for increasing the recognition capability of the system, especially for the uttered speech signal under addition Gaussian noise. It is shown that the recognition rate of the systems by using 2D-MFCC, with or without GA optimization is comparable with that of 1D-MFCC for uttered speech signal under normal conditions. However, these recognition rates are decreased significantly when a Gaussian noise is added to the uttered speech signal. Further analysis shows that the 1st-coefficient of both 2D-MFCC and 1D-MFCC are largely influence by the addition of the Gaussian noise, and by eliminating this coefficient, the performance of the 2D-MFCC is greatly change to show higher recognition rate, i.e. 70.5% and 88.5% for 2D-MFCC without GA and 2D-MFCC with GA, respectively, compare with 59.4% for 1D-MFCC method. Further analysis of these coefficients for the system performance is now under investigation in order

to develop more robust speaker identification system, especially under harsh noise environments.

6. Acknowledgement

REFERENCES

[1] Z. Li, J. Sun, J. Han, f. Chu and Y. He, Parametric bispectrum analysis of cracked rotor based on blind identification of time series models, *IEEE Proceeding of Intelligent Control and Automation*, Vol. 2, 2006, pp.5729-5733.

[2] I. Jouny, E.D. Garber and R.L. Moses, Radar target identification using the bispectrum: a comparative study, *IEEE Trans. Aerospace and Electronic Systems*, Vol. 31, No. 1, 1995, pp. 69-77.

[3] E.S. Fonseca, R.C. Guido, A.C. Silvestre and J.C. Pereira, Discrete wavelet transform and support vector machine applied to pathological voice signals identification, *IEEE Proceeding of International Symposium on Multimedia*, 2005

[4] Z. Wang and H. Wang, Voice identification system based on server, *IEEE Proceeding of Intern. Conf. on Computer Application and System Modeling*, Vol. 9, 2010, pp. 384-387.

[5] M. Abdollahi, E. Valavi and H.A. Noubari, Voice-based gender identification via multiresolution frame classification of spectro-temporal maps, *IEEE Proceeding of Intern. Joint Conf. on Neural Networks*, 2009, pp. 1-4.

[6] T.D. Ganchev, *Speaker Recognition*, PhD Dissertation, Wire Communications Laboratory, Department of Computer and Electrical Engineering, University of Patras Greece, 2005

[7] B. Kusumoputro, A. Triyanto, M.I. Fanany and W. Jatmiko, Speaker identification in noisy environment using bispectrum analysis and probabilistic neural networks, *IEEE Proceeding of Intern. Conf. on Computational Intelligence and Multimedia Application*, 2001, pp.118-123.

[8] C.L. Nikeas and A.P. Petropulu, *Higher order spectra analysis: A Nonlinear Signal Processing Framework*, Prentice-Hall, Inc. New Jersey, 1993.

[9] T.E. Ozkurt and T. Akgul, Robust text-independent speaker identification using bispectrum slice, *IEEE Proceeding of Signal Processing and Communications Applications*, 2004, pp. 418-421.

[10] L. Luo and L.F. Chaparro, Parametric identification of systems using a frequency slice of the bispectrum, *IEEE Proceeding of Intern. Conf. on Acoustic, Speech and Signal Processing*, Vol. 3, 1991, pp. 3481-3484

[11] L. Rabiner. A Tutorial on Hidden Markov Model and Selected Applications in Speech Recognition. *Proceeding IEEE*, Vol 77 No. 2. February 1989.

[12] Cornaz, C. dan U. Hunkeler. An Automatic Speaker Recognition System. Mini-Project. *http://www.ifp.uiuc.edu/~minhdo/teaching/speaker_recognition*, access : August, 15, 2008.

[13] Zbigniew M. *Genetic Algorithms + Data structures = Evolution Programs*, 3th Edition, Springer, 1996.