SIViP
DOI 10.1007/s11760-015-0826-9

ORIGINAL PAPER

A comparative approach between cepstral features for human authentication using heart sounds

M. Abo-Zahhad[1] · Mohammed Farrag[1] · Sherif N. Abbas[1] · Sabah M. Ahmed[1]

Received: 26 May 2015 / Revised: 7 September 2015 / Accepted: 29 September 2015

Abstract The main objective of this paper is to provide a comparative study between different cepstral features for the application of human recognition using heart sounds. In the past 10 years, heart sound, which is known as phonocardiogram, has been adopted for human biometric authentication tasks. Most of the previously proposed systems have adopted mel-frequency and linear frequency cepstral coefficients as features for heart sounds. In this paper, two more cepstral features are proposed. The first one is based on wavelet packet decomposition where a new filter bank structure is designed to select the appropriate bases for extracting discriminant features from heart sounds. The other is based on nonlinear modification for mel-scaled cepstral features. The four cepstral features are tested and compared on two databases: One consists of 21 subjects, and the other consists of 206 subjects. Based on the achieved results over the two databases, the two proposed cepstral features achieved higher correct recognition rates and lower error rates in identification and verification modes, respectively.

Keywords Heart sounds · PCG biometric authentication · Wavelet denoising · Cepstral features · Wavelet packet decomposition · Linear discriminant analysis

✉ Sherif N. Abbas
sherif.siha13@eng.au.edu.eg

M. Abo-Zahhad
zahhad@yahoo.com

Mohammed Farrag
mohammed.farrag@eng.au.edu.eg

Sabah M. Ahmed
sabahma@yahoo.com

[1] Department of Electrical and Electronics Engineering, Faculty of Engineering, Assiut University, Assiut, Egypt

1 Introduction

Due to the increasing need for security, research studies are seeking alternative biometric traits that can complement current biometrics such as fingerprint and facial characteristics. Heart sound signal, which is known as phonocardiogram (PCG), is considered one of the emerging techniques for human authentication applications because it has some advantages over traditional biometric traits. A study conducted by Matsumoto et al. [1] showed that a fingerprint system can be easily breached using fake fingerprints. A true acceptance rate ranging from 65 up to 100 % was achieved using gummy fingers tested over 11 different commercial fingerprint systems. However, PCG signals are less likely to be artificially generated and fed to a sensor to spoof it. Also, unlike face recognition technique, PCG signals are not exposed and therefore can not be captured at a distance [2]. Furthermore, they are easy to process because the heart sound is a one-dimensional signal with low frequency range and have high universality because every person has a beating heart. In the next paragraph, the origin of the PCG signal is briefly discussed.

Phonocardiography is the process of recording the cardiac acoustics vibration by means of a digital stethoscope [3]. The human heart is a four-chamber pump as shown in Fig. 1 [4]. Two sets of valves control the flow of blood: the atrioventricular valves which open to let the blood flow into the heart (diastole–relaxation) and the semilunar valves which open to let the blood flow out of the heart (systole–contraction). These mechanical processes produce vibrations and acoustic signals that can be recorded over the chest wall. The heart sounds consist mainly of two sounds denoted by $S1$ and $S2$, which are detected at the beginning of the systolic and diastolic cycles, respectively, as illustrated in Fig. 2 [5]. The first sound, $S1$, is caused by the closure of the atrioventricular

Published online: 24 October 2015

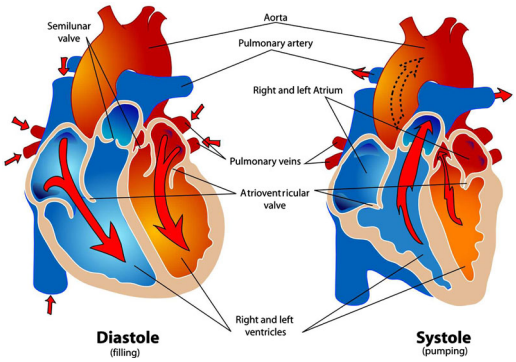

Fig. 1 The anatomy of the human heart

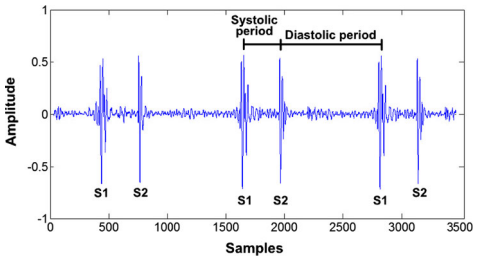

Fig. 2 Typical waveform of $S1$ and $S2$ sounds

valves, while the second sound, $S2$, is caused by the closure of the semilunar valves. $S1$ lasts for an average period of 100–200 ms, and its frequency components lie in the range of 25–45 Hz. $S2$ lasts about 120 ms, with a frequency of 50 Hz which is typically higher in frequency than $S1$.

In this paper, a comparative study between different types of cepstral features is presented, where two new cepstral features are proposed and compared with two well-known cepstral features: mel-frequency cepstral coefficient (MFCC) features and linear frequency cepstral coefficient (LFCC) features. The first proposed cepstral features are based on replacing the mel-scaled triangular filter banks with wavelet packet filter banks designed to match the acoustic nature and the frequency range of PCG signals. The extracted features using wavelet packet decomposition (WPD) are denoted as wavelet packet cepstral coefficients (WPCC). The second proposed cepstral features are based on modifying the mel-scaled triangular filter banks by increasing the nonlinearity of the triangular filters in the frequency range of PCG signals. The extracted features using nonlinear triangular filters are denoted as nonlinear frequency cepstral coefficients (NLFCC).

The outline of this paper is organized as follows. Section 2 presents a literature review of the proposed algorithms for heart sound biometric authentication systems. Section 3 describes the proposed heart sound biometric authentication system, specifically the four cepstral feature extraction techniques, in details. Different parameter settings for the proposed system are tested in Sect. 4 for the identification and verification modes. Finally, Sect. 5 summarizes the main conclusions and future directions.

2 The literature review

Various studies already introduced heart sounds as a biometric trait. The first attempt to use cardiac sounds for human recognition purpose was proposed in 2007 by Beritelli et al. [6]. Features were extracted using z-Chirp Transform (CZT) and classified using an Euclidean distance matching algorithm. False acceptance and rejection rates of 2.2 and 5 % were achieved, respectively, over a population of 20 individuals. Another study was followed by Phua et al. in 2008 [7]. Phua et al. developed an authentication system based on LFCC features with best correct recognition rate of 96 % over a database of 10 subjects. In 2010, a wavelet-based approach was adopted by Fatemian et al. [8,9] for PCG biometric identification and verification systems. Using a wavelet transform for the preprocessing stage and MFCC for feature extraction, Fatemain et al. were able to identify subjects with accuracy up to 100 % and to verify subjects with an equal error rate (EER) of 33 %. In 2013, Zhao et al. [10] presented a heart sound biometric system based on marginal spectrum analysis of PCG signals. A recognition rate up to 94.4 % was achieved over 40 participants. In the same year, Spadaccini and Beritelli [11] presented an evolution of a biometric identity verification system based on heart sounds. For preprocessing, the endpoints of $S1$ and $S2$ signals were detected. Then, the ratio between the power of $S1$ and $S2$ signals, known as First-to-Second Ratio (FSR), was computed. A statistical system was implemented using LFCC and FSR for feature extraction and GMM for classification. The best result obtained by Beritelli, computed over a database of 206 people, was an EER of 13.66 %. In the following section, our proposed PCG biometric authentication system is discussed in detail.

3 Proposed PCG biometric authentication system

Any biometric authentication system consists of four basic modules: database, preprocessing module, feature extraction module, and classifier and decision-making module [12]. The proposed algorithms for these modules are discussed in detail as follows.

3.1 Databases

In this paper, the four cepstral features are evaluated over two databases which are described hereafter.

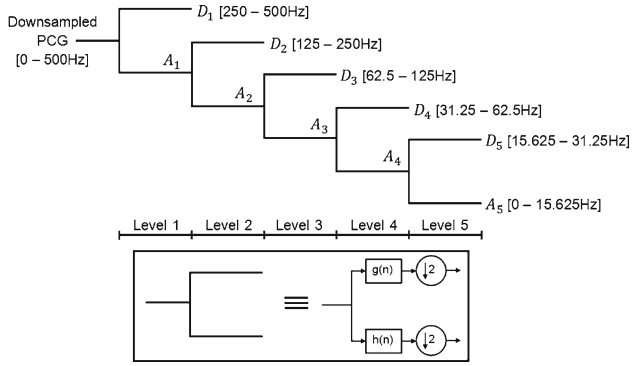

Fig. 3 DWT of downsampled PCG signal up to the fifth level

3.1.1 BioSec. PCG database

The BioSec. PCG database is a database of 21 subjects which was collected by the Biometrics Security Laboratory (BioSec. lab) at the University of Toronto, Canada [8]. The PCG signals were collected using the Littmann electronic stethoscope model 4100WS with a sampling rate of 8000 Hz and 16 bits per sample. For each subject, 6 recordings each of length 8 s under the rest condition were collected. The recording position was specified for each subject, either aortic or pulmonary according to the strength of the heard heart sound [13]. This database is not available for public access; however, it was used in this paper under a signed agreement between the University of Toronto in Canada and Assiut University in Egypt.

3.1.2 HSCT-11 database

The HSCT-11 is an open database which was previously collected by the University of Catania, Italy [11]. This database is freely available for public access [14]. The HSCT-11 database is, to the best of our knowledge, the largest heart sound database in terms of the number of people that contributed to it (206 subjects). The sensor used for acquisition is a ThinkLabs Rhythm digital electronic stethoscope; the files were acquired using a sampling frequency of 11,025 Hz and 16 bits per sample and are stored using the WAV format. This database contains only sequences recorded near the pulmonary valve.

3.2 Preprocessing

The main objective of the preprocessing stage is to provide a noise reduction scheme for further analysis of the PCG signal. Due to the nonstationary property of the PCG signal, the discrete wavelet transform (DWT) is used for the preprocessing stage. DWT provides a strong tool that is particularly applicable for analyzing problems of this type [3,9,15].

Table 1 Description of the four threshold selection rules [17]

Threshold selection rule	Description
rigrsure	Selection using the principle of Stein's Unbiased Risk Estimate (SURE) [18]
sqtwolog	Fixed form threshold which equals to $\sqrt{2 \times \text{length(wavelet coefficients)}}$
heursure	Selection using a mixture of the first two options mentioned
minimaxi	Threshold selection using minimax principle

Table 2 Description of the three threshold rescaling values

Threshold rescaling	Description
one	Performs no rescaling for the threshold value
sln	Uses a single-level estimation of the scaling value based on energy of the first-level wavelet coefficients
mln	Uses multilevel dependent estimation of the scaling values based on energy of each wavelet coefficients

The frequency components of PCG signals are concentrated below 250 Hz [7]. So, the sampling frequencies of the HSCT-11 (11,025 Hz) and the BioSec. (8000 Hz) databases are very high, consuming large time for computations. Therefore, the PCG signal is downsampled by a factor of 11 for the HSCT-11 and 8 for the BioSec. database to get a sampling frequency of approximately 1000 Hz which is greater than the Nyquist rate. Since most of the energy in PCG signal is concentrated in the range 15–200 Hz, the PCG signal is decomposed using the DWT up to the fifth level ($i = 5$) which is appropriate for PCG signal. This is illustrated in Fig. 3. The denoising is performed by applying a thresholding scheme for the wavelet coefficients. The threshold selection rules used in this paper are previously proposed in [16]. In [16], four threshold selection rules and three threshold rescaling values were adopted for optimal wavelet denoising of PCG signals as summarized in Tables 1 and 2. The four threshold selection rules with the three rescaling values are evaluated and compared in the experiments provided in Sect. 4.1. After threshold estimation, the wavelet coefficients are thresholded using soft thresholding. Finally, the denoised PCG signal is reconstructed again using the thresholded wavelet coefficients.

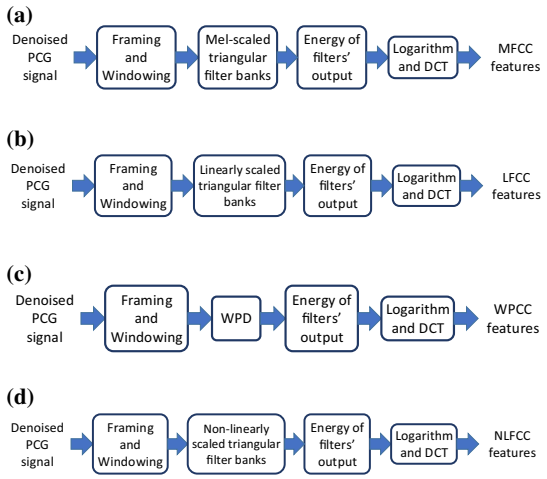

Fig. 4 Block diagram of the four cepstral features. **a** MFCC. **b** LFCC. **c** WPCC. **d** NLFCC

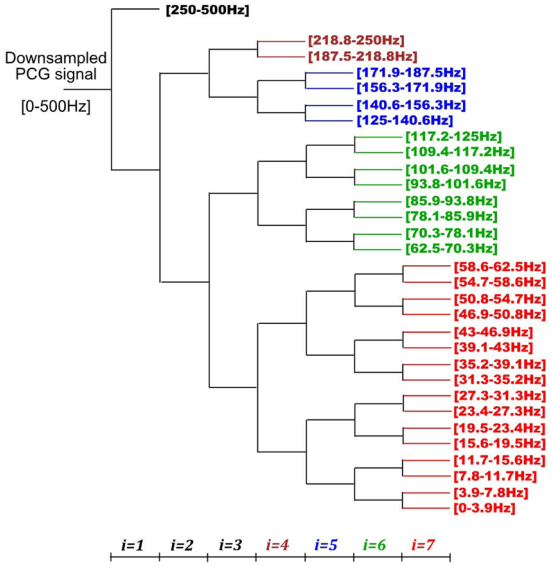

Fig. 5 The proposed filter bank structure for WPCC features

3.3 Feature extraction

As mentioned earlier, four cepstral features are evaluated and compared in this paper. The block diagram of the four cepstral features is illustrated in Fig. 4. As shown, the four types of cepstral features are similar in all blocks except for the type of the filter banks. Each block of cepstral features is discussed hereafter.

3.3.1 Framing and windowing

The first step in computing cepstral features is dividing the PCG signals into smaller frames. The average normal heart rate ranges from 80 to 110 beats per minute [19]. Therefore, a frame length of 1 s with a frame shift of 0.5 s is chosen for our proposed PCG biometric system so that each frame has at least one period of PCG. In the next step, each frame is multiplied by Hamming window to decrease the effect of discontinuities at the borders of each frame.

3.3.2 Filter banks

- *Mel-filter banks* MFCC features are obtained by passing the framed signal through a set of triangular filters which are equally spaced along the mel-frequency scale. Mel-scale is a nonlinear frequency scale that is related to the linear frequency scale by the following equation

$$f_{\text{mel}} = 2595 \log\left(\frac{1 + f_{\text{lin}}}{700}\right) \tag{1}$$

  where f_{mel} and f_{lin} are the mel-scale and linear frequency scale. In speech recognition, it has been shown that the human ear resolves frequencies nonlinearly across the audio spectrum. Therefore, MFCC has proved to have effective performance in automatic speaker and speech recognition applications [20–22]. As PCG signals are acoustic by nature, MFCC was adopted in many previously implemented PCG biometric authentication systems [7,8,23].
- *Linear filter banks* LFCC is similar to MFCC; however, the set of triangular filter banks is equally spaced in the linear frequency scale.
- *Wavelet packet filter banks* Unlike MFCC and LFCC, WPCC uses wavelet packet filters instead of triangular filters. WPCC features were previously used for automatic speech recognition systems [24–26]; however, they have not been yet adopted for PCG biometric authentication systems according to authors' knowledge. When dealing with wavelets, a problem usually arises which is choosing the best wavelet function to represent the PCG signal. MATLAB provides many wavelet families in its wavelet toolbox like Daubechies family (db), Symlets family (sym), Coiflets family (coif), biorthogonal and reverse biorthogonal spline families (bior, rbio), and discrete Meyer wavelet (dmey) [27]. All these wavelet families are tested and compared in Sect. 4.1. The proposed wavelet filter banks are illustrated in Fig. 5. The downsampled PCG signal is decomposed using WPD to the third level. This splits the range 0–250 Hz of the signal into 4 bands each of 62.5 Hz bandwidth. The bands (0–62.5 Hz), (62.5–125 Hz), (125–187.5 Hz), and (187.5–250 Hz) are further decomposed up to the seventh, sixth, fifth, and fourth levels, respectively. The total number of selected bases is 30 (16 + 8 + 4 + 2). The proposed filter bank structure has two advantages. Firstly, the lower band (0–62.5 Hz) has the finest resolution which gives

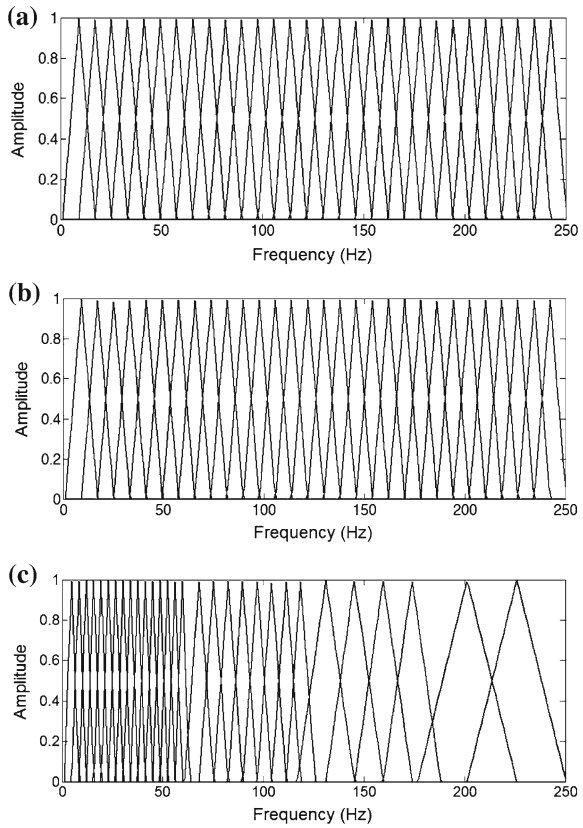

Fig. 6 Comparing between triangular filters for MFCC, LFCC, and NLFCC features. **a** Filter banks equally spaced in mel-scale. **b** Filter banks equally spaced in linear scale. **c** Proposed nonlinear filter banks

best representation for the primary sounds of PCG signal. Secondly, the bandwidth of the selected bases changes nonlinearly as the frequency increases from 0 to 250 Hz which matches the acoustic nature of the PCG signal [28,29].

- *Nonlinear filter banks* The nonlinearity of mel-frequency scale, as provided in Eq. 1, increases for frequencies above 1000 Hz, and the mel-scale is considered linear below this value. Therefore, MFCC is more suitable for speech signals than PCG signals because speech signals have higher frequency ranges (20–8000 Hz) than PCG which is concentrated below 250 Hz. Figure 6 shows a comparison between mel-scaled (Fig. 6a) and linearly scaled (Fig. 6b) triangular filters in the frequency range where most of PCG energy is concentrated (0–250 Hz). As shown, the two filter banks are identical because below 1000 Hz the mel-scale is approximately linear. So, to increase the nonlinearity, a new triangular filter bank is proposed as in Fig. 6c. The range (0–250 Hz) is divided equally into four bands each of 62.5 Hz. The bands (0–62.5 Hz), (62.5–125 Hz), (125–187.5 Hz), and (187.5–250 Hz) are passed through 16, 8, 4, and 2 equally spaced triangular filters, respectively.

3.3.3 Logarithm of energy and discrete cosine transform (DCT)

The energy of each filter output is calculated as follows

$$E_m = \frac{1}{N_m} \sum_{n=1}^{N_m} |y_m(n)|^2, \quad m = 1, 2, 3, \ldots, M \tag{2}$$

where $y_m(n)$ is the output of the filter bank number m and N_m is the total number of samples of the output signal. E_m is the output energy of the filter m, and M is the total number of filter banks (in this paper $M = 30$). Another method for calculating energy is adopted for WPCC features based on the Teager Energy Operator (TEO). TEO-based cepstral features have been shown to achieve better speech recognition rates than the standard energy calculation method (Eq. 2) in the presence of car noise [30,31]. The Teager energy of each filter output is calculated as

$$TE_m = \frac{1}{N_m} \sum_{n=1}^{N_m} |\Psi[y_m(n)]|, \quad m = 1, 2, 3, \ldots, M \tag{3}$$

where $\Psi[y_m(n)]$ is the TEO of the filter output which can be approximated by [30]

$$\Psi[y_m(n)] = y_m(n)^2 - y_m(n+1)y_m(n-1) \tag{4}$$

Finally, cepstral features are obtained by taking the logarithm and applying DCT to the energy values as follows:

$$CC_k = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} \log(E_m) \cos\left(\frac{\pi k}{M}\left(m - \frac{1}{2}\right)\right) \tag{5}$$

where CC_k represents the kth cepstral coefficient ($k = 0, 1, 2, \ldots, N_c - 1$) and N_c is the number of cepstral coefficients which has a maximum value of M (number of filter banks). The term E_m in Eq. 5 is replaced by TE_m to obtain TEO-WPCC features.

3.4 Classification

To achieve best discrimination between different classes, a suitable classifier that best fits the scattering distributions of these classes has to be used. Linear discriminant analysis (LDA) is adopted as a classifier for the proposed system due to its high performance and low computation time. LDA assumes that the features extracted from every class (subject), c, have a multivariate Gaussian distribution with mean, μ_c [32]. Moreover, LDA assumes that all the classes have the same covariance, i.e., $\Sigma_c = \Sigma$ for all values of c. The classifier decision is performed using the optimum Bayes rule which maximizes the posterior probability or its logarithm as shown in the following equation:

$$G(x) = \underset{c}{\operatorname{argmax}}\left[-\frac{1}{2}\left(x_f - \mu_c\right)^T \Sigma^{-1}\left(x_f - \mu_c\right) + \log(f_c)\right] \tag{6}$$

where f_c is the prior probability of class c (which is assumed to be uniform) and $G(x)$ is the logarithmic probability that the unknown feature vector x_f belongs to class c. Equation 6 represents a linear equation where the boundaries between classes are determined by a straight line.

4 Experimental setup and results

The experimental setup of the proposed system is divided into two phases: training and testing. For evaluating the proposed system, 55 feature vectors are selected randomly for each subject from the two databases. Then, 50 feature vectors out of the 55 vectors are selected randomly for the training phase to train the classifier model. The remaining 5 feature vectors are averaged to generate the testing sample. The performance of the system is evaluated under two modes: identification and verification as discussed in the following subsections.

4.1 Identification mode

In identification mode, the system is evaluated under different parameter settings, namely threshold selection rule, threshold rescaling, wavelet filter type, and number of cepstral features (N_c). In the test phase, the features of each subject are compared with the templates, computed in the training phase, of the whole dataset to find the best match. The performance of the system is evaluated using the average Correct Recognition Rate (CRR_{av}). CRR_{av} is calculated as follows

$$\text{CRR}(j) = \frac{\text{Number of correctly identified subjects in trial } j}{\text{Total number of subjects}} \tag{7}$$

$$\text{CRR}_{\text{av}} = \frac{1}{J}\sum_{j=1}^{J}\text{CRR}(j) \tag{8}$$

and J is the number of times the experiment is repeated. In this paper, each experiment is run for $J = 800$ and $J = 200$ times for the BioSec. and the HSCT-11 database, respectively. Then, CRR_{av} is obtained using Eq. 8.

Tables 3 and 4 show the performance of the proposed system using the four threshold selection rules and the three rescaling options over the two databases under different values of Signal-to-Noise Ratio (SNR). Based on the achieved results over the two databases, the 'rigrsure' and 'heursure' threshold selection schemes are superior to the others. However, 'rigrsure' is slightly better than 'heursure' for most of the cases. Also, the 'sln' method outperforms the other two rescaling options. Therefore, the parameter settings achieving highest accuracy for denoising are 'rigrsure' for threshold selection and 'sln' for threshold rescaling. This test is performed using the fifth order Daubechies wavelet (db5) and 24 WPCC feature coefficients ($N_c = 24$). Note that the same wavelet function is used for preprocessing and feature extraction stages.

Tables 5 and 6 summarize the performance of the four types of cepstral features over the HSCT-11 and the BioSec. databases, respectively, using different types of wavelet fil-

Table 3 Comparing CRR_{av} (%) using different threshold selection rules and rescaling schemes as proposed in [16] under different SNR values for the HSCT-11 database

SNR value (dB)	rigrsure			heursure			minimaxi			sqtwolog		
	sln	mln	one	sln	mln	one	sln	mln	one	sln	mln	one
10	**80.83**	78.74	22.46	80.63	80.19	19.13	74.98	67.2	19.55	72.17	62	19.14
5	76.86	73.99	19	**77.04**	75.1	15.77	62.27	59.65	16.34	62.95	54.01	15.74
0	**70.01**	66.31	15.69	67.39	64.99	12.05	54.2	46.39	12.62	47.9	40.7	12.11

Bold values represent the highest recognition rates achieved

Table 4 Comparing CRR_{av} (%) using different threshold selection rules and rescaling schemes as proposed in [16] under different SNR values for the BioSec. database

SNR value (dB)	rigrsure			heursure			minimaxi			sqtwolog		
	sln	mln	one	sln	mln	one	sln	mln	one	sln	mln	one
10	**79.93**	76.38	44.99	76.83	70.32	41.69	70.44	65.84	41.98	67.96	60.93	40.61
5	75.13	**75.67**	43.82	66.33	58.98	39.67	60.42	55.73	42.28	57.03	53.48	39.07
0	**70.02**	63.71	42.56	58.95	47.86	38.34	48.13	48.84	39.3	46.11	45.98	38.45

Bold values represent the highest recognition rates achieved

Table 5 A Comparison between the minimum and maximum CRR_{av} (%) of the four cepstral features for different wavelet filters using the HSCT-11 database

Wavelet filters	Daubechies (db1–db45)		Symlets (sym1–sym20)		Biorthogonal (bior1.1–bior6.8)		Reverse biorthogonal (rbior1.1–rbior6.8)		Coiflets (coif1–coif5)		Discrete Meyer (dmey)
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	
MFCC	91.3	**92.05**	91.3	91.63	91.3	91.53	91.3	92	91.4	91.48	91.61
LFCC	90.94	**91.61**	90.94	91.17	90.94	91.12	90.81	91.51	90.99	91.09	91.15
WPCC	58.37	89.93	58.37	89.3	38.45	85.74	58.37	85.83	72.09	87.88	**90.26**
NLFCC	92.24	**92.94**	92.24	92.52	92.23	92.45	92.22	92.78	92.3	92.37	92.51

Bold values represent the highest recognition rates achieved

Table 6 A Comparison between the minimum and maximum CRR_{av} (%) of the four cepstral features for different wavelet filters using the BioSec. database

Wavelet filters	Daubechies (db1–db45)		Symlets (sym1–sym20)		Biorthogonal (bior1.1–bior6.8)		Reverse biorthogonal (rbior1.1–rbior6.8)		Coiflets (coif1–coif5)		Discrete Meyer (dmey)
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	
MFCC	96.56	97.35	96.56	**97.36**	96.51	97.35	96.44	**97.36**	97.17	97.32	97.31
LFCC	96.15	**97.05**	96.15	97.02	96.04	97	95.63	97.01	96.89	96.98	96.94
WPCC	75.17	96.63	75.17	96.22	64.64	91.49	75.17	91.62	79.76	94.1	**97.02**
NLFCC	97.82	**98.1**	97.82	**98.1**	97.66	**98.1**	97.5	98.08	97.98	98.08	98.05

Bold values represent the highest recognition rates achieved

Table 7 Comparing the robustness of cepstral features under different SNR values of additive white Gaussian noise

Database	HSCT-11 database					BioSec. database				
SNR value (dB)	MFCC	LFCC	NLFCC	WPCC	TEO-WPCC	MFCC	LFCC	NLFCC	WPCC	TEO-WPCC
10	90.11	89.43	**91.37**	89.08	89.04	86.33	84.69	**92.18**	88.82	89.14
5	88.28	87.69	**89.67**	87.02	86.94	79.6	77.93	**88.12**	85.02	84.53
0	83.12	82.51	**84.76**	81.17	80.98	69.83	67.71	**80.48**	75.24	73.92

Bold values represent the highest recognition rates achieved

Table 8 Comparing the robustness of cepstral features under different SNR values of additive background noise

Database	HSCT-11 database					BioSec. database				
SNR value (dB)	MFCC	LFCC	NLFCC	WPCC	TEO-WPCC	MFCC	LFCC	NLFCC	WPCC	TEO-WPCC
10	90.52	90.04	**90.79**	88.73	88.95	89.7	88.02	**92.71**	89.38	89.59
5	88.12	87.69	**88.54**	85.57	86.09	83.54	80.9	**88.46**	85.4	84.35
0	83.15	82.46	**84.15**	79.23	79.81	71.62	68.22	**79.34**	78.58	75.57

Bold values represent the highest recognition rates achieved

ters. In general, the higher order wavelets show better performance than the lower order ones; however, the computational time increases as the order of the wavelet increases. As a compromise between time computation requirements and correct recognition rate, (dmey) is considered the best wavelet choice as it needs low computation time and achieves high recognition rate in comparison with other wavelets. Also, MFCC, LFCC, and NLFCC features do not show a great variation in accuracy compared with WPCC features because these features involved signal reconstruction; however, WPCC features are based directly on wavelet coefficients. Moreover, the CRR_{av} using MFCC and LFCC are approximately equal for the proposed system over the two databases with MFCC is slightly better than LFCC features. Also, increasing the nonlinearity for the triangular filter banks in the NLFCC features showed better CRR_{av} values than MFCC over the two tested databases.

Tables 7 and 8 show a comparison between the four types of cepstral features under different levels of additive white Gaussian and background noises, respectively. The back-

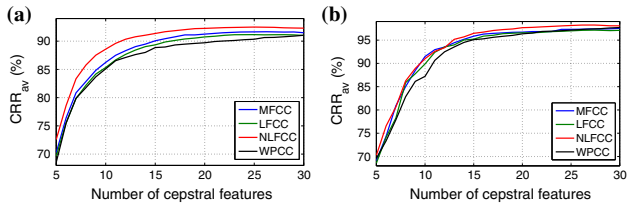

Fig. 7 The relation between average correct recognition rate and number of cepstral features. **a** HSCT-11 database. **b** BioSec. database

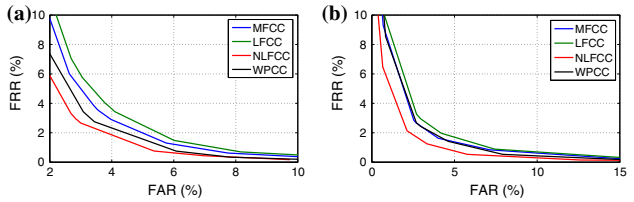

Fig. 8 Comparing the four cepstral features in verification mode. **a** FAR and FRR for the HSCT-11 database. **b** FAR and FRR for the BioSec. database

ground noise recording used in this paper is freely available on the Internet (http://soundbible.com/954-Background-Noise.html). It was downsampled to 1000 Hz (to be the same sampling frequency of the PCG signal) and then added to the PCG signal. Based on the achieved results, the proposed system using WPCC and NLFCC features shows better CRR_{av} values for high noise levels compared to MFCC and LFCC features using the BioSec. database. However, for the HSCT-11 database, the performance using the four cepstral features decreases with the same level by increasing noise. Moreover, using the Teager operator for energy estimation improves the performance of the system over different levels of additive background noise as shown in Table 8 using the HSCT-11 database. However, the Teager energy estimation did not show performance improvement over white Gaussian noise which is similar to the results achieved in [30].

Figure 7 shows the performance of the proposed system using different values for cepstral coefficients (N_c) over the two databases. The number of cepstral coefficients is varied from 5 to 30. In general, as N_c changes from 5 to 10, the CRR_{av} increases dramatically, and then, the CRR_{av} increases slightly to $N_c = 25$. After using 25 cepstral feature coefficients, the CRR_{av} remains approximately constant. This means that adding extra features has no effect on the system's performance. Therefore, it is sufficient to use 25 cepstral coefficients for the proposed system. This test is performed using (dmey) wavelet.

4.2 Verification mode

In the testing phase, the user, x_s, provides a claimed identity, I, and the database is divided into two classes: the I-related (H_0) and I-nonrelated (H_1) classes. The testing sample of x_s is then classified to one of the two classes as follows:

$$D(x_s, I) = \frac{p(x_s|H_0)}{p(x_s|H_1)} \begin{cases} \geq T_s, & x_s \in I(\text{genuine}) \\ < T_s, & x_s \notin I(\text{imposter}) \end{cases} \quad (9)$$

where $p(x_s|H_0)$ ($p(x_s|H_1)$) is the probability that x_s belongs (does not belong) to I. This probability is calculated in the logarithmic scale as in Eq. 6. T_s is a threshold which defines the degree of security of the system. The proposed system in this mode is evaluated using two parameters, namely false acceptance rate (FAR) and false rejection rate (FRR). These error rates depend on T_s; as T_s increases, FAR decreases and FRR increases making the system more secure. However, as T_s decreases, FAR increases and FRR decreases making the system less secure. At a given value of T_s, FAR and FRR will have the same values, i.e., equal error rate. In this paper, the four cepstral features are compared in verification mode over the two databases for T_s values ranging from 0.1 to 3. Figure 8 shows the achieved FAR and FRR for the four cepstral features. As shown, NLFCC and WPCC features achieved lower error rates for the two databases except for WPCC features in the BioSec. database which achieved same error rates as MFCC features. This test is performed using 30 cepstral feature coefficients and (dmey) wavelet for all cepstral features. As a summary, Table 9 presents a comparison between the four cepstral features in terms of EER over the two databases.

5 Conclusion and future directions

This paper presented a comparative study between different cepstral features for PCG biometric authentication application. In this paper, two new cepstral features were proposed and compared with MFCC and LFCC features. The two proposed features are based on replacing the mel-scaled triangular filters with wavelet packet filters and increased nonlinear triangular filter banks as discussed in Sect. 3.3.2. The proposed system with the four cepstral features was evaluated over two PCG databases. Of the four cepstral features, the proposed NLFCC features outperform the other cepstral features over the two databases in the two modes of authentication: identification and verification. This agrees with the acoustic nature of PCG and the fact that the human ear

Table 9 EER for the proposed cepstral features over the two databases

Database	HSCT-11 database				BioSec. database			
Features	MFCC	LFCC	NLFCC	WPCC	MFCC	LFCC	NLFCC	WPCC
EER (%)	3.55	3.87	**2.88**	3.2	2.68	2.95	**2.13**	2.69

Bold values represent the lowest error rates achieved

resolves frequencies nonlinearly across the audio spectrum. Also, NLFCC and WPCC features showed better performance than MFCC and LFCC features under high noise levels as discussed in Sect. 4.1.

In this paper, four cepstral features were adopted and compared over two PCG databases. For improving the performance, a feature level fusion can be adopted using feature fusion techniques such as canonical correlation analysis (CCA) [33] or partial least squares (PLS) [34]. Although the proposed system was found to achieve good performance on a large database of 206 subjects with short testing PCG sample of 3 s, a database of thousands of subjects should be built and tested for the implementation of commercial PCG biometric systems. Also, numerous factors have to be considered in practical applications such as age, gender, heart rate, and heart disease status.

Acknowledgments The authors would like to thank Dimitrios Hatzinakos, Professor in the Electrical and Computer Engineering Department, University of Toronto, for his help in providing the BioSec. PCG database that made this work possible.

References

1. Matsumoto, T., Matsumoto, H., Yamada, K., Hoshino, S.: Impact of artificial gummy fingers on fingerprint systems. In: Electronic Imaging 2002. International Society for Optics and Photonics, pp. 275–289 (2002)
2. Nixon, K.A., Aimale, V., Rowe, R.K.: Spoof Detection Schemes. In: Jain, A.K., Flynn, P., Ross, A.A. (eds.) Handbook of Biometrics. Springer, pp. 403–423 (2008)
3. Abbas, A.K., Bassam, R.: Phonocardiography signal processing. Synth. Lect. Biomed. Eng. **4**(1), 1–194 (2009)
4. Abo-Zahhad, M., Ahmed, S.M., Abbas, S.: Biometric authentication based on pcg and ecg signals: present status and future directions. Signal Image Video Process. **8**(4), 739–751 (2014)
5. Beritelli, F., Spadaccini, A.: Human identity verification based on heart sounds: recent advances and future directions. In: Yang, J. (ed.) Biometrics. INTECH Open Access Publisher, pp. 217–234 (2011)
6. Beritelli, F., Serrano, S.: Biometric identification based on frequency analysis of cardiac sounds. IEEE Trans. Inf. Forensics Secur. **2**(3), 596–604 (2007)
7. Phua, K., Chen, J., Dat, T.H., Shue, L.: Heart sound as a biometric. Pattern Recog. **41**(3), 906–919 (2008)
8. Fatemian, S.Z., Agrafioti, F., Hatzinakos, D.: Heartid: Cardiac biometric recognition. In: 2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS). IEEE, pp. 1–5 (2010)
9. Fatemian, S.Z.: A wavelet-based approach to electrocardiogram (ecg) and phonocardiogram (pcg) subject recognition. Master's thesis, University of Toronto (2009)
10. Zhao, Z., Shen, Q., Ren, F.: Heart sound biometric system based on marginal spectrum analysis. Sensors **13**(2), 2530–2551 (2013)
11. Spadaccini, A., Beritelli, F.: Performance evaluation of heart sounds biometric systems on an open dataset. In: 2013 18th International Conference on Digital Signal Processing (DSP). IEEE, pp. 1–5 (2013)
12. Delac, K., Grgic, M.: A survey of biometric recognition methods. In: Electronics in Marine 2004, Proceedings Elmar 2004, 46th International Symposium. IEEE, pp. 184–193 (2004)
13. Biosec. pcg database. http://www.comm.utoronto.ca/~biometrics/databases.html
14. Hsct-11 open database. http://www.diit.unict.it/hsct11/
15. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press, Waltham (1999)
16. Messer, S.R., Agzarian, J., Abbott, D.: Optimal wavelet denoising for phonocardiograms. Microelectron. J. **32**(12), 931–941 (2001)
17. Guo, D.-f., Zhu, W.-h., Gao, Z.-m., Zhang, J.-q.: A study of wavelet thresholding denoising. In: 5th International Conference on Signal Processing Proceedings, 2000. WCCC-ICSP 2000, vol. 1. IEEE, pp. 329–332 (2000)
18. Donoho, D.L., Johnstone, J.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994)
19. Opthof, T.: The normal range and determinants of the intrinsic heart rate in man. Cardiovasc. Res. **45**(1), 177–184 (2000)
20. Murty, K., Yegnanarayana, B.: Combining evidence from residual phase and mfcc features for speaker recognition. IEEE Signal Process. Lett. **13**(1), 52–55 (2006)
21. Vergin, R., O'shaughnessy, D., Farhat, A.: Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition. IEEE Trans. Speech Audio Process. **7**(5), 525–532 (1999)
22. On, C.K., Pandiyan, P.M., Yaacob, S., Saudi, A.: Mel-frequency cepstral coefficient analysis in speech recognition. In: International Conference on Computing & Informatics, 2006. ICOCI'06. IEEE, pp. 1–5 (2006)
23. Beritelli F., Spadaccini, A.: Human identity verification based on mel frequency analysis of digital heart sounds. In: 2009 16th International Conference on Digital Signal Processing. IEEE, pp. 1–5 (2009)
24. Erzin, E., Yardimci, Y., et al.: Subband analysis for robust speech recognition in the presence of car noise. In: 1995 International Conference on Acoustics, Speech, and Signal Processing, 1995. ICASSP-95, vol. 1. IEEE, pp. 417–420 (1995)
25. Pavez, E., Silva, J.F.: Analysis and design of wavelet-packet cepstral coefficients for automatic speech recognition. Speech Commun. **54**(6), 814–835 (2012)
26. Farooq, O., Datta, S.: Mel filter-like admissible wavelet packet structure for speech recognition. IEEE Signal Process. Lett. **8**(7), 196–198 (2001)
27. Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.-M.: Matlab wavelet toolbox user's guide. Version 3 (2004)
28. Quatieri, T.F.: Discrete-Time Speech Signal Processing: Principles and Practice. Pearson Education India, Gurgaon (2002)
29. Rabiner, L.R., Juang, B.-H.: Fundamentals of Speech Recognition, vol. 14. PTR Prentice Hall, Englewood Cliffs (1993)
30. Jabloun, F., Erzin, E., et al.: Teager energy based feature parameters for speech recognition in car noise. IEEE Signal Process. Lett. **6**(10), 259–261 (1999)
31. Jabloun, F., et al.: The teager energy based feature parameters for robust speech recognition in car noise. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999, vol. 1. IEEE, pp. 273–276 (1999)
32. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Academic Press, Waltham (2008)
33. Abo-Zahhad, M., Ahmed, S.M., Abbas, S.N.: Pcg biometric identification system based on feature level fusion using canonical correlation analysis. In: 2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, pp. 1–6 (2014)
34. Sun, Q.-S., Jin, Z., Heng, P.-A., Xia, D.-S.: A novel feature fusion method based on partial least squares regression. In: Singh, S., Singh, M., Apte, C., Perner, P. (eds.) Pattern Recognition and Data Mining. Springer, pp. 268–277 (2005)