International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:4, No:8, 2010

Comparison of MFCC and Cepstral Coefficients as a Feature Set for PCG Biometric Systems

Justin Leo Cheang Loong, Khazaimatol S Subari, Muhammad Kamil Abdullah, Nurul Nadia Ahmad and Rosli Besar

Abstract—Heart sound is an acoustic signal and many techniques used nowadays for human recognition tasks borrow speech recognition techniques. One popular choice for feature extraction of accoustic signals is the Mel Frequency Cepstral Coefficients (MFCC) which maps the signal onto a non-linear Mel-Scale that mimics the human hearing. However the Mel-Scale is almost linear in the frequency region of heart sounds and thus should produce similar results with the standard cepstral coefficients (CC). In this paper, MFCC is investigated to see if it produces superior results for PCG based human identification system compared to CC. Results show that the MFCC system is still superior to CC despite linear filter-banks in the lower frequency range, giving up to 95% correct recognition rate for MFCC and 90% for CC. Further experiments show that the high recognition rate is due to the implementation of filter-banks and not from Mel-Scaling.

Keywords—Biometric, Phonocardiogram, Cepstral Coefficients, Mel Frequency

I. Introduction

A HUMAN identification system is a system that is able to recognize an individual when certain data which is specific to the individual is presented to it. In the past, such systems were based on handwriting, speech, fingerprints and facial features to perform its tasks. However too much reliance cannot be placed onto these biometrics as they can be forged by others. As such researchers have looked deeper into the human body to search for alternatives which cannot be easily falsified.

The human heart sound has been traditionally used as a means of identifying diseases through its analysis. Phonocardiogram (PCG) is the digitally recorded heart sound. Recently PCG has been explored to see if it carries information specific to individuals. Studies have shown that heart sound possesses the capability of being a biometric [1], [2], [9].

The heart sound is categorised by the two loudest sounds referred to as S1 and S2. S1 typically lasts for a duration of 150 ms with a frequency between 25 to 45 Hz and S2 lasts for 120 ms bearing a frequency of about 50 Hz. S1 is produced through the sudden closure of the mitral and tricuspid valves during isovolumetric contraction to pump blood into the aorta and pulmonary artery. The sound of S1 is akin to a low and slightly prolonged “lub”. S2, which is a short, high-pitched “dup”, is then caused by closure of the aortic and pulmonary valves during isovolumetric relaxation when the ventricles end ejection and starts the diastole. Together they make up the “lup-dup” sound that is often thought of when referring to a heart beat. An example of a PCG signal containing S1 and S2 is shown in Figure 1.

Justin Leo Cheang Loong, Khazaimatol S Subari, Muhammad Kamil Abdullah and Nurul Nadia Ahmad are with the Faculty of Engineering, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Selangor, Malaysia.
Rosli Besar is with the Faculty of Engineering and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia.

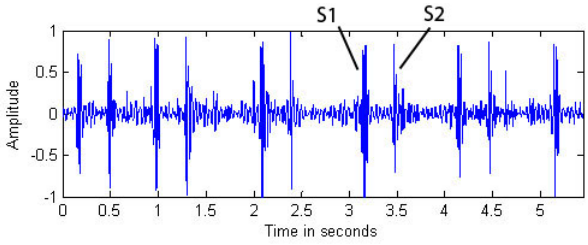

Fig. 1. Example of a PCG waveform where S1 and S2 are clearly visible.

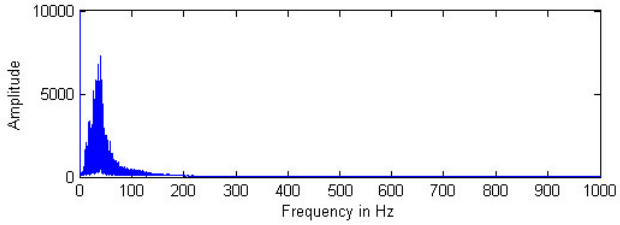

Fig. 2. Spectrum of a PCG signal. The frequency range of the signal is concentrated in the low frequency region.

The idea of using Mel Frequency Cepstral Coefficients (MFCC) as the feature set for a PCG biometric system comes from the success of MFCC for speaker identification [5] and because PCG and speech are both accoustic signals. MFCC differentiates itself from the standard cepstral coefficients (referred to as CC from here on) as it maps the spectrum of the signal onto the Mel-Scale which replicates the human hearing perception. However, in the low frequency range of up to 1000 Hz, the Mel-Scale is linear , therefore heuristically MFCC should be identical to CC. This is shown in Figure 2 where the energy of the spectrum is concentrated in the low frequency range. MFCC is researched to see if it allows for improved system performance compared to CC for PCG signals identification using a Gaussian Mixture Model (GMM) classification algorithm.

II. Cepstral Coefficients and Mel Frequency Cepstral Coefficients

The cepstrum is defined as the inverse Fourier transform of the log-magnitude Fourier spectrum. It is used to separate

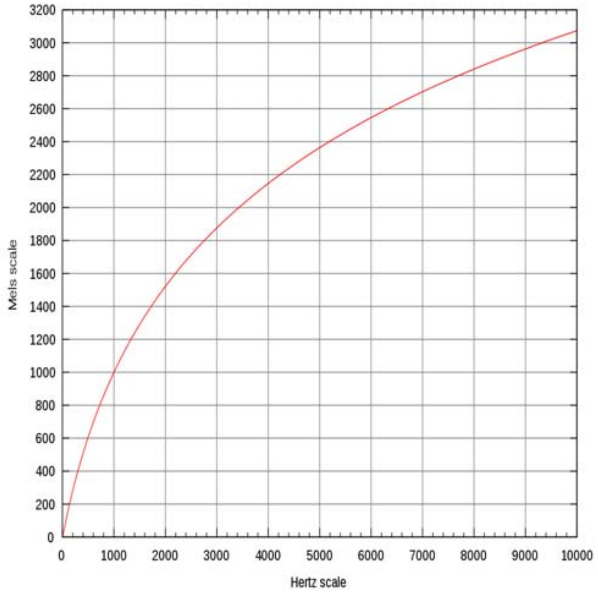

Fig. 3. The Mel scale versus the Hertz scale.

the transfer function and the excitation signal which exists in the low quefrency and high quefrency respectively. The coefficients that make up the resulting cepstrum are known as the cepstral coefficients. In the past, CC has only been used for identifying echoes that are present in an accoustic signal [3] but later on CC have been shown to be a feasable set of features for speaker identification [7], [8] and even musical instrument identification [4].

Stevens et al. (1937) proposed the Mel scale which is a scale of pitches judged to be equal in distance from one another according to human perception [11]. According to the scale, larger and larger intervals of frequency are judged by listeners to produce equal pitch increments. For example the perceptual interval between 100 Hz to 200 Hz is approximately the same as the perceptual interval between 10 kHz to 20 kHz. Figure 3 shows that as you go up the Hertz scale, the same interval on the Mel scale will require increasingly larger Hertz intervals.

Mel-frequency cepstrum is actually a cepstrum with its spectrum mapped onto the Mel-Scale before the log and inverse fourier transform is taken. As such, the scaling in Mel-frequency cepstrum mimics the human perception of distance in frequency and its coefficients are known as the MFCC. MFCCs are now widely used for speaker recognition tasks [5] and has been shown to yield excellent results [6], [10]. In [6], it is also shown that MFCC outperforms normal cepstral coefficients for speaker identification.

A similar study by Phua et al. (2008) has shown successful implementation of MFCC of PCG sound as a biometric. However no comparison has been done to determine if there is an increase in performance when using MFCC against normal cepstral coefficients as in speaker recognition.

III. Methodology

This experiment consists of four distinct parts: i) experimental setup, ii) preprocessing, iii) feature extraction and iv) classification. Each part will be explained in detail in the following subsections.

A. Experimental Setup

For the purpose of this experiment, a database of PCG recordings were created. The signals were recorded using a Dong Jin Medical i-Scope 200 digital stethoscope with a sampling frequency of 11025 Hz at 16-bits per sample. The stethoscope was connected to a computer and the PCG was captured using the Matlab Data Aqcuisition Toolbox. Participants were required to sit on a reclining chair and remain calm and relaxed throughout the whole procedure. The stethoscope was placed on the pulmonary auscultation site on the participants' chest.

There were a total of 6 participants and they were required to attend 6 separate recording sessions. Each session consisted of 10 trials with each trial lasting for approximately 60 seconds. The sessions were spaced at least one day apart from each other. Thus there was a total of 60 PCG recordings for each participant. The first and last 5 second of each recordings were discarded therefore 50 seconds of the signal was obtained from each trial. The signals were randomly divided into training and testing sets each time the system is executed whereby 50% of the signals were used for training and 50% for testing.

B. Preprocessing

Preliminary processing is done in order to prepare the signal for the feature extraction stage. The signal may be affected by the noise caused by internal organs, body or hand movements and also bursty interferences. The following steps were done during preprocessing:

1) Low-pass filter: An elliptic low-pass filter with a cut-off frequency of 300 Hz is used to remove unwanted high frequency noise.
2) Spike removal: Values that are higher than a certain threshold are set to the threshold value to minimize the effects of bursty interference.
3) Amplitude normalization: All the signals are normalized to a range between -1 and +1 using the following equation:

$$x_n[n] = \frac{x_i[n] - \mu_x}{\max(|x|)} \quad (1)$$

where $x_i[n]$ is the input signal, μ_x is the signal mean, $\max(x)$ is the maximum amplitude of the signal and $x_n[n]$ is the normalized signal.

C. Feature Extraction

For feature extraction, the signal will be put through the CC and MFCC method so that the performance of the system using these two feature sets can be compared.

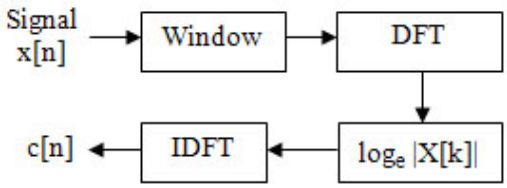

Fig. 4. Cepstral coefficient feature extraction process.

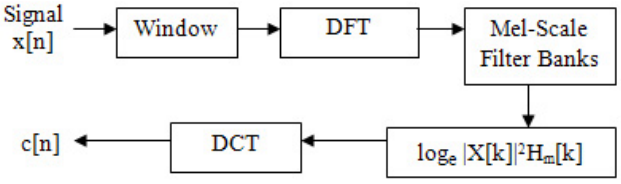

Fig. 5. MFCC feature extraction process.

1) Cepstral Coefficients: The process of extracting the cepstral coefficients are shown in Figure 4. First the signal is divided into frames of 512 ms and each frame is multiplied with a Hamming window. Then the signal is converted from the time domain into the frequency domain by using the discrete Fourier transform (DFT).

As the PCG signal has a very small bandwidth in the lower frequency regions, only the spectrum ranging from 0 Hz to 100 Hz is utilized to maximize the information contained within the cepstral coefficients for better seperability and also to cut back on computational costs.

Following this, the natural log of the magnitude of the spectrum is computed and finally, the inverse discrete Fourier transform (IDFT) is computed. Only the first few cepstral coefficients are selected as the higher order coefficients represents the excitation process which is less useful [9]. A normal cepstrum operates on a linear frequency scale while Mel-frequency cepstrum operates on the Mel-Scale which will be discussed in the following section.

In general, the equation for determining the cepstrum of a signal can be written as follows:

$$c[n] = \mathrm{IDFT}(\log(|\mathrm{DFT}(x[n].w[n])|)) \tag{2}$$

where $x[n]$ is the signal, $w[n]$ is the window function, n is the frequency index and $c[n]$ is the computed cepstrum. The first 50 coefficients are used for this experiment.

2) Mel Frequency Cepstral Coefficients: The main difference between computation of the MFCC and the cepstral coefficients is the inclusion of Mel-Scale filter-banks, as shown in Figure 5. The Mel-Scale filter-banks are computed as follows:

$$m = 1127 \log_e \left(\frac{f}{700} + 1 \right) \tag{3}$$

where f is the frequency in the linear scale and m is the resulting frequency in Mel-Scale. The frequency scale ranging from 0 Hz to 100 Hz is converted to the Mel-Scale and 51 centre frequencies are then spaced linearly throughout the range in Mel-Scale. These centre frequencies are then converted back to the normal linear scale using the inverse of Equation (3) and they will now be spaced logarithmically. Triangular overlapping filter-banks are then constructed based on these centre frequencies.

The power spectral density (PSD) of the spectrum is mapped onto the Mel-Scale by multipying it with the filter-banks constructed earlier and the log of the energy output of each filter is calculated as follows:

$$S[m] = \log \left(\sum_{k=0}^{N-1} |X[k]|^2 H_m[k] \right) \tag{4}$$

where $H_m[k]$ is the filter-banks and m is the number of the filter-bank. Finally, the discrete cosine transform (DCT) of the spectrum to obtain the MFCC is computed:

$$c[n] = \sum_{m=0}^{M-1} S[m] \cos \left(\frac{\pi n}{M} \left(m - \frac{1}{2} \right) \right), n = 0, 1, 2, \ldots, M \tag{5}$$

where M is the total number of filter banks. The first coefficient is discarded, and the remaining is used for testing and training.

D. Classification

The classification method used for this comparison is the GMM. It is actually a probabilistic model with a normal (Gaussian) distribution. Every class will have its own GMM. During the testing phase, the signal is compared to the available GMMs and classification is made according to the GMM which gives the maximum likelihood estimation. Studies have shown that GMM not only make an excellent classifier for speech recognition [10] but also for PCG signal recognition as well [9]. It has also been shown that GMM outperforms Vector Quantization (VQ) for PCG biometric systems [9] and thus GMM is selected as the classifier in this biometric system.

For a D-dimensional feature vector denoted as x, the GMM of a person's heart sound is given by the weighted sum of M component densities:

$$p(x|\lambda) = \sum_{i=1}^{M} p_i b_i(x) \tag{6}$$

Each component density, $b_i(x)$ is a uni-modal Gaussian density function given by:

$$b_i(x) = \frac{1}{\sqrt{(2\pi)^D} \sqrt{|\sum_i|}} e^{\left(-\frac{1}{2}(x-\mu_i)^T \sum_i^{-1} (x-\mu_i)\right)} \tag{7}$$

where μ_i is the mean vector, $\sum_i$ is the covariance matrix and the mixture weight satisfies the constraint $\sum_{i-1}^{m} p_i^s = 1$. Therefore the GMM of a person's heart sound feature vector is denoted as $\lambda = \{p_i, \mu_i, \sum_i\}$, $i = 1, \ldots, M$. The Expectation-Maximization (EM) algorithm is used for estimating the maximum likelihood model parameters.

For recognition tasks, the cepstral coefficients and the MFCCs will be put through the different GMMs and classified according to the maximum likelihood.

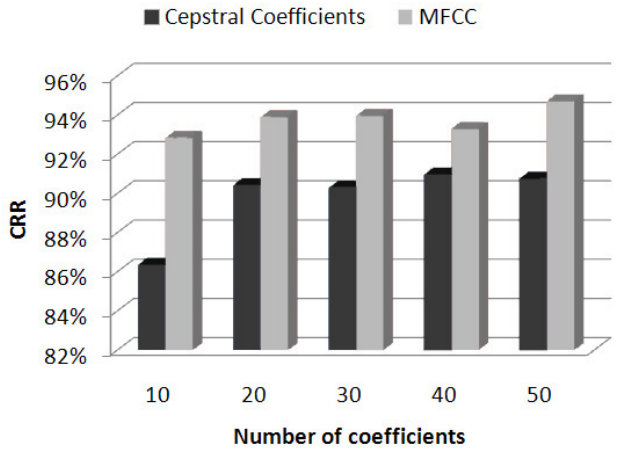

Fig. 6. CRR of the system using different number of cepstral coefficients and MFCCs. The number of GMM components was set to 13.

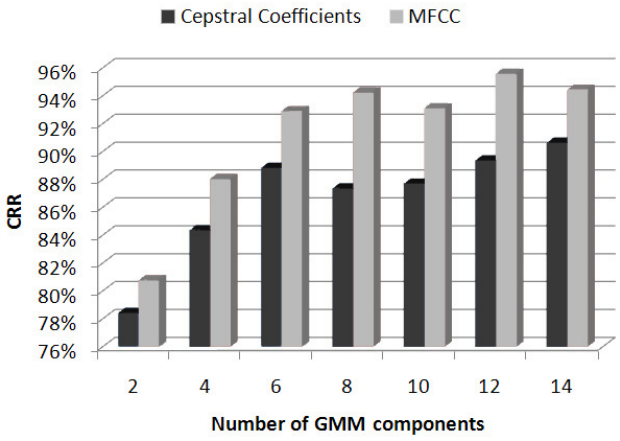

Fig. 7. CRR of the system using cepstral coefficients and MFCC for different values of GMM components. The number of coefficients was set to 50.

IV. Results and Discussion

The results of this experiment will be based on the correct recognition rate (CRR):

$$CRR = \frac{C_n}{T_n} \times 100 \tag{8}$$

where C_n is the number of correct classifications and T_n is the total number of testing samples. Since the samples used for training and testing for every run are randomized, the CRR for every test is taken as an average of 10 runs.

Preliminary experiments were conducted in order to find the i) optimum number of coefficients and ii) the optimum number of mixture components needed for high classification rates. In the first experiment, the number of mixture components, M, was set to 13 and the number of cepstral coefficients and MFCCs were varied. Figure 6 shows that the performance of both systems remain relatively stagnant from 20 coefficients onwards. It is interesting to note is that despite the fact that the MFCC is mostly linear below 100 Hz, it shows a constant performance gain over the cepstral coefficient counterpart.

In the second experiment, the amount of mixture components used to model the GMM classifier was varied. It was observed in Figure 7 that as the number of mixture components increase, both systems show improved performance. Nonetheless the results follow the trend from the previous experiment whereby the MFCCs show better performance gain over cepstral coefficients.

These results are rather surprising, considering that Mel-Scale is almost linear within the bandwdth where heart sound resides and therefore the results should not differ much from that of the cepstral coefficients. Another test is done where expand the bandwidth for obtaining the coefficients is expanded to the full frequency range of up to 5512.5 Hz. Additionally another system is created for extraction of cepstral coefficients using filter-banks that are separated linearly henceforth referred to as linear filter cepstral coefficients (LFCC).

Table I shows that there is almost no difference in the CRR between the cepstral coefficient system and the LFCC system when the full frequency range is used. However a large increase in CRR of about 10% can be seen when using the MFCC. This can be attributed to the property of the Mel-Scaled filter-banks having denser filter-banks in the lower frequency regions, thus extracting more useful information compared with the previous systems. As can be seen in Figure 8, LFCC only has 2 filter-banks below 100 Hz while MFCC has 3 when 21 filter-banks are implemented over a range of 0 to 1000 Hz.

TABLE I
CRR using full bandwidth of signal (0 Hz to 5512.5 Hz).

Type of coefficients	CRR (%)
Cepstral Coefficients	61.67
LFCC	58.93
MFCC	74.53

Since LFCC has linear scaling, when the specified bandwidth of 0 to 100 Hz is used, it should produce the same results as CC. However Table II show that although LFCC is linear in the specified bandwidth, the performance is still superior to CC. Based on these results, it can infered that when the specified frequency range is used, Mel-Scaling of MFCC does not provide any substantial improvements for heart sounds and using a linear scale gives similar results. The gains in performance for LFCC over CC is speculated to be the result of the implementation of filter-banks in the MFCC algorithm in the specified bandwidth.

TABLE II
CRR for different number of filter-banks for MFCC and LCC up to 100 Hz of signal.

Type of coefficients	CRR (%)
MFCC (51 filter-banks)	95.53
MFCC (11 filter-banks)	91.93
LFCC (51 filter-banks)	95.20

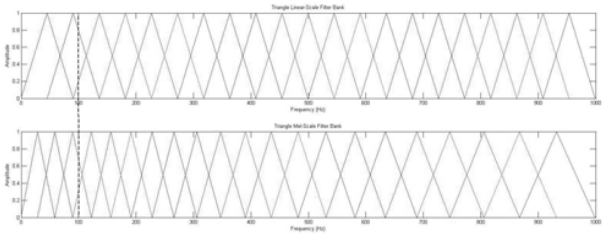

Fig. 8. Linear and Mel-Scaled filter-banks from 0 to 1000 Hz.

V. Conclusion

In this paper, the performance of two widely used feature extraction techniques for speaker identification in a PCG biometric system is compared. In the MFCC system, the optimum configuration gives a CRR of 95% while in the cepstral coefficient system, the CRR peaks at 90%. Overall, MFCC makes a better feature set as compared to cepstral coefficients. Experimental results show that Mel-Scaling does not provide any substantial benefit when the feature extraction is restricted to the low frequency region ($\leq$100 Hz). The performance difference in performance is speculated to be the result of the implementation of filter-banks in the MFCC algorithm and not because of the Mel-Scaling.

As such, future works can be done to incorporate filter-banks in other feature extraction algorithms in the low frequency region to determine if the filter-banks improve on those algorithms as well.

Acknowledgment

The authors would like to thank the Ministry of Science, Technology and Innovation, Malaysia (MOSTI) for funding the research.

References

[1] F. Beritelli. A multiband approach to human identity verification based on phonocardiogram signal analysis. In *Biometrics Symposium*, pages 71–76, Tampa, FL, 2008.

[2] F. Beritelli and S. Serrano. Biometric identification based on frequency analysis of cardiac sounds. *IEEE Transactions on Information Forensics and Security*, 2(3):596–604, 2007.

[3] B. P. Bogert, M. J. R. Healy, and J. W. Tukey. *The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking*, chapter 15, pages 209–243. Wiley, NY, 1963.

[4] J. C. Brown. Computer identification of musical instruments using pattern recognition with cepstral coefficients as features. *The Journal of the Acoustical Society of America*, 105(3):1933–1941, 1999.

[5] T. Ganchev, N. Fakotakis, and G. Kokkinakis. Comparative evaluation of various MFCC implementations on the speaker verification task. In *Proceedings of the Speech and Computer*, pages 191–194, Patras, Greece, 2005.

[6] M. R. Hasan, M. Jamil, M. G. Rabbani, and M. S. Rahman. Speaker identification using mel frequency cepstral coefficients. In *Proceedings of the 3rd International Conference on Electrical and Computer Engineering*, pages 565 – 568, Dhaka, Bangladesh, 2004.

[7] A. Kinney and J. Stevens. Wavelet packet cepstral analysis for speaker recognition. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 206–209, Monterey, CA, 2002.

[8] M. Nazar. Speaker identification using cepstral analysis. In *Proceedings of the IEEE Students Conference*, volume 1, pages 139–143 vol.1, Lahore, Pakistan, 2002.

[9] K. Phua, J. Chen, T. H. Dat, and L. Shue. Heart sound as a biometric. *Pattern Recognition*, 41(3):906 – 919, 2008.

[10] D. A. Reynolds. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication*, 17(1-2):91 – 108, 1995.

[11] J. Volkmann, S. S. Stevens, and E. B. Newman. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3):208–208, 1937.

Justin Leo Cheang Loong received his B.Eng (Hons) majoring in Computer from Multimedia University in 2009. He is currently working as a research officer in the same university while pursuing his M.Sc. His research interests are biomedical signal processing and biometrics systems.

Khazaimatol S Subari received her B.Eng (Hons) and M.Sc from Vanderbilt University in 1999 and 2001 respectively. She obtained her Ph.D from the University of Southampton in 2006. She is currently a lecturer and researcher at the Faculty of Engineering in Multimedia University, Cyberjaya. Her research interest is in biomedical signal processing and biometrics systems.

Muhammad Kamil Abdullah received B.Eng (Hons) majoring in Multimedia at Multimedia University in 2009. Currently working as a research officer in Centre for Multimedia Security and Signal Processing in Faculty of Engineering, Multimedia University. His research interest is biomedical and audio signal processing. He has been pursuing his M.Eng.Sc since 2009 in the same university.

Nurul Nadia Ahmad received the Diploma in Electronics Engineering from the Institute of Telecommunications and Information Technology (ITTM), B.Eng (Hons) in Electronics and Communications Engineering from the University of Bristol and the Ph.D degree from the University of Southampton. Her current research interests are Wireless communications and Image Processing.

Rosli Besar received the B.Eng (Hons) and M.Sc degrees from the University of Science Malaysia (USM), Malaysia, in 1990 and 1993, respectively and the Ph.D degree from the Multimedia University, Malaysia, in 2004. His current interests include Signal and Image Processing, Medical Imaging, Digital Signal Processing, and Telemedicine.