Available online at www.sciencedirect.com
ScienceDirect
Biomedical Signal Processing and Control 2 (2007) 80–96
Biomedical Signal Processing and Control
www.elsevier.com/locate/bspc

Wavelet energy based diagnostic distortion measure for ECG

M. Sabarimalai Manikandan *, S. Dandapat

Department of Electronics and Communication Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

Received 4 December 2006; received in revised form 26 April 2007; accepted 2 May 2007
Available online 26 June 2007

Abstract

In this paper, a novel Wavelet Energy based diagnostic distortion (WEDD) measure is proposed to assess the reconstructed signal quality for ECG compression algorithms. WEDD is evaluated from the Wavelet coefficients of the original and the reconstructed ECG signals. For each ECG segment, a Wavelet energy weight vector is computed via five-level biorthogonal discrete Wavelet transform (DWT). WEDD is defined as the sum of Wavelet energy weighted percentage root mean square difference of each subband. The effectiveness of this measure is validated by linear (linear polynomial and cubic polynomial) and nonlinear (logistic) regression analysis between the computed WEDD values and the mean opinion score (MOS) given by cardiologists. WEDD provides a better prediction accuracy and exhibits a statistically better monotonic relationship with the MOS ratings than Wavelet based weighted percentage root mean square difference (PRD) measure (WWPRD), PRD and other objective measures. Standard correlation coefficient and Spearman rank-order correlation coefficient (SROCC) between the WEDD/MOS ratings is 0.969 and 0.9624, respectively.

Keywords: Electrocardiogram; ECG; Diagnostic distortion measures; WEDD; WWPRD; PRD; Nonlinear regression

1. Introduction

Electrocardiogram signal is widely used for diagnosis of cardiovascular diseases. In recent years, the growth in digital technology and cost reduction of miniaturized digital acquisition units and communication devices have incited to acquire and handle the ECG signals in digital form. For cardiac analysis, ECG data are continuously recorded up to 72 h to monitor ischemia, ventricular and supra-ventricular dysrhythmias, conduction abnormalities, QT interval, and heart rate variability. For high resolution ECG data acquisition, a broad bandwidth of 0.05–300 Hz is considered in order to capture all diagnostic information. As a result, ECG signals are typically sampled and digitized at a sampling frequency of 1000 Hz with resolution ranging from 8 to 12 bits for good signal reconstruction [1]. The amount of ECG data to be stored or transmitted depends upon sampling frequency, number of quantization levels, number of channels and duration of ECG recording. Consequently, ECG record management system and telecardiology application requires more reliable and efficient compression algorithms for ECG data while preserving clinical information for the accurate detection and classification of ECG features [2–4].

Various lossy and lossless ECG compression algorithms are reported in the literature [5–21,23,25]. Lossless compression methods permit near perfect reconstruction of the original ECG signal while compression ratio is poor. For analysis of late potentials, it may be convenient to use lossless ECG compression. On the other hand, lossy data compression methods can significantly reduce the amount of data. The main challenge in an ECG compression method is to minimize the storage requirements without losing the clinically significant information. The evaluation of reconstructed ECG signal quality is of critical importance in the field of ECG compression. Most commonly, the mean squared error (MSE) between the original and the decompressed ECG signal is used for the evaluation compression methods. Some of the variants of MSE are normalized MSE [7], root MSE (RMSE) [9], normalized RMSE (NRMSE) [10], percentage root mean square difference [5,6,11–20] and signal to noise ratio (SNR) [17]. These measures fail to capture the diagnostic distortion of the local waves in a reconstructed signal.

* Corresponding author at: Electro Medical and Speech Technology Laboratory, Electronics and Communication Engineering Department, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India.
Tel.: +91 361 2582505; fax: +91 361 2582542.
E-mail addresses: msm@iitg.ernet.in (M.S. Manikandan), samaren@iitg.ernet.in (S. Dandapat).

1746-8094/$ – see front matter
doi:10.1016/j.bspc.2007.05.001

In the area of ECG signal compression, little attention has been paid towards the evaluation of distortion of clinical information. A suitable diagnostic distortion measure can help proper evaluation of the ECG compression schemes. The best criterion for the evaluation of ECG signal quality is the subjective test which is based on the visual perception of experienced cardiologist [22,24]. Mean opinion score (MOS) is used as a metric for the subjective quality test which is referred as gold standard [22]. Though the subjective quality test is the most efficient criterion [22,24,25] for the clinical quality assessment, it is time-consuming and expensive. Thus, a simple objective error measure for estimating subjective quality of the diagnostic features are desirable in compression algorithms. It is also essential to have an objective distortion measure which can give an immediate and reliable estimate of the anticipated visual quality of a particular compression algorithm. But this objective measure must be validated by comparing with subjective quality test to make it a standard metric. A good objective measure is the one which effectively predicts the subjective quality rating given by the cardiologists. Naturally, the objective diagnostic distortion measure is acceptable only when it is well correlated with cardiologists' perception.

To evaluate the performances of the objective error measure, the ECG signals are taken from the MIT-BIH Arrhythmia (`mita`) database[1] and the MIT-BIH ECG compression test (`mitect`) database.[2] The `mita` database contains 48 records of two-channel ambulatory ECG recordings. Each record is slightly longer than 30 min and is sampled at 360 samples/s with resolution of 11 bits per sample. All the sample values are in the range of −1024 to 1023. Therefore, baseline of 1024 is added to the original signal for storage purpose while it is subtracted from each sample in error measure. This gives the sample values ranging from 0 to 2047. The `mita` database consists of different rhythms, QRS complex morphologies, ectopic beats and noisy ECG signal and this database records are widely used for testing purpose in most of the compression methods reported [5,6,9,12–24]. The `mitect` database contains 168 short ECG recordings (20.48 s each). The sampling rate and the resolution are 250 samples/s and 12 bits, respectively. This database contains a wide variety of arrhythmias, conduction disturbances, and noise. In the following section, the performance of the conventional reconstruction error measures are discussed with usage of these database records.

2. Review of ECG signal distortion measures

In this section, various distortion measures which are used for the evaluation of performance of the ECG compression methods are discussed.

2.1. Subjective error measures

Signal quality for medical use is evaluated by visual inspection of the diagnostic features such as amplitudes, durations and shape of the waves. Cardiologists are ultimately the judges of ECG signal quality and expresses their judgements of diagnostic feature qualities according to a given MOS scale. The MOS is a mapping of the level of diagnostic features such as amplitude, duration, and the shape of the PQRST complexes distortion into either the descriptive terms bad, not bad, good, very good and excellent or into equivalent numerical ratings in the range of 1–5. Finally, the scores are averaged across subjects to obtain the final MOS. This measure is correlated with diagnostic information in the signal but it is time-consuming and expensive. In this paper, the mean opinion score (MOS) is used as a basis for validating our proposed objective measure.

2.2. Objective error measures

There are basically two classes of objective quality or distortion assessment methods. The first class includes non-diagnostic distortion measures which evaluates the global and local errors in the reconstructed signal. The second class of quality measures consider diagnostic features in an attempt to incorporate visual inspection quality by cardiologists. These measures are discussed in the following subsections.

2.2.1. Non-diagnostic error measures

This subsection discusses some of the global error and local error measures. One such a distortion criteria or error measure is the mean square error. Given an original signal, $x(n)$, consisting of N samples, and a reconstructed approximation to this signal, $\tilde{x}(n)$, the mean square error is given by

$$\mathrm{MSE} = \frac{1}{N}\sum_{n=1}^{N}[x(n) - \tilde{x}(n)]^2 \tag{1}$$

Geometrically, this distortion measure is the mean of the square of Euclidean distance between the input and the output vectors. In some papers, normalized form of MSE is also used as an error measure [7]. NMSE is expressed as

$$\mathrm{NMSE} = \frac{\sum_{n=1}^{N}[x(n) - \tilde{x}(n)]^2}{\sum_{n=1}^{N}[x(n)]^2} \tag{2}$$

Generally, normalization process is done to make the error measure independent of the amplitude scale of the original signals. For example, ECG signals taken from `mita` database have different maximum and minimum amplitude. Basically, these amplitudes vary for different subjects and leads. Therefore, measured error value should be normalized to conform within a standard range.

Another error measure criterion employed to evaluate the distortion introduced by compression and quantization is the root mean square error [9] which is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[x(n) - \tilde{x}(n)]^2} \tag{3}$$

[1] http://www.physionet.org/physiobank/database/mitdb.
[2] http://www.physionet.org/physiobank/database/cdb/.

The modified version of RMSE, normalized root mean square error, is used as an error measure [10]. NRMSE is defined as

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{n=1}^{N}[x(n)-\tilde{x}(n)]^2}{\sum_{n=1}^{N}[x(n)]^2}} \tag{4}$$

In general, the percentage rms difference is a normalized value which indicates the error between original and reconstructed signals. It is given by

$$\mathrm{PRD} = \left(\frac{\mathrm{rms_e}}{\mathrm{rms_v}}\right) \times 100 \tag{5}$$

where $\mathrm{rms_e}$, $\mathrm{rms_v}$ are the rms error and the rms value of the ECG signal, respectively. In most ECG compression algorithms, PRD is computed with different normalized value [11–20]. Based on the normalization process, PRD measures are defined accordingly as PRD1 [17–20], PRD2 [12–16] and PRD3 [5,19]. Most of the compression methods tested using `mita` database adopts the addition of 1024-baseline for storage purposes. Also the records in `mita` database have different mean value. Therefore, in [17–20], PRD1 is calculated after subtracting the mean value and the baseline of 1024 from the original signal. Here, PRD1 is defined as

$$\mathrm{PRD1} = \sqrt{\frac{\sum_{n=1}^{N}[x(n)-\tilde{x}(n)]^2}{\sum_{n=1}^{N}[x(n)-\mu_o-1024]^2}} \times 100 \tag{6}$$

where μ_o is the mean of the original signal. In [12–16], PRD2 defined as

$$\mathrm{PRD2} = \sqrt{\frac{\sum_{n=1}^{N}[x(n)-\tilde{x}(n)]^2}{\sum_{n=1}^{N}[x(n)-1024]^2}} \times 100 \tag{7}$$

PRD2 is determined with the inclusion of mean value of the ECG signal. In [5,19], PRD3 is defined as,

$$\mathrm{PRD3} = \sqrt{\frac{\sum_{n=1}^{N}[x(n)-\tilde{x}(n)]^2}{\sum_{n=1}^{N}[x(n)]^2}} \times 100 \tag{8}$$

PRD3 measure calculation includes both the mean value, μ_o and the base line of 1024. PRD3 value is very low for the given set of original and reconstructed signal as compared to the value of PRD1 and PRD2. But the reconstruction with low PRD value does not necessarily imply diagnostic acceptance. However, it is meaningless to compare the value of PRDs with different offsets. To illustrate this, ECG signal with block length of 1024 samples is taken from the `mita` record 100 for testing purpose. The mean value of ECG signal is −0.3135 and the respective baseline is 959.82. The `mita` record 100 is compressed at compression ratio of 8:1 using Wavelet based ECG compression with large zero zone quantizer [6]. The reconstruction error is measured using PRD1, PRD2 and PRD3 for different value of mean and baseline. Finally, the experimental results are shown in Fig. 1. It can be measured that the value of PRD1, PRD2 and PRD3 are 5.55%, 2.78% and 0.2146%, respectively. Here, the value of PRD3 is smaller than PRD1 and PRD2 due to the different normalized value

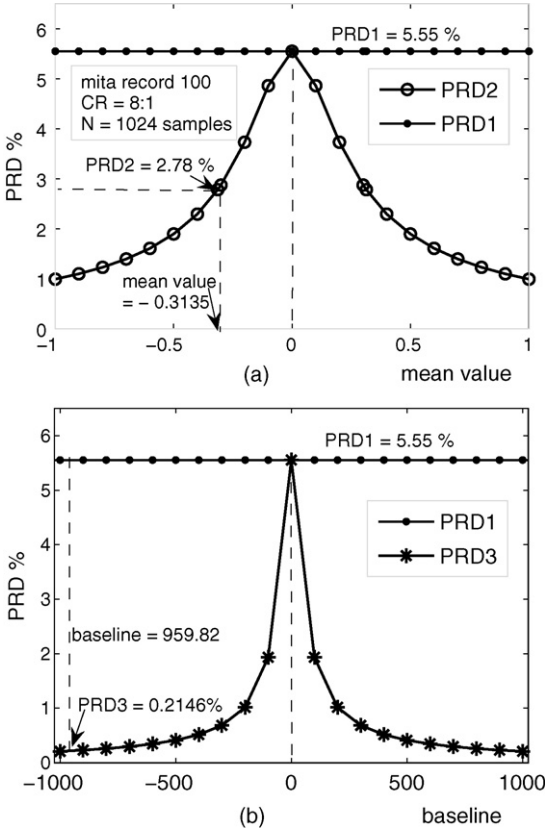

Fig. 1. The effects of different mean values and baseline on the PRD measure. (a) PRD as a function of mean value. (b) PRD as a function of baseline. Here, the `mita` record 100 is selected for testing purpose.

(denominator) but the rms error values (numerator) are equal. This experiment shows that the comparison of PRDs with different offsets is meaningless. It will be more pertinent if the mean value and the baseline of 1024 is removed from the original signal.

In [17], SNR is also employed to measure the distortion which is defined as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{n=1}^{N}[x(n)-\mu_o]^2}{\sum_{n=1}^{N}[x(n)-\tilde{x}(n)]^2}\right) \tag{9}$$

The MSE and its modified versions are global performance measures. For local error criterion, the maximum amplitude error (MAX) or peak error (PE) is used in ECG compression methods to find out the maximum error and determine its position within a cycle [8,17]. MAX is defined as

$$\underset{i}{\mathrm{MAX}} = \max_{n=1}^{N_{ci}}\{|x(n)-\tilde{x}(n)|\} \tag{10}$$

where, N_{ci} is the number of samples within an ith cycle. Another version of MAX, normalized maximum amplitude error (NMAX), gives the largest absolute value of errors normalized by the dynamic range of the signal within a cycle [10]. The NMAX for the ith cycle is

$$\mathrm{NMAX}_i = \frac{\max_{n=1}^{N_{ci}}\{|x(n)-\tilde{x}(n)|\}}{\max_{n=1}^{N_{ci}}\{x(n)\} - \min_{n=1}^{N_{ci}}\{x(n)\}} \tag{11}$$

The mean NMAX for the entire signal is determined by averaging over all the cycles. The maximum absolute error (MAX) measure also provides local error within each ECG beat. The maximum error occurs in isoelectric regions due to the presence of noise, which is depicted in Fig. 2 (c). This can be improved by weighting each error sample by the absolute value of its original sample, which is shown in Fig. 2(d). Now the maximum error occurs in the QRS complex location. But the average MAX for an ECG segment is obtained by averaging over all the blocks. From Fig. 2(d), it can be observed that the MAX error in the isoelectric regions also contributed some amount to the average MAX. The error measure can be improved further by weighting the error sample by the energy value of its original sample. The improvement in the error distribution is shown in Fig. 2(e). But it will not give which local waves are distorted within an ECG cycle because the local waves are spatially time variant. In this paper, the later concept is exploited and implemented in Wavelet transform domain because it provides good spatial frequency localization.

Another local error measure is the standard deviation of errors, StdErr, which is expressed as [16]

$$\mathrm{StdErr} = \sqrt{\frac{1}{N_c - 1}\sum_{n=1}^{N_c}[\mathrm{error}(n) - \mu_e]^2} \tag{12}$$

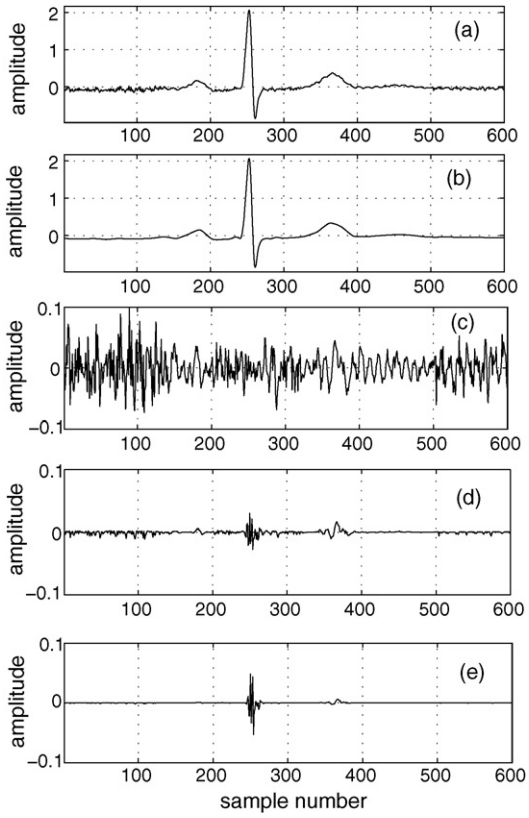

Fig. 2. Performance of the quality measure in time-domain. (a) The original signal from `mita` record 123 with introduced AWGN of −30 dB. (b) The reconstructed signal. (c) Error sample. (d) Error sample weighted by absolute value of its original sample. (e) Error sample weighted by energy value of its original sample.

where N_c is the number of error samples within a cycle, error the difference between the original sample and the reconstructed sample and μ_e is the mean of the errors.

In some papers, the cross-correlation measure is employed to evaluate the similarity between the original signal and the reconstructed signal. It is defined as [21]

$$\mathrm{CC} = \frac{(1/N)\sum_{n=1}^{N}[x(n) - \mu_o]\sum_{n=1}^{N}[\tilde{x}(n) - \mu_r]}{\sqrt{(1/N)\sum_{n=1}^{N}[x(n) - \mu_o]^2}\sqrt{(1/N)\sum_{n=1}^{N}[\tilde{x}(n) - \mu_r]^2}} \tag{13}$$

where μ_o and μ_r are the mean values of the original signal and the reconstructed signal, respectively. It is referred hereafter as normalized cross-correlation (NCC) in this paper.

The MSE and other similar error measures are mathematically convenient but have many disadvantages. Since MSE is only an average measure, it cannot adequately quantify the performance of an ECG compression algorithm. The MSE does not give any idea about how the error is distributed within any ECG cycle which is important in determining the clinical acceptability of the reconstructed signal [10]. The MSE measure fails to characterize the local distortion of an ECG signal. These non-diagnostic error measures are not well correlated with diagnostic features distortion of an ECG signal [13,14,17,20–25]. Therefore, these error measures are not good for evaluation of compression error. However, it is easy to calculate and thus, is widely used. Among these, the PRD is the most widely used measure for evaluation of the compression methods [17].

2.2.2. Diagnostic error measures

To solve the above mentioned problems, researchers have proposed new objective error measures that considered the diagnostic distortion of the local waves such as P-wave, Q-wave, QRS complex, ST segment and T-wave of the reconstructed ECG signal. Chen and Itoh [21] suggested a new distortion measure, the weighted PRD to improve the local distortion measure for evaluating the fidelity of the reconstructed ECG signal. The *weighted* PRD (WPRD) is defined as

$$\mathrm{WPRD} = \sqrt{\frac{\sum_{k=1}^{M} w_k \gamma_k}{\sigma}} \tag{14}$$

where w_k are the weights, γ_k the reconstruction MSE of the diagnostic features such as P-wave, Q-wave, QRS-wave and ST-wave, M is the number of diagnostic features and σ is the power of the original signal. The weights can be chosen according to the desired diagnostic feature error measure. However, there is no procedure reported to select the optimal weights for each local waves [21]. The WPRD measure depends on the accurate extraction of local waves within each beat and the weights for the significant waves.

The weighted diagnostic distortion (WDD) measure was proposed and implemented by considering the diagnostic features of the the original ECG signal and the reconstructed signal [22]. The measure uses diagnostic feature vector which consists of 18 features (durations, amplitudes and shapes). The

WDD (in percentage) is defined as

$$\mathrm{WDD}(\beta, \tilde{\beta}) = \Delta\beta^{\mathrm{T}} \cdot \frac{w}{tr[w]} \cdot \Delta\beta \times 100 \tag{15}$$

where β and $\tilde{\beta}$ denote the diagnostic features vector of the original and reconstructed signal, respectively. $\Delta\beta$ is the normalized difference vector and w is a diagonal weighting matrix. This matrix is used to emphasize certain regions in the ECG complex. Although the WDD measure correlates well with visual inspection, it suffers from high computational complexity mainly due to the accurate evaluation of all diagnostic features and the calculation of optimal weights for the significant features. The non-stationarity of ECG signal and the artifacts may lead to a false classification of diagnostic features. The source error due to the classification and the comparison of irregularity of the wave shapes may degrade the accuracy of the WDD measure.

In [23], nine diagnostic parameters such as amplitudes and durations are extracted within each ECG cycle and then, the average absolute error (AAE) is calculated without giving any weight to the local waves. The AAE is defined as

$$\mathrm{AAE} = \frac{1}{K}\sum_{i=1}^{K}\lambda_i \tag{16}$$

where λ_i is the normalized absolute error for ith diagnostic parameter and K is the number of diagnostic parameters.

A new Wavelet based quality measure, WWPRD, is proposed in [24]. The measure is based on decomposition of the segment of interest into subbands and the weighted score is given to the band depending on the dynamic range and its diagnostic significance. The PRD measure is used as error measure for each band, which is called as WPRD. The Wavelet based weighted PRD (WWPRD) is given as

$$\mathrm{WWPRD} = \sum_{j=1}^{L+1} w_j \,\mathrm{WPRD}_j \tag{17}$$

where, L is the number of decomposition level, w_j the weight of the jth subband and WPRD_j is the PRD value of the jth subband. In WWPRD measure, the weight of each subbands is calculated as the ratio of sum of the absolute value of coefficients within that band and the sum of absolute value of Wavelet coefficients in all the bands. Qualitative and statistical analysis have proved that the WWPRD is well correlated with clinically evaluated results.

In general the compression algorithms do not have any prefiltering process to remove the noise in the original signal. ECG compression is directly performed on the raw signal. In transformed compression algorithms, the high frequency noise component is smoothed in the reconstruction. This will produce a large Wavelet PRD value because the amplitude of the high frequency component is very small. The calculated weight value for this subband is also small. If the WWPRD is used as measure for the compression error between the reconstructed and the original noisy signal, it gives high values due to the presence of noise. But this may not reflect the actual distortion in the reconstruction. Finally, the contribution of the insignificant error of some subbands are more in WWPRD measure. Consequently, this makes the signal to fall into wrong signal quality group. The measure will increase the overlapping of the quality groups and leads to confusion in the judgement of the reconstructed signal quality.

Recently, various efficient Wavelet based ECG compression algorithms are proposed for ECG signals. For evaluating such algorithms, a quality measure based on Wavelet transform that has a good correlation with the diagnostic information loss will be very useful. In this paper, a modified Wavelet based diagnostic distortion measure, WEDD, is proposed where the dynamic weight of each subband is calculated from the energy of the Wavelet coefficients instead of absolute value of coefficients within that band. The weights are dynamic in nature because these are determined from the Wavelet coefficients of the non-stationary ECG signal whose PQRST complexes have a time varying characteristics. These weights make the error value of the diagnostic features more significant at the same time it minimizes the error value introduced by the presence of noise in the original signals. Under noisy conditions the WEDD error measure provides relevant diagnostic distortion whereas the WWPRD measure fails to do so. In this work, the distortion in the local diagnostic features are analyzed using Wavelet transform. The ECG feature extraction technique using Wavelet transform is established in [2–4]. The concept of energy compaction for an ECG compression is discussed in [5,6].

The rest of the paper is organized as follows. The proposed Wavelet Energy based diagnostic distortion measure is presented in Section 3. In Section 4, qualitative and quantitative measures for evaluating the performance of the objective distortion measures are performed. Finally, in Section 5, summary and conclusions are drawn.

3. Proposed Wavelet energy based diagnostic distortion (WEDD) measure

Recording and analysis of electrical potentials of the heart is extensively used in diagnosis of the heart function. The basic characteristic features of an ECG signal are the P, Q, R, S, T and U waves. From the study of the heart's electrical conduction systems, QRS complex is the most significant component. It is characterized by sharp slopes in ECG signal [26]. QRS complex has strong energy distribution which is mostly between 1 and 40 Hz, and even more if the Q, R, and S waves have very sharp morphologies. QRS complex originating from an impulse in the SA node will have a sharp morphology and significant energy spectrum in the higher frequency band. QRS complex resulting from an ectopic site in the ventricles will have a rounded morphology and its stronger energy will be in the lower spectrum of the ECG band. The P and T waves have significant proportion of their energy only up to 10 Hz. There is also an overlapping problem of the P and T waves' spectral components [27].

The Wavelet transform describes signals in terms of coefficients, representing their energy content in a specified

time–frequency region. The signal $f(t)$ can be decomposed as [28–32,34]:

$$f(t) = \sum_j \sum_k d_{j,k}\psi_{j,k}(t) = \sum_j f_j(t) \tag{18}$$

where $j,k \in Z$ and $\psi(t)$ is a mother Wavelet. The Wavelet coefficients $d_{j,k}$ is the inner product

$$d_{j,k} = \langle f(t), \psi_{j,k}(t)\rangle = \frac{1}{\sqrt{2^j}} \int f(t)\psi(2^{-j}t - k)\,\mathrm{d}t \tag{19}$$

From the Wavelet coefficients $d_{j,k}$, the energy of the details of f at level j can be expressed as:

$$E_j = \sum_k d_{j,k}^2 \tag{20}$$

If the total energy of the details is denoted as $E_t = \sum_j E_j$, then the percentile energy corresponding at level j is:

$$\varepsilon_j = \frac{E_j}{E_t} \times 100 \tag{21}$$

The level j is associated with frequency band, ΔF, given by:

$$2^{-j-1}F_s \le \Delta F \le 2^{-j}F_s \tag{22}$$

where F_s is the sampling frequency and $j = 1, 2, 3, \ldots,$ The Wavelets are band-pass functions with zero dc-components and constructed from the mother Wavelet $\psi(t)$. ECG signals are generally non-stationary and their local waves are often spatially variant. By decomposing a signal into the elementary building blocks that are well localized both in time and frequency, the Wavelet transform is an efficient tool for characterizing local features of the signals [34]. It is obvious that the WT at small scales reflects the high frequency components of the signal and, at large scales reflects the low frequency components of the signal. The WT at small scales reflects the high frequency components of the signal and at large scales reflects the low frequency components of the signal. The WT allows the representation of the temporal features of a signal at different resolution [4].

3.1. Wavelet transform of ECG

The generic form of one-dimensional biorthogonal discrete Wavelet transform is shown in Fig. 3. A two channel perfect reconstruction filter bank, which is also known as a biorthogonal filter bank, consists of an analysis filter bank combined with downsampling operators and synthesis filter bank combined with upsampling operators. In Fig. 3, $H_o(z)$, $G_o(z)$ are analysis and synthesis low-pass filters, respectively while $H_1(z)$, $G_1(z)$ are analysis and synthesis high-pass filters, respectively. The Wavelet coefficients are obtained by convolution with the one-dimensional filters $H_o(z)$ and $H_1(z)$ followed by decimation (Fig. 3). The reconstruction is performed by upsampling and convolution with filters $G_o(z)$ and $G_1(z)$. The resulting transformed ECG signal consists of signal components at different resolution levels. The four filters are related in Z-domain as follows:

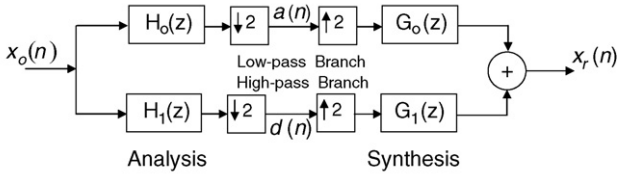

Fig. 3. Convolution structure of a discrete Wavelet transform (DWT): two channel, perfect reconstruction filter bank (analysis and synthesis).

$$G_o(z) = H_1(-z) \tag{23}$$

$$G_1(z) = -H_o(-z) \tag{24}$$

$$H_o(z)H_1(-z) - H_1(z)H_o(z) = 2 \tag{25}$$

The first two conditions eliminate the aliasing and the last one eliminates the phase and amplitude distortion in the reconstructed signal. The rule of selecting proper decomposition level depends on the characteristics of ECG signal. Each level corresponds to its frequency range. The decomposition levels are fixed based on the level which will be more sensitive to the change of amplitude. The number of decomposition levels is calculated from the formula (10) in [24]. Here, five-level decomposition structure is chosen because it is sufficient to cover the spectrum of the ECG signal from the `mita` and `mitect` database and was also successively employed for ECG compression [5,6,18]. For five-level decomposition, the coarse approximation $a(n)$ passes through the filter bank whose analysis stage is identical to that of Fig. 3. The DWT of the original signal is then obtained by concatenating all the Wavelet coefficients, $a(n)$ and $d(n)$, starting from the last level of the decomposition.

3.2. Energy based weighted Wavelet diagnostic distortion measure

The mean value of the original ECG signal and the reconstructed signal is measured and then subtracted from them for diagnostic distortion measure purpose. Both signals are decomposed using Daubechies 9/7 biorthogonal Wavelet filters [31] up to L level. The decomposition bands are the approximation band A_L at level L, which is given by the scaling function and the detail bands D_L, D_{L-1}, $D_{L-2}, \ldots, D_1$, which are given by the Wavelets. Low levels contain high-frequency information about the signal and high levels contain low-frequency information about the signal. In this work, the five-level decomposition structure is used for the decomposition of the ECG signals. This level is sufficient for the ECG signal sampled between 250 and 360 Hz [24]. Basically, the Q and S waves are high frequency and low amplitude waves and their energies are mainly at small scales (highest frequency bands). The first scale contains high-frequency noise and the QRS complex. The QRS complex in the ECG signal usually has the largest amplitude and the widest spectrum and therefore the QRS complex is visible at all WT scales. However, it is most prominent in the second and third scales. The P and T waves are not visible at the first two WT scales. The energies of T and P waves are mainly at scales of four and five.

According to formula (22), for the sampling frequency of 360 Hz, the frequency bands A_5, D_5, D_4 and D_3 occupies the region of frequencies ranging from 0 to 45 Hz. According to the power spectra of the ECG signal, noise and artifact [27], this frequency range has most of the energy of the local waves of the ECG. Fig. 4 shows the T-wave, U-wave, P-wave, S-wave, Q-wave and different QRS complex morphologies, together with details band (D_1–D_5) and approximation band A_5. It can be observed that small scales reflect the high frequency components of the signal and large scales reflect the low frequency components of the signal. The effect of high-frequency artifact can be seen in bands D_1 and D_2. It can be observed that the small Q and S wave energy is mostly between bands D_1 and D_2. These bands are weighted with small value because the energy contribution to the spectrum is low. In most cases the weight of the band D_1 is less than the weight given to the band D_2. The band D_3 consists of high frequency portion (spikes) of the QRS complex and its weight is higher than the other high frequency bands. The band D_4 consists of most of the QRS complex contribution and low percentage of the P-wave. The weight of the band D_4 depends on the amplitude of the QRS complex. The band D_5 have large contribution of the P-wave and part of the T-wave and QRS complex. The weight of the band D_5 is mostly maintained by the P-wave and low percentage of the T wave and QRS complex. Finally, the approximation band A_5 contains large contribution of the T-wave and some contribution of the P-wave and its weight includes the significance of the T and P waves and also the shape of the waves.

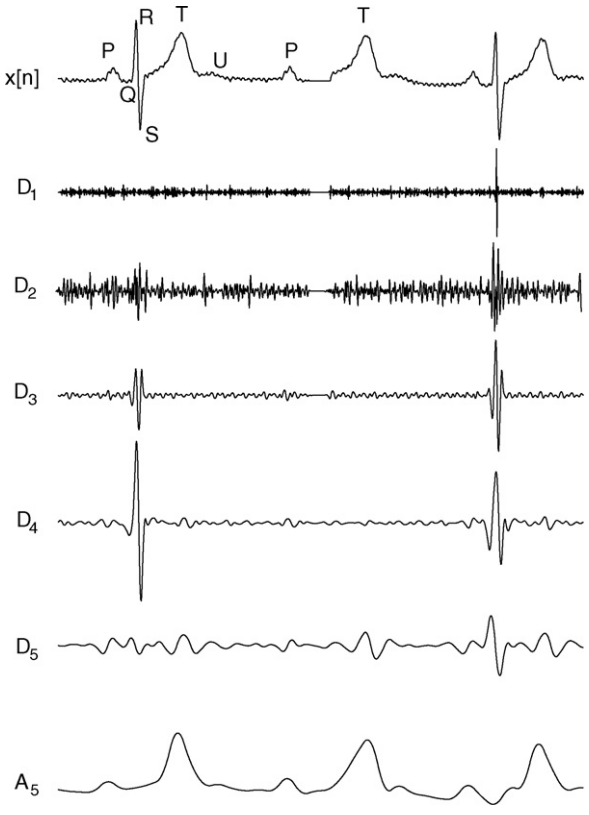

Fig. 4. Wavelet transform at the first five scales of ECG-like simulated wave. Here, ECG wave is extracted from the mita record 117 for testing purpose.

3.2.1. Weight calculation

Above observations suggest that the weights can either be assigned based on the significance of the diagnostic features or they can be objectively determined from the Wavelet coefficients. The first case reflects poor diagnostic distortion measure when different abnormal signals are analyzed. In the second case, the weights are dependent on the amplitude and shape of the local waves in the signal. The dynamic weight calculation was proposed in [24]. The weights, which were assigned for the band D_1 and D_2, were inefficient to bring the insignificant error to low value. The calculated WWPRD value was high even when the diagnostic features are preserved exactly. In most cases, the resulted WWPRD by the band D_1 and D_2 are high due to the presence of noise. In [24], the total error is calculated by adding all the weighted errors which are contributed by the band A_5 and from D_5 to D_1. This may mislead the judgement of the signal quality when the signal contains more noise. In most compression algorithms, there is no prefiltering to eliminate the noise. The above mentioned problem is rectified by the proposed weight which is calculated based on the energy contribution. The percentile energy contribution of the details is given in equation (21). The proposed weight can be expressed as:

$$w_l = \frac{\sum_{k=1}^{K_l} d_l^2(k)}{\sum_{m=1}^{L+1}\sum_{k=1}^{K_l} d_m^2(k)}, \quad l = 1, 2, \ldots, (L+1) \tag{26}$$

where w_l is the weight for lth subband, K_l denotes the number of Wavelet coefficients in lth subband, $d_l(k)$ the kth Wavelet coefficient of the original signal in lth subband and L is the decomposition level. The denominator of Eq. (26) indicates the sum of the energy of each subband. The determined weights are dynamic in nature because they are dependent on the dynamic range of the ECG signal.

The error between the Wavelet coefficients of the original signal and Wavelet coefficients of the reconstructed signal is calculated by percentage root mean square difference, which is referred to as Wavelet PRD. The WPRD is defined as

$$\mathrm{WPRD}_l = \sqrt{\frac{\sum_{k=1}^{K_l}[d_l(k) - \tilde{d}_l(k)]^2}{\sum_{k=1}^{K_l}[d_l(k)]^2}}, \quad l = 1, 2, 3, \ldots, (L+1) \tag{27}$$

where WPRD_l is the error in the lth subband, $d_l(k)$ the kth Wavelet coefficient of the original signal in lth subband and $\tilde{d}_l(k)$ is the kth Wavelet coefficient of the reconstructed signal in lth subband. This measure gives the idea about the structured error estimation since it calculates the error between the local waves of the ECG signal. It will localize the diagnostic feature distortion estimation for reconstructed signals rather than the distributed error within the ECG beat. The WPRD value is high in higher bands. Hence, the dynamic weight is important to focus the local error significantly for checking the diagnostic quality of the reconstructed signal. The WEDD is calculated

using (26) and (14), given by

$$\mathrm{WEDD}_l = \sum_{l=1}^{L+1} w_l \mathrm{WPRD}_l \tag{28}$$

3.3. Preliminary evaluation of the proposed WEDD measure

For a preliminary evaluation of the proposed WEDD measure, some simple tests are performed by applying the measure to three different types of distorted ECG signals with predictable results: (i) ECG signal reconstruction after zeroing specified Wavelet band coefficients; (ii) ECG with noise; (iii) ECG with isoelectric region distortions. These types of distortions are considered because these occurs during the zeroing of samples/coefficients based on threshold value in ECG compression algorithms, the improper quantization of samples/coefficients with small amplitude and the ECG signal transmission.

3.3.1. Zeroing of Wavelet coefficients

This experiment is carried out here because recently reported efficient Wavelet based compression algorithms [5,6,9,10,13,19,21] uses thresholding process to retained a specified number of Wavelet coefficients while zeroing other Wavelet coefficients. The zeroing Wavelet coefficients may lead to diagnostic features distortion. The reflection of local errors due to the zeroing of Wavelet coefficients of particular subband is performed using the `mita` record 103. The original signal of the tested ECG segment is shown in Fig. 5 (a). The reconstructed signals after zeroing the detail bands D_1, D_2, D_3, D_4, D_5, the approximation band A_5 and the combined detail bands D_1 and D_2 are shown in Fig. 5(b–h). The WWPRD [24] and the proposed WEDD for each band are measured. The PRD1, MAX and NCC are also measured. The value of WEDD for each band, WWPRD, PRD1, MAX and NCC values are summarized in Table 1. From the results presented in Table 1, it can be observed that all measures detect differences from the original and zeroed band signals. The WEDD value is low for the zeroed detail band D_1 and D_2 case whereas WWPRD values are large. The reconstructed signal is shown in Fig. 5(b and c), respectively. It can be seen that the reconstructed signal qualities are good. The error in these bands may be referred to as insignificant error since it does not distort the diagnostic features except small amplitude reduction in the QRS complex. Zeroing the band D_3 slightly distorts the QRS complex, which is shown in Fig. 5(d). The QRS complex is severely distorted when the band D_4 is zeroed and results in high WEDD value of 44.7802%, as shown in Fig. 5(e). The T-wave becomes nearly flat line when the A_5 coefficients are zeroed and its effect is shown in Table 1 and Fig. 5(g). WEDD values in eight column of Table 1 shows that the WEDD measure emphasizes the value of the significant error with respect to the amplitude of the features in time domain ECG signal while it diminishes the insignificant error. The above experimental results of the

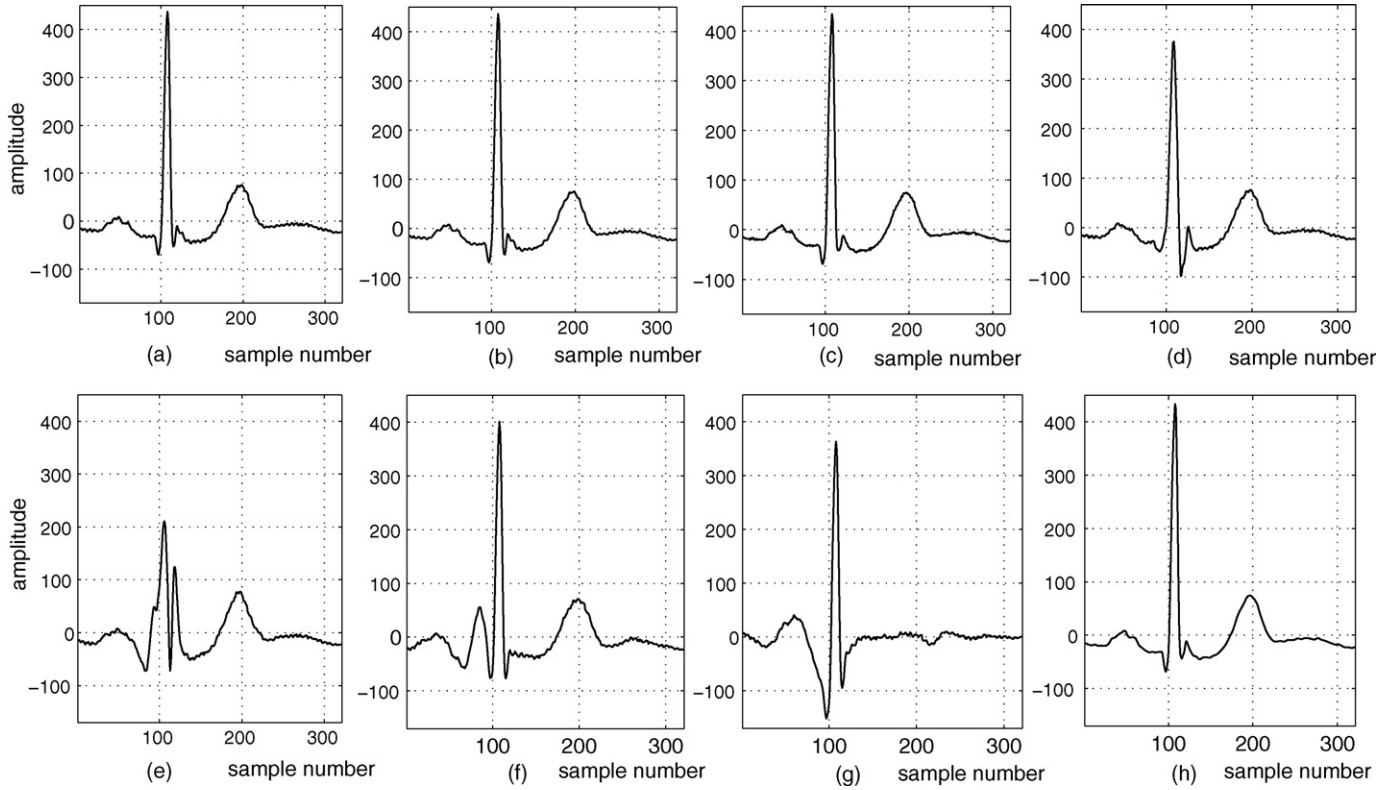

Fig. 5. (a) Original signal from the `mita` record 103. Reconstructed signal after zeroing Wavelet coefficients in (b) detail band (D_1), (c) detail band (D_2), (d) detail band (D_3), (e) detail band (D_4), (f) detail band (D_5), (g) approximation band (A_5) and (h) detail bands (D_1 and D_2).

Table 1
Performance of the proposed measure when a particular subband is zeroed

Zeroed band	WEDD: weighted PRD (%) of each band							WWPRD [24]: weighted PRD (%) of each band							PRD1 (%) [18]	MAX [17]	NCC [21]
	A_5	D_5	D_4	D_3	D_2	D_1	Total	A_5	D_5	D_4	D_3	D_2	D_1	Total			
D_1	0.0258	0.0039	0.0283	0.0411	0.0035	0.0312	0.134	0.032	0.005	0.013	0.0622	0.1331	3.745	3.991	1.727	0.032	0.9999
D_2	0.3355	0.0343	0.0363	0.0138	0.1202	0.0019	0.542	0.417	0.044	0.0167	0.0209	4.5196	0.224	5.242	3.553	0.084	0.9994
D_3	0.6075	0.0401	0.0805	8.5993	0.0048	0.0003	9.333	0.7551	0.0515	0.0371	12.999	0.1794	0.0323	14.055	32.037	0.781	0.9473
D_4	0.3662	0.0335	44.780	0.0176	0.0007	0.0001	45.198	0.4551	0.0429	20.617	0.0266	0.0277	0.0071	21.176	68.360	1.38	0.7304
D_5	3.762	8.748	0.1498	0.0418	0.0015	0.0001	12.703	4.6759	11.215	0.069	0.0632	0.0572	0.0145	16.095	31.533	0.430	0.9491
A_5	38.845	0.1377	0.1314	0.0414	0.0015	0.0001	39.157	48.282	0.1765	0.0605	0.0626	0.0555	0.013	48.650	53.977	0.451	0.8418
D_1, D_2	0.3112	0.0349	0.061	0.0458	0.120	0.0313	0.6044	0.3869	0.0447	0.0281	0.0692	4.524	3.7549	8.808	3.936	0.078	0.9992

proposed WEDD measure proves that the error in the diagnostic features is reflected in their subbands, respectively. Whereas in the WWPRD measure, the error value is high for the insignificant features. Hence it may not reflect the diagnostic distortion and leads to confusion in the quality judgement with the usage of measured WWPRD value. Similarly the PRD1, MAX, and NCC measure gives the value of global error, local error and similarity, respectively. Since global error measures are only an average measure, it alone cannot adequately quantify the performance of an ECG compression algorithm. These measures may not give which diagnostic feature is severely distorted even when the measure is applied block-wise, since the local waves of ECG are spatial variant. The proposed WEDD measure provides more substantive results than MSE based measures, since it provides errors of the local waves. And also these local errors are weighted dynamically based on the amplitude and shape of the ECG waves in the original signal. Thus, the proposed WEDD measure is more effective for assessing the quality of the diagnostic features in the reconstructed signal.

3.3.2. Proposed WEDD and other distortion measures under noisy condition

In many situations, the ECG signal contains noise which is not important for clinical diagnosis but it provides a nonzero value in all the above mentioned error measures including WWPRD [24]. In WEDD measure, the error introduced by the noise is weakened by the proposed Wavelet energy based weights, which are very small. These problems are considered and illustrated with the following example. The ECG segments from the noisy `mita` records 100, 108, 117 and 123 are compressed/decompressed by [6] and the reconstructed signal distortion is computed using non-diagnostic and diagnostic distortion measures. But for the testing purpose, the additive white gaussian noise of −30 dB is introduced in isoelectric region of the ECG segment from the `mita` record 123. Here, other `mita` records 100, 108 and 117 are processed with their noise level in the records. The Wavelet decomposition of these records have high frequency noise components in subbands D_2 and D_1. These subbands Wavelet coefficients are zeroed when the thresholding process is employed in the compression algorithms. Now, a distortion measure is needed to measure the amount of diagnostic feature distortion without the inclusion of insignificant error. The measured objective error values using non-diagnostic and diagnostic distortion measures are summarized in Table 2. For the visual inspection, the reconstructed signals are shown in Fig. 6 (b), (d), (f) and (h). Table 2 illustrates the actual behavior of local and global error measures. The local errors introduced by the approximation band A_5 and the detail bands D_5- D_1 are computed using the WWPRD [24] and the proposed WEDD measure for the effective comparison. In WWPRD, the error contributed to the total error value by the bands D_1 and D_2 is high compared to other Wavelet bands value given in the table. These two bands provide high error value due to the smoothing of background noise. Hence, the WWPRD measure may not reflect the local wave distortion. For all the tested ECG segments, the resulting WWPRD value is very high. These values indicate that some

Table 2
Performance of the proposed measure when the signal with noise and distortion in isoelectric region

Measure	Record	Weighted PRD (%) of each band							Non-diagnostic measures: for mita records 100, 108, 117 and 123, respectively		
		A_5	D_5	D_4	D_3	D_2	D_1	Total	PRD1	MAX	NCC
WWPRD [24]	100	1.1121	0.247	0.8651	1.4829	3.5722	5.2015	12.4808			
	108	1.6623	0.4749	1.6152	3.0294	8.0238	6.5371	21.3427			
	117	1.5995	0.5718	0.8029	5.1285	2.9298	4.5792	15.6117	4.983	0.0279	0.9988
	123	0.9925	0.7628	1.0169	6.0858	6.1846	12.7361	27.7787	7.201	0.0387	0.9974
Proposed WEDD	100	0.3845	0.5753	1.331	0.7529	0.6117	0.0756	3.7311	5.731	0.0557	0.9984
	108	2.2423	0.7283	0.3058	0.253	0.3773	0.096	4.0026	8.3533	0.1011	0.9965
	117	1.9963	0.4777	1.125	0.2772	0.0623	0.0419	3.9804	–	–	–
	123	0.6399	1.1024	2.8355	0.7264	0.3853	0.3794	6.069	–	–	–

Note: Additive white gaussian noise (AWGN) of −30 dB is introduced in isoelectric region of mita record 123. Other mita records 100, 108, and 117 are processed with their noise level in their respective records.

diagnostic features are severely distorted within ECG segment of the tested records. But the diagnostic features are preserved well, which is shown in Fig. 6. But the proposed measure provides low error value for these two bands. Consequently, the total WEDD exactly reflects the amount of diagnostic distortion. For example, the WWPRD value for the tested mita record 108 is 21.3427% while the WEDD value is 4.0026%. The diagnostic features are preserved in reconstructed signal of mita record 108, which are shown in Fig. 6(d). The WEDD for the test record 123 is 6.069%, which indicates that some of the diagnostic features are distorted. From Fig. 6 and Table 2, it can be observed that the tabulated WEDD values indicate the amount of diagnostic features reconstruction error. Consequently, the judgement of the reconstructed signal quality is better using the estimated WEDD value than the WWPRD value under noisy condition. It can be also observed that the proposed WEDD measure is more sensitive to diagnostic features distortion while it is less sensitive to distortion in the isoelectric regions and the presence of noise as compared to other error measures such as PRD1, MAX, NCC and WWPRD. The above experiments show that the proposed measure performs quite well for a wide range of distortion types.

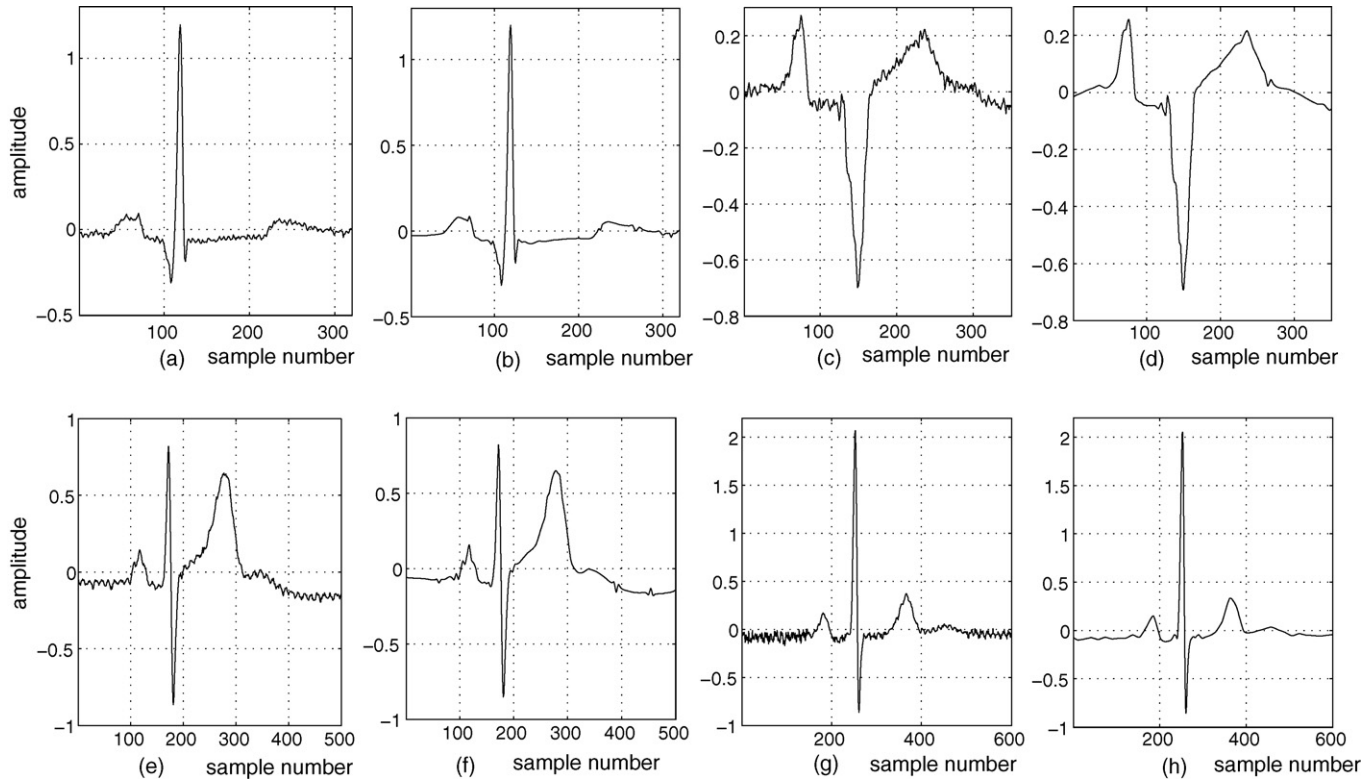

Fig. 6. Performance of the proposed WEDD and other error measures for the noisy test signal. Original signal: (a, c, e and g) are from the mita record 100, 108, 117 and 123, respectively. Reconstructed signal: (b, d, f) and (h) of the compressed mita record 100, 108, 117 and 123, respectively. Here, the raw ECG signal of 100, 108, and 117 are processed. And the additive white gaussian noise of −30 dB is introduced in isoelectric region of ECG signal from record 123 for testing purpose.

3.4. Mean opinion score (MOS) test

In order to analyze the proposed WEDD measure with reference to subjective quality test, the MOS test is performed in this section. Subjective test based distortion measure is used as reference because it represents the true quality of the reconstructed signal (gold standard) [22]. These tests can be visually performed by the cardiologist and the biomedical signal processing group by reviewing the diagnostic features of the original and reconstructed one. In order to find a qualitative distortion measure for each of the tested signals, MOS test was performed, which consists of a blind and a semi-blind tests [22]. The blind test analyzes both the original and reconstructed signals individually and the different ECG signal features are interpreted by the observers. The semi-blind test compares the original and reconstructed signals together. In this test, the observer provides an evaluation of the similarity between the original and reconstructed signal features. In this paper, semi-blind test is conducted for the evaluation of MOS rating. Here, the MOS rating is defined according to the quality of the diagnostic features P-wave, PR-segment, PR-interval, QRS complex, T-Wave, ST-segment and QT-interval for every tested signal. The quality rating ranges from 1 (bad) to 5 (excellent) for each tested signal. The observer examines a set of ECG features and assigns a numerical rank or classification for signal under test. All such scores are recorded. The average rank of each ECG segment is later determined. This average rating is known as a mean opinion score (MOS). The MOS^{s} of one ECG segment from the ith cardiologist is given by

$$\mathrm{MOS}^{(s)}(i) = \frac{1}{N_f}\sum_{n=1}^{N_f} R(n) \tag{29}$$

where $R(n)$ is the rating of the nth diagnostic features and N_f is the number of diagnostic features which is seven. The overall MOS for the tested signals is the average MOS of all the segments of the signal. The MOS of one segment from all N_c cardiologists is given by

$$\mathrm{MOS} = \frac{1}{N_c}\sum_{i=1}^{N_c} \mathrm{MOS}^{s}(i) \tag{30}$$

Using Eq. (30), $\mathrm{MOS}_{\mathrm{error}}$ for the semi-blind test of one segment from all N_c cardiologists is calculated as

$$\mathrm{MOS}_{\mathrm{error}} = \frac{5 - \mathrm{MOS}}{5} \times 100 \tag{31}$$

In this work, the $\mathrm{MOS}_{\mathrm{error}}$ measure is used as a gold standard [22]. It is necessary to determine a gold standard that can represent the diagnostic truth of each reconstructed signal. A signal that is judged to be of deficient quality would have a high $\mathrm{MOS}_{\mathrm{error}}$, whereas one that is thought to be of good quality would receive a low $\mathrm{MOS}_{\mathrm{error}}$. One can estimate the effectiveness of a metric by comparing the numerical values reported by an ECG quality system to those assigned by the subjective test group. The values of a good metric should be similar to the mean opinion scores. MOS for the tested ECG segment (in Fig. 6(h)) from record 123 is given in Table 3.

Table 3
Semi-blind MOS test applied to `mita` record 123 (Fig. 6(h))

ECG features	Score 1	Score 2	Score 3	Average
P-wave	4	5	4	4.33
PR-segment	4	5	5	4.67
PR-interval	5	5	4	4.67
QRS complex	5	5	4	4.67
ST-segment	5	4	5	4.67
T-wave	4	4	5	4.33
QT-interval	5	5	4	4.67
MOS	4.57	4.71	4.43	4.57

4. Results and discussion

In this section, the performance of the proposed measure is compared with other diagnostic and non-diagnostic distortion measures. The relationship between the subjectively evaluated values and the objectively estimated values is also studied quantitatively and qualitatively.

4.1. Test ECG signal

In order to assess ability of the proposed WEDD measure in predicting the $\mathrm{MOS}_{\mathrm{error}}$ and to compare with WWPRD, PRD and other non-diagnostic distortion measure, the reconstructed signals with different qualities are obtained using five different compression algorithms AZTEC, FAN, DCT, SPIHT [13], DSI [23]. For evaluation purpose, the ECG signals are extracted from the 48 records included in the `mita` database, 08730_01, 08730_02, 08730_05 and 08730_07. These signals are chosen because they consist of a large variety of pathological cases and noisy ECG signals. Also, these signals are used for testing different compression algorithms [5,6,12,13,18,19,23]. The mean is removed from every ECG signal. Using the above mentioned records, 210 reconstructed signals are produced for the evaluation purpose.

4.2. Analysis of correlation between $\mathrm{MOS}_{\mathrm{error}}$ and each distortion measure

For every reconstructed signal, the non-diagnostic distortion measures and diagnostic distortion measures are computed: RMSE, PRD1, PRD2, SNR, MAX, NMAE, CC, WWPRD and the proposed WEDD. To evaluate the objective measures, the gold standard $\mathrm{MOS}_{\mathrm{error}}$ is calculated from the results of the semi-blind test of two cardiologists for every tested signal. To observe the relationship between gold standard and distortion measure, the scatter plots of $\mathrm{MOS}_{\mathrm{error}}$ versus objective distortion measures are shown in Figs. 7–9 . The mean opinion score errors are linearly scaled to a nominal range of [0, 100], where 0 and 100 represent the best and the worst ratings, respectively. The horizontal axis and the vertical axis represents the objective distortion measure and $\mathrm{MOS}_{\mathrm{error}}$, respectively. Each sample point in the scatter plot indicates the error value of

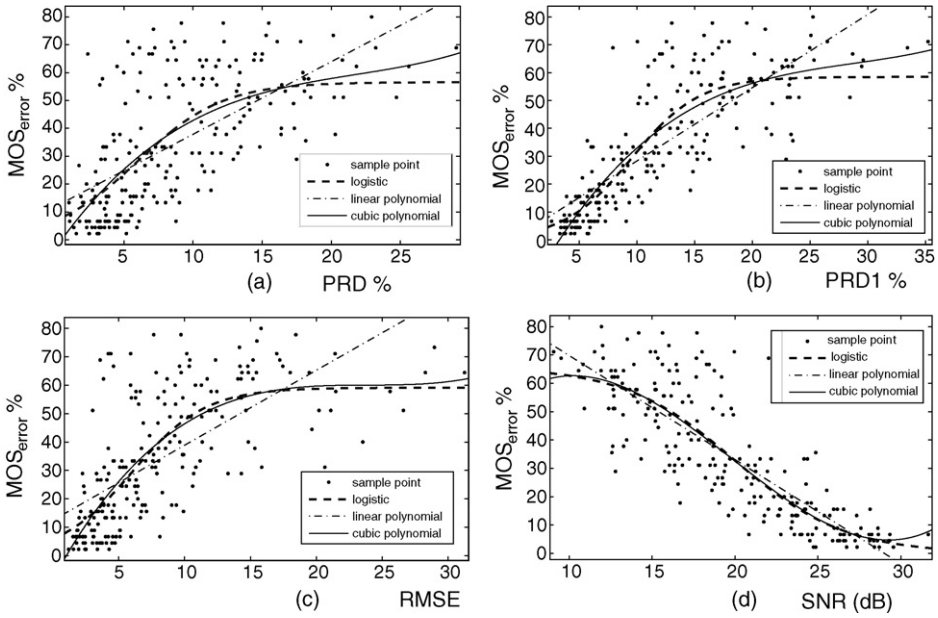

Fig. 7. Relationship between measured gold standard (MOS_{error}) values vs. values estimated via objective non-diagnostic distortion measures. Each point represents one test signal. (a) MOS_{error} vs. PRD2, CC = 0.6342. (b) MOS_{error} vs. PRD1, CC = 0.799. (c) MOS_{error} vs. RMSE, CC = 0.6639. (d) MOS_{error} vs. SNR, CC = −0.8354.

one reconstructed signal. Scatter plot shows at a glance whether a relationship exists between two sets of sample points or not. The scatter plots of the global error measures are shown in Fig. 7. The scatter plot of MOS_{error} versus PRD2 is shown in Fig. 7(a). From the figure, it is observed that the sample points are dispersed over wide space. Hence, forming the strong relationship between subjective/objective scores is difficult in PRD2 case. Fig. 7(b) shows the scatter plot of MOS_{error} versus PRD1. At higher values of PRD1, the points are widely spread. Hence, it results in poor correlation. But at low values of PRD1, the points are closer. Therefore, the PRD1 has good correlation compared to PRD2. The performance of PRD1 is better than PRD2 because it is measured without mean of the original signal. This relationship between two variables is analyzed by linear and nonlinear regression. Fig. 7(c) shows the scatter plot of MOS_{error} versus RMSE. It is observed that the data points are widely spread. In the above plots, the data points make a straight line going from the origin out to high x- and y-values, then the variables MOS_{error} and PRD1/PRD2/RMSE are said to have a positive correlation. Fig. 7(d) shows the scatter plot of MOS_{error} versus SNR. The sample points are not widely distributed. Here, the line goes from a high-value on the y-axis down to a high-value on the x-axis and hence, the variables MOS_{error} and SNR have a negative correlation.

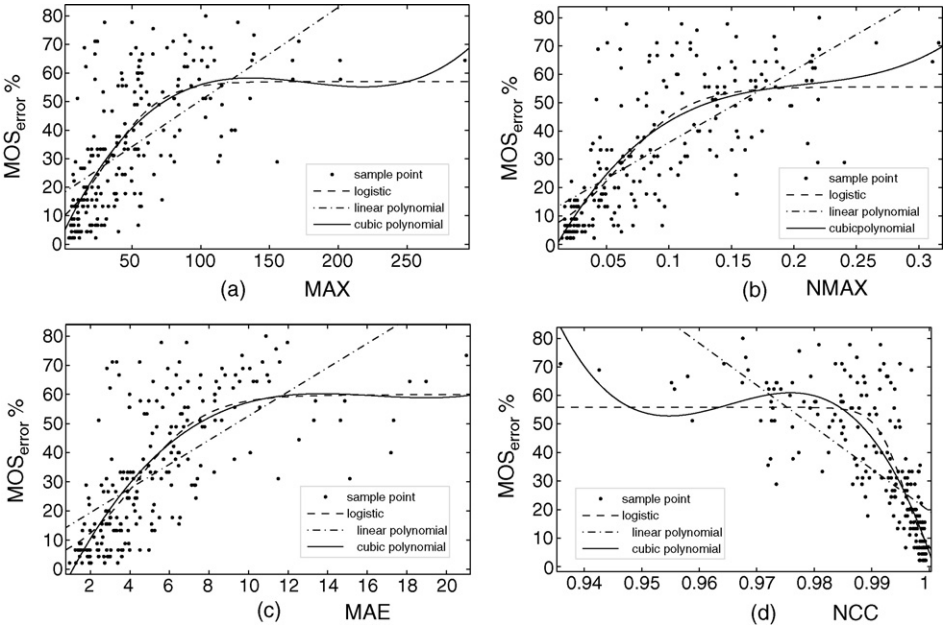

Fig. 8. Relationship between measured gold standard (MOS_{error}) values vs. values estimated via objective non-diagnostic distortion measures. Each point represents one test signal. (a) MOS_{error} vs. MAX, CC = 0.6173. (b) MOS_{error} vs. NMAX, CC = 0.7215. (c) MOS_{error} vs. MAE, CC = 0.647. (d) MOS_{error} vs. NCC, CC = −0.7016.

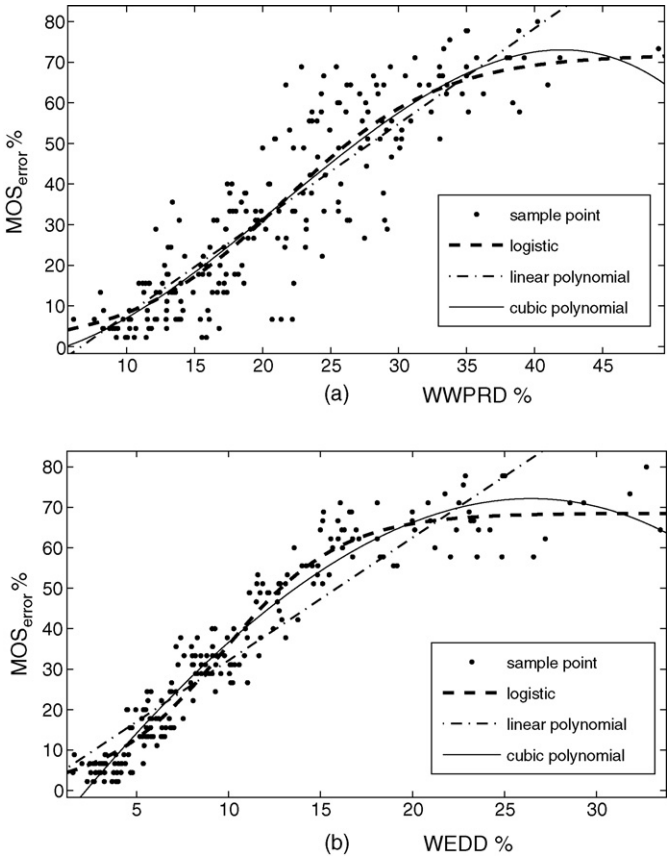

Fig. 9. Relationship between measured gold standard (MOS_{error}) values vs. values estimated via objective diagnostic distortion measures. (a) MOS_{error} vs. WWPRD, CC = 0.8811. (b) MOS_{error} vs. WEDD, CC = 0.9197.

The scatter plots of local error measures (MAX, NMAX), MAE and similarity measure (NCC) are shown in Fig. 8. Fig. 8(a) shows the scatter plot of MOS_{error} versus MAX. Fig. 8(b) shows the scatter plot of MOS_{error} versus NMAX. Fig. 8(c) shows the scatter plot of MOS_{error} versus MAE. From these figures, it is observed that the sample points are dispersed widely. The correlation between the variables is very poor in MAX/NMAX/MAE case. The correlation values are 0.6173, 0.7215 and 0.647, respectively. But these measures have positive correlation since its line goes from a high-value on the y-axis down to a high-value on the x-axis. Fig. 8(d) shows the scatter plot of MOS_{error} versus NCC. The correlation is good at high NCC values while NCC performs poor at low values. In this scatter plot, the line goes from a high-value on the y-axis down to a high-value on the x-axis and hence, the variables MOS_{error} and NCC have a negative correlation. The correlation value of this measure is -0.7016. Fig. 9 shows the scatter plot of MOS_{error} versus diagnostic distortion measures. Fig. 9(a) shows the scatter plot of MOS_{error} versus WWPRD. It is observed that the sample points are not tightly clustered together and is spread more widely along the horizontal direction. Hence it results in poor correlation. Fig. 9(b) shows the scatter plot of MOS_{error} versus proposed WEDD. Here the sample points are tightly packed and the plot resembles a line rising from left to right. The closer the sample points are, the higher is the correlation between the two variables. Since the slope of the line is positive, there is a positive correlation between the two sets of points. The performance of the proposed WEDD measure is better than WWPRD when the noisy signal is compressed and decompressed. However, to evaluate the significance of the proposed measure, both quantitative and qualitative analysis are performed.

4.3. Statistical measures for evaluating the predictability of each distortion measure

In order to assess the ability of the proposed WEDD measure and other measures to predict MOS ratings, the linear and nonlinear predictors are fitted to the various scatter plots. Table 4 gives the linear and nonlinear function which is used to provide a linear and nonlinear mapping between the objective and subjective scores. The linear and cubic polynomials are used as linear predictor which exactly fits the sample points as line. Logistic function is used for the nonlinear mapping. In this work, the model parameters are found numerically using a nonlinear regression process with MATLAB optimization toolbox. After the linear and nonlinear mapping, the following analysis is performed to evaluate the performance of the objective distortion measures.

Firstly, a measure of deviation from the regression line between subjective and objective measures is used for the evaluation of predicted quality compared to gold standard [22]. The regression line expresses the best prediction of the dependent variable (MOS_{error}) for the given independent variables (objective error measures). The deviation of a particular point from the regression line (its predicted value) is called the prediction error. The smaller the variability of the error values around the regression line, the better is the prediction. The VAR [22] measure is given as

$$\mathrm{VAR} = \frac{1}{N_s}\sum_{n=1}^{N_s}[d(n) - \mu_d]^2 \tag{32}$$

where $d(n)$ is the vertical deviation of the nth sample point from the regression line, N_s the total number of sample points and μ_d is the mean value of $d(n)$, $n = 1, 2, 3, \ldots, N_s$.

For each regression line (RL), the variance of the error between the actual and predicted MOS_{error} are summarized in Table 5. The VAR is zero for a good predictor. The VAR is low for the nonlinear predictor (logistic function) compared to polynomial predictor. Thus, the logistic predictor gives the best fit. The variance between the WWPRD and the MOS_{error} is 106.49 and the variance between the WEDD and the MOS_{error} is 31.70, which is three times better than the WWPRD case.

Table 4
Linear and Nonlinear regression function

Type of fit	Regression function
Linear polynomial	$f_1(x) = b_1x + b_0$
Cubic polynomial	$f_2(x) = b_3x^3 + b_2x^2 + b_1x + b_0$
Logistic	$f_3(x) = \frac{b_0}{1+b_1\exp(-x/b_2)}$

Note: Where x is the objective score and b_0, b_1, b_2 and b_3 are the model parameters.

Table 5
Variance (VAR) of the error calculated between distortion measures and their regression lines

Regression line (RL)	Non-diagnostic distortion measure (non-DDM)								DDM-Wavelet weights	
	Global-OM					Local-OM		SM	Absolute w_l (WWPRD [24])	Energy w_l (WEDD)
	RMSE	PRD2	PRD1	SNR	MAE	MAX	NMAX	NCC		
Linear RL	290.26	310.29	187.68	156.85	301.75	321.24	248.89	263.59	116.08	80.00
Cubic RL	238.31	288.71	151.68	150.15	240.67	265.02	208.67	159.26	107.18	35.42
Nonlinear RL	239.20	288.44	150.75	149.21	240.11	213.46	267.78	155.71	106.49	31.70

Note: OM: Obejective measure; SM: Similarity measure.

Secondly, to investigate the performance of the proposed measure, various statistical measures like root mean square prediction error, mean absolute prediction error (MAE), standard error (S.E.) and correlation coefficient (CC) between $\mathrm{MOS}_{\mathrm{error}}$ and each distortion measure are computed. The prediction error is the difference between the true $\mathrm{MOS}_{\mathrm{error}}$ and the predicted $\overline{\mathrm{MOS}}_{\mathrm{error}}$ for an objective error. The regression line seeks to minimize the sum of the squared errors of the prediction. The square root of the average squared error of prediction, standard error, is used to measure the accuracy of predictions made with regression line. The standard error is given by

$$\mathrm{S.E.} = \sqrt{\frac{\sum_{n=1}^{N_s}\left[\mathrm{MOS}_{\mathrm{error}}(n) - \overline{\mathrm{MOS}}_{\mathrm{error}}(n)\right]^2}{N_s - 2}} \tag{33}$$

where N_s is the number of sample points.

4.4. Analysis of prediction accuracy and monotonicity of each measure

Performance of the objective measures is also evaluated with respect to two aspects of their ability to estimate subjective assessment of signal quality. The first measure is the prediction accuracy which gives the ability to predict the subjective quality ratings with low error. The second measure is the prediction monotonicity. The objective measure score should increase and decrease monotonically as the MOS increases and decreases. After the nonlinear mapping, the correlation coefficient between the predicted and true subjective scores is calculated to evaluate prediction accuracy.

The prediction monotonicity is evaluated by Spearman rank order correlation coefficient (SROCC). The monotonic relation is expressed using rank-order numbers instead of the values. As a statistical test to check whether a relation between two variables exists, SROCC test is better than the standard correlation coefficient because the latter will only work when there is a linear relation between the variables. In practical situations, assuming a linear relation is often unrealistic. For more details about these measures, readers can refer to [33].

The statistical evaluation results for all the distortion measures are summarized in Table 6. Using regression lines, the statical measures namely RMSE, MAE, S.E. and CC are calculated for each objective distortion measures. Comparison of the values given in column four (PRD2), five (PRD1) and six

Table 6
Performance comparison of ECG signal distortion measures through the statistical measures

Type of regression line (RL)	Metrics	Non-diagnostic distortion measure (non-DDM)								DDM-Wavelet, w_l	
		Global-OM					Local-OM		SM	WWPRD [24]	WEDD
		RMSE	PRD2	PRD1	SNR	MAE	MAX	NMAX	NCC		
Linear RL (linear polynomial)	M1	16.9979	17.5748	13.6683	12.4953	17.3311	17.8821	15.7403	16.1984	10.7497	8.9238
	M2	13.8897	14.367	10.3427	9.5397	14.3147	14.6177	12.0759	12.9466	8.3673	7.0685
	M3	17.0764	17.6559	13.7315	12.553	17.4111	17.9647	15.813	16.2732	10.7994	8.9651
	M4	0.6639	0.6342	0.799	0.8354	0.647	0.6173	0.7215	0.7016	0.8811	0.9197
Linear RL (cubic polynomial)	M1	15.4021	16.9526	12.2876	12.2256	15.478	16.2421	14.4123	12.5912	10.3293	5.9385
	M2	11.756	13.7934	9.2926	9.1102	12.0152	12.5307	10.7778	9.4747	7.9886	4.8341
	M3	15.4375	16.9916	12.3159	12.2537	15.5136	16.2795	14.4454	12.6201	10.3531	5.9521
	M4	0.7354	0.6662	0.8413	0.843	0.7324	0.6996	0.7733	0.8326	0.8908	0.9653
Nonlinear RL (logistic)	M1	15.4309	16.9448	12.2503	12.1874	15.4603	16.3268	14.5779	12.4512	10.2963	5.6198
	M2	11.7827	13.7173	9.2793	9.1423	11.8726	12.661	11.0407	9.552	8.0041	4.5321
	M3	15.5022	17.023	12.3069	12.2437	15.5317	16.4023	14.6452	12.5087	10.3439	5.6458
	M4	0.7343	0.6666	0.8424	0.8441	0.7331	0.6959	0.7674	0.8368	0.8916	0.969
SROCC		0.7433	0.68	0.8504	−0.8486	0.7439	0.7129	0.7799	−0.8479	0.8865	0.9624

Note: M1: Root mean squared prediction error; M2: Mean absolute prediction error; M3: Standard error; M4: Correlation coefficient; SROCC: Spearman rank-order correlation coefficient.

(SNR) of Table 6 shows that the performance of the PRD1 measure is better than the PRD2. But the SNR performance is better than PRD1 and PRD2 case. The numbers in M4 row of Table 6 have the following meanings. M4 is the correlation coefficient between the actual and predicted MOS values. Perfect prediction would yield the value CC = 1. The correlation coefficient of WWPRD is 0.8916, which is slightly better than PRD1 with CC of 0.8424. But it results in poor prediction for the noisy input signal. For the same test signal, the CC of WEDD is 0.969, which shows better prediction accuracy.

The row (M3) represents the standard error. The standard error should be zero for a good prediction. The S.E. between the actual MOS_{error} and predicted MOS_{error} from WEDD is 5.6458 and from WWPRD is 10.3439. The performance of WEDD is approximately two times better than WWPRD.

In order to justify the effectiveness of the proposed WEDD measure, the rank-order correlation between the WEDD and MOS_{error} is examined and it is compared with the rank-order correlation between the WWPRD and the MOS_{error}. The SROCC is 0.9624 for the proposed measure, which is better than the WWPRD with SROCC value of 0.8865. These results prove that the subjective quality scores obtained from the proposed WEDD measure gives good prediction accuracy (higher correlation coefficient) and better prediction monotonicity (higher Spearman rank-order correlation coefficient) than PRD1, which is most widely used in the ECG compression methods.

4.5. Qualitative analysis of the proposed WEDD, WWPRD and PRD1 measure

A rough classification of signal quality is defined by dividing the MOS_{error} into five quality groups. The quality groups are: excellent, very good, good, not bad and bad. These quality groups and classified MOS_{error} ranges are shown in Table 7. This classification may suggest a reasonable means of evaluating the measures and comparing its findings to those of an independent observer. In this work, the methods followed in [22,24] for the determination of thresholds of each prediction range is performed. For the proposed measure, five prediction ranges are determined. Fig. 10 shows the PRD1, WWPRD and WEDD values of each quality group of signals and the prediction ranges. The horizontal axis and the vertical axis represent the quality groups and objective distortion measure, respectively. Each sample point in the scatter plot indicates the quality of the reconstructed signal. It is observed that the overlapping of quality groups in Fig. 10(a and b) is more compared to quality groups in Fig. 10(c). Overlapping of quality groups leads to confusion. For PRD1 and WWPRD measure, the prediction ranges given in [24] is used in this paper. The thresholds of each prediction range are determined manually making sure that all the signals in the prediction range falls within a minimum number of quality groups.

Table 7
Quality groups defined by MOS_{error}

MOS_{error} (%)	ECG signal quality
0–10	Excellent
10–25	Very good
25–40	Good
40–55	Not bad
>55	Bad

The prediction ranges of the PRD1, WWPRD, and WEDD measure and their quality groups are depicted in Table 8. This is referred to as confusion matrix [22]. The number (r_{mn}) inside

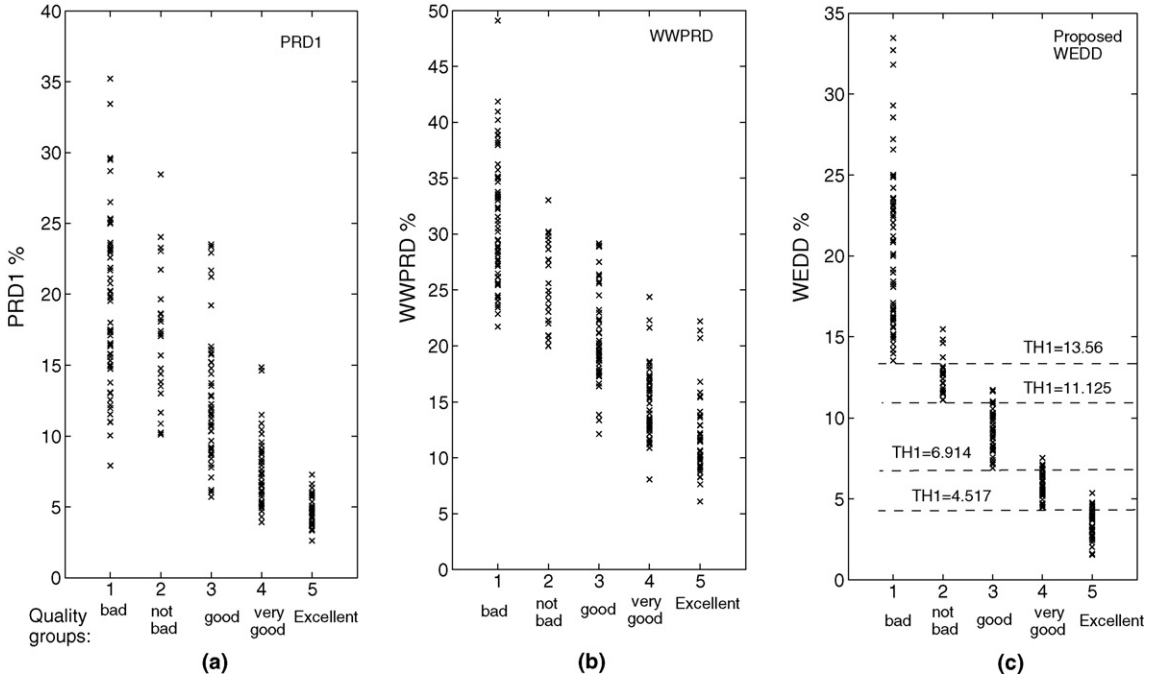

Fig. 10. The scattering and threshold values of the prediction range for each quality group of reconstructed signals. (a) for the PRD1 measure, (b) for the WWPRD measure, (c) for the proposed WEDD measure.

Table 8
Quality groups and their prediction range of PRD1, WWPRD and WEDD measure

Distortion measure	Prediction range	Quality groups					MCP (%)	NPE (%)
		Excellent	Very good	Good	Not bad	Bad		
PRD1 (%)	0–4.33	89	11	0	0	0		
	4.33–7.8	43	48	9	0	0		
	7.8–11.59	0	38	41	8	13	54.4	7.22
	11.59–22.57	0	3	28	22	47		
	>22.57	0	0	12	16	72		
WWPRD [24] (%)	0–7.4	100	0	0	0	0		
	7.4–15.45	55	39	6	0	0		
	15.45–25.18	6	27	42	14	11	50.2	7.5
	25.18–37.40	0	0	15	20	65		
	>37.40	0	0	0	50	50		
Proposed WEDD (%)	0–4.517	100	0	0	0	0		
	4.517–6.914	2	98	0	0	0		
	6.914–11.125	0	7	93	0	0	95	0.6876
	11.125–13.56	0	0	9	91	0		
	>13.56	0	0	0	7	93		

MCP: Mean correct prediction; NPE: Normalized prediction error.

the 5×5 confusion matrix indicates the percentage of signals from the mth prediction range and the nth quality group. Table 8 also demonstrates the prediction ability of the three distortion measures. The diagonal element of the confusion matrix indicates the correct prediction by the distortion measure. From Table 8, it is observed that the determination of quality group is difficult when the tested signal has a PRD1 value between 7.8% and 11.59% or a WWPRD value between 15.45% and 25.18%. But a WEDD value between 6.914% and 11.125% predicts a quality group of very good or good. The mean correct prediction (MCP) and the normalized prediction error (NPE) are computed as defined in [22]. The mean correct prediction of the proposed measure is 95%, which is higher than the measures PRD1 and WWPRD with an MCP of 54.4% and 50.2%, respectively. The normalized prediction error of the proposed measure is 0.6876%, which is ten times less than the PRD1 and WWPRD case.

5. Conclusion

In this paper, a novel objective distortion measure is proposed for evaluating the reconstructed signals of various ECG compression methods. Subjective measures and several objective quality measures in time and frequency domain are discussed. The subjective measure used is based on semi-blind test on reproduced local waves within an ECG cycle. The relationship between the subjectively evaluated values and the objectively estimated values is studied quantitatively and qualitatively. The performance of a quality measure is gauged by its prediction accuracy and monotonicity. The quantitative results demonstrate that the proposed WEDD measure is superior than PRD1 and WWPRD measure. The performance of the WWPRD measure is degraded when the test signals contain noise. Both qualitative and quantitative results prove that the proposed WEDD measure is more effective for evaluating the quality of the diagnostic features in the reconstructed signal. The proposed WEDD is well correlated with the cardiologist's perception. WEDD measure is simple and robust under noisy conditions.

References

[1] S. Jalaleddine, C. Hutchens, R. Stratran, W. Coberly, ECG data compression techniques—A unified approach, IEEE Trans. Biomed. Eng. 37 (4) (1990) 329–343.
[2] J.P. Martnez, R. Almeida, S. Olmos, A.P. Rocha, P. Laguna, A Wavelet-based ECG delineator: evaluation on standard databases, IEEE Trans. Biomed. Eng. 51 (4) (2004) 570–581.
[3] S. Kadambe, R. Murray, G.F.B. Bartels, Wavelet transform-based QRS complex detector, IEEE Trans. Biomed. Eng. 46 (7) (1999) 838–847.
[4] C. Li, C. Zheng, C. Tai, Detection of ECG characteristic points using Wavelet transforms, IEEE Trans. Biomed. Eng. 42 (1) (1995) 21–28.
[5] B.A. Rajoub, An efficient coding method for the compression of ECG signals using the wavelet transform, IEEE Trans. Biomed. Eng. 49 (4) (2002) 355–362.
[6] M.S. Manikandan, S. Dandapat, Wavelet based ECG Compression with Large Zero Zone Quantizer, accepted for IEEE INDICON 2006, Delhi, India, 2006.
[7] E. Berti, F. Chiaraluce, N.E. Evans, J.J. McKee, Reduction of Walsh-transformed electrocardiograms by double logarithmic coding, Trans. Biomed. Eng. 47 (11) (2000) 1543–1547.
[8] W. Philips, G.D. Jonghe, Data compression of ECGs by higher-degree polynomial approximation, IEEE Trans. Biomed. Eng. 39 (4) (1992) 330–337.
[9] B. Bradie, Wavelet packet-based compression of single lead eCG, IEEE Trans. Biomed. Eng. 43 (5) (1996) 493–501.
[10] A.G. Ramakrishnan, S. Saha, ECG coding by Wavelet-based linear prediction, IEEE Trans. Biomed. Eng. 44 (12) (1997) 1253–1261.
[11] G. Nave, A. Cohen, ECG compression using long-term prediction, IEEE Trans. Biomed. Eng. 40 (9) (1993) 877–885.
[12] H. Lee, K.M. Buckley, ECG data compression using cut and align beats approach and 2-D transform, IEEE Trans. Biomed. Eng. 46 (1999) 556–564.
[13] Z. Lu, D.Y. Kim, W.A. Pearlman, Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm, IEEE Trans. Biomed. Eng. 47 (7) (2000) 849–856.
[14] A. Bilgin, M.W. Marcellin, M.I. Altbach, Compression of electrocardiogram signals using JPEG, IEEE Trans. Consum. Electron. 49 (4) (2003) 833–839.

[15] S.C. Tai, C.C. Sun, W.C. Yan, A 2-D ECG compression method based on Wavelet transform and modified SPIHT, IEEE Trans. Biomed. Eng. 52 (6) (2005) 999–1008.
[16] H.H. Chou, Y.J. Chen, Y.C. Shiau, T.S. Kuo, An effective and efficient compression algorithm for ECG signals with irregular periods, IEEE Trans. Biomed. Eng. 53 (6) (2006) 1198–1205.
[17] Y. Zigel, A. Cohen, A. Katz, ECG signal compression using analysis by synthesis coding, IEEE Trans. Biomed. Eng. 47 (10) (2000) 308–316.
[18] S.G. Miaou, H.L. Yen, Quality driven gold washing adaptive vector quantization and its application to ECG data compression, IEEE Trans. Biomed. Eng. 47 (2) (2000) 209–218.
[19] M. Blanco-Velasco, F. Cruz-Roldán, J.I. Goldino-Llorente, K.E. Barner, ECG compression with retrieved quality guaratnteed, IEE Electron. lett. 40 (23) (2004) 1466–1467.
[20] E.B. De Lima Filho, A.B. Eduardo, M.B. Da Silva, W.S. De Carvalho, J. Da Silva Jnior, Koiller, Electrocardiographic signal compression using multiscale recurrent patterns, IEEE Trans. Circuits Syst. I: Regular Papers 52 (12) (2005) 2739–2753.
[21] J. Chen, S. Itoh, A wavelet transform-based ECG compression method guaranteeing desired signal quality, IEEE Trans. Biomed. Eng. 45 (12) (1998) 1414–1419.
[22] Y. Zigel, A. Cohen, A. Katz, The weighted diagnostic distortion (WDD) measure for ECG signal compression, IEEE Trans. Biomed. Eng. 47 (11) (2000) 1422–1430.
[23] M.S. Manikandan, S. Dandapat, ECG signal compression using discrete sinc interpolation, in: Proc. 3rd IEEE Int. Conf. Intelligent Sensing and Information Processing, Bangalore, India, (2005), pp. 14–19.
[24] A.S. Al-Fahoum, Quality Assessment of ECG compression techniques using a Wavelet-based diagnostic measure, IEEE Trans. Inf. Technol. Biomed. 10 (1) (2006) 182–191.
[25] R.S.H. Istepanian, A.A. Petrosian, Optimal zonal wavelet-based ECG data compression for a mobile telecardiology system, IEEE Trans. Inf. Technol. Biomed. 4 (3) (2000) 200–211.
[26] V.X. Afonso, W.J. Tompkins, T.Q. Nguyen, S. Luo, ECG beat detection using filter banks, IEEE Trans. Biomed. Eng. 46 (2) (1999) 192–202.
[27] N.V. Thakor, J.G. Webster, W.J. Tompkins, Estimation of QRS complex power spectrum for design of a QRS filter, IEEE Trans. Biomed. Eng. 31 (1984) 702–706.
[28] I. Daubechies, Orthonormal bases of compactly supported Wavelets, Commun. Pure Appl. Math. 41 (1988) 909–996.
[29] A. Cohen, I. Daubeehies, J.C. Feauveau. Biorthogonal bases of compactly supported wavelets. ATT Bell Lab., Tech. Rep. TM 11217–900529-07, 1990.
[30] S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, IEEE Trans. Patt. Anal. Mach. Intel. 11 (1989) 574–693.
[31] M. Antonini, Image coding using wavelet transform, IEEE Trans. Image Proc. 1 (2) (1992) 205–220.
[32] B.E. Usevitch, A tutorial on modern lossy wavelet image compression: Foundation of JPEG, IEEE Signal Process. Mag. 18 (5) (2001) 22–35.
[33] VQEQ, Final report from the video quality experts group on the validation of objective models of video quality assessment, March 2000. http://www.vqeg.org/.
[34] S. Mallat, S. Zhong, Characterization of signals from multi-scale edges, IEEE Trans. Patt. Anal. Mach. Intell. 14 (7) (1992) 710–732.