CHAPTER 3
ECG DATA COMPRESSION METHODS

3.1 Introduction

A typical computerized medical signal processing systems acquires a large amount of data that is difficult to store and transmit[20],[47],[70]. It is very desirable to find a method of reducing the quantity of data without loss of important information. All data compression algorithms seek to minimize data storage by eliminating redundancy where possible. The compression ratio is defined as the ratio of the number of bits of the original signal to the number stored in the compressed signal. A high compression ratio is wanted, typically, but using this alone to compare data compression algorithms is not acceptable. Generally the bandwidth, sampling frequency, and precision of the original data very much effect the compression ratio.

A data compression algorithm must also represent the data with acceptable fidelity. In biomedical data compression, the clinical acceptability of the reconstructed signal has to be determined through visual inspection. The residual between the reconstructed signal and the original signal may also be measured by a numerical measure. A lossless data compression algorithm produces zero residual, and the reconstructed signal exactly replicates the original signal[48],[71]. However, clinically acceptable quality is neither guaranteed by a low nonzero residual nor ruled out by a high numerical residual.

3.2 Direct signal Compression methods

Direct signal compression methods are also known as time domain techniques dedicated to compression of ECG signal. The mode of operation is to
extract a subset of significant samples from the original sample set. Which signal samples are significant, depends on the underlying criterion for the sample selection process. To get a high performance time-domain compression algorithm, much effort should be put in designing intelligent sample selection criteria. The original signal is reconstructed by an inverse process, most often by drawing straight lines between the extracted samples. This category includes the FAN, CORTES, AZTEC, AZTDIS, SAIES, TP, SLOPE, SAPA, CORNER algorithms. The CCSP(Cardinality Constrained Shortest Path) technique, presented in Chapter 4, is also fit into this category.

3.2.1 The FAN algorithm

The basic idea of this algorithm is to identify signal segments where a straight line serves as a close approximation, and to discard all but the terminal

Figure 3.1 Illustration of the FAN algorithm

points along this line[19]. When significant deviations from this line are detected, the corresponding samples are included in the extracted signal samples. The FAN
algorithm accomplishes the above idea by initially accepting the very first sample point as shown in Figure 3.1.

Next it computes a range within which succeeding samples must be found if they are to be fit by a straight line. This is done by drawing two lines \((U_1; L_1)\) between the initial point and the next sample point plus a specified threshold \((\pm \varepsilon)\) as shown in Figure 3.1. If the third sample point falls within the area bounded by the two lines, new slopes \((U_2; L_2)\) are drawn between the initial point and the third sample point plus the same specified threshold. These new lines \((U_2; L_2)\) are then compared to the previously stored lines \((U_1; L_1)\) and the most restrictive lines are retained \((U_1; L_2)\). The process is repeated, comparing future sample values to the most restrictive lines. Whenever a sample falls outside the area bounded by the most restrictive lines, its predecessor is accepted as a significant sample, and the procedure above is repeated from this point on. The FAN method offers control of the absolute error by guaranteeing that the error between the reconstructed and original signal is less than or equal to the threshold, \(\varepsilon\). The FAN algorithm is computationally an efficient algorithm and provides a better compression ratio with less PRD.

3.2.2 The AZTEC algorithm

The AZTEC algorithm[18] converts raw ECG sample points into plateaus and slopes. The AZTEC plateaus (horizontal lines) are produced by utilizing the zero-order interpolation. The stored values for each plateau are the amplitude value of the line and its length (the number of samples with which the line can be interpolated within aperture \(\varepsilon\)). The production of an AZTEC slope starts when the number of samples needed to form a plateau is less than three. The slope is saved
whenever a plateau of three samples or more can be formed. The stored valued of the slope are the duration (number of samples of the slope) and the final elevation (amplitude of last sample point). Even though the AZTEC provides a high data reduction ratio, the reconstructed signal has poor fidelity mainly because of the discontinuity (step-like quantization) of the waves.

3.2.3 The TP algorithm

The turning point (TP) data reduction algorithm[69] was developed for the purpose of reducing the sampling frequency of an ECG signal from 200 to 100 Hz. The algorithm processes three data points at a time: a reference point \(X_0\) and two consecutive data points \(X_1\) and \(X_2\). Either \(X_1\) or \(X_2\) is to be retained. This depends on which point preserves the slope of the original three points. In this method, only the amplitudes are to be stored but not their locations.

3.2.4 The CORTES algorithm

CORTES algorithm[37] is a hybrid of the AZTEC and TP algorithms. In this algorithm, the ability of the TP is exploited to track the fast changes in the signal, and the ability of the AZTEC is exploited to compress effectively isoelectric regions. CORTES applies the TP algorithm to the high frequency regions (QRS complexes), whereas it applies the AZTEC algorithm to the lower frequency regions and to the isoelectric regions of the ECG signal.

3.2.5 The SAPA algorithm

The Scan-Along Polygonal Approximation (SAPA) algorithm[35], is based on first-order interpolation. In this method, the compressor searches for the most distant sample (on the time axis), such that if a line is drawn between it and the last stored sample, the local error along the line will be lower than a specific error
tolerance - ε. The location and the amplitude of this sample are stored, and this process recurs. In this method, the reconstructed signal looks like a broken line, and its fidelity depends on the error threshold (ε). The greater the threshold is, the better the compression ratio, but the reconstructed signal has poorer fidelity.

3.2.6 The SLOPE algorithm

The SLOPE algorithm[66] attempts to delimit linear segments of different lengths and different slopes in the ECG signal. It considers some adjacent samples as a vector, and this vector is extended if the coming samples falls in a fan spanned by this vector and a threshold angle; otherwise, it is delimited as a linear segment. Similar to the SAPA and FAN algorithms, the SLOPE reconstructed signal looked like as continuous broken line.

3.2.7 The CORNER algorithm

The CORNER algorithm[65] selects “corner points” by using the curvature of a sample and its displacement from an encoded linear segment as criteria. The curvature is estimated using the second-order difference signal.

3.3 Transformation methods

Transformation techniques[6] have generally been used in multi-lead ECG compression and require preprocessing of the input signal by a linear orthogonal transformation and encoding of the output (expansion coefficients) using an appropriate error criterion. For signal reconstruction, an inverse transformation is carried out and the ECG signal is recovered with some error.
In principle, if the sample sequence of the current ECG beat is considered as an N-dimensional vector \( x \), the transform of \( x \) is given by the \( N \)-dimensional vector, \( y \),

\[
y = Ax
\]

where \( A \) is the transform matrix \((N*N)\). The original signal \( x \) can be obtained from the transform vector by the inverse transform,

\[
x = By
\]

where,

\[
B = A^{-1}
\]

Many orthogonal transform compression algorithms[6] for ECG signals have been presented in the last thirty years, such as the Fourier Transform[63], Walsh Transform[39], Discrete Cosine Transform[13], and Karhunen-Loeve Transform (KLT)[59]. The typical performances of the transform methods are compression ratio between 3:1 to 12:1, where the KLT has the best compression ratio. The KLT is an optimal transform in the sense that the fewest orthonormal functions are required to represent the signal.

In recent years, the wavelet transform has been used for ECG data compression. Many ECG compression algorithms based on wavelet transform have been developed[8],[15]. The compression ratios obtained are ranging from 13.5:1 to 22.9:1 with the corresponding PRD values between 5.5% and 13.3%.

### 3.4 Parametric methods

Although most of the reported ECG compression algorithms belong to the first two methods, i.e. direct techniques and transformation techniques, more
and more ECG compression algorithms based on parametric techniques have been proposed in recent years. Some of these algorithms are hybrids of direct and parametric techniques or transformation and parametric techniques. The compression algorithms based on parametric techniques require a preprocessing stage, which is sometimes heavy in the sense of calculation, but this is not a problem for computers today. Some examples of this technique are ECG compression algorithms using beat code book, ECG compression algorithms using Artificial neural networks, and peak picking and vector quantization techniques[8],[17],[22],[43],[50].

3.5 Compression and distortion measures

The criterion for testing performance of compression algorithms includes three components: compression measure, reconstruction error and computational complexity. The compression measure and the reconstruction error are usually dependent on each other and are used to create the rate-distortion function of the algorithm[57],[58]. The computational complexity component is part of the practical implementation consideration but it is not part of any theoretical measure.

3.5.1 Error criterion and distortion measure

One of the most difficult problems in ECG compression applications and reconstruction is defining the error criterion[74],[75]. The purpose of the compression system is to remove redundancy, the irrelevant information (which does not contain diagnostic information – in the ECG case). Consequently the error criterion has to be defined such that it will measure the ability of the
reconstructed signal to preserve the relevant information. Such a criterion has been defined in the past as "diagnostability". A similar problem exists in synthesized speech signals, in which the criterion "intelligibility" has been defined. Today the accepted way to examine diagnostability is to get cardiologists’ evaluations of the system's performance. This solution is good for getting evaluations of coders’ performances, but it can not be used as a tool for designing ECG coders and certainly, can not be used as an integral part of the compression algorithm. However, in order to use such a criterion for coders design, one has to give its mathematical model. As yet, there is no such mathematical structure to this criterion, and all accepted error measures are still variations of the Mean Square Error or absolute error, which are easy to compute mathematically, but are not always diagnostically relevant.

In most ECG compression algorithms, the Percent Root-mean-square Difference (PRD) measure is employed:

$$PRD = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - \bar{x}(n))^2}{\sum_{n=1}^{N}x^2(n)}} \times 100$$

where $x(n)$ is the original signal, $\bar{x}(n)$ is the reconstructed signal, and $N$ is the length of the window over which the PRD is calculated. In some of the articles a fixed version of PRD definition is used:

$$PRD = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - \bar{x}(n))^2}{\sum_{n=1}^{N}(x(n) - \bar{x})^2}} \times 100$$

This definition is independent in the DC level of the original signal. One
can use the first definition, but the original signal has to have a zero mean.

In the literature, there are some other error measures[76] for comparing original and reconstructed ECG signals, such as the Root Mean Square error (RMS):

\[
RMS = \sqrt{\frac{\sum_{n=1}^{N} (x(n) - \bar{x}(n))^2}{N}}
\]

Another distortion measure is the Signal to Noise Ratio, which is expressed as:

\[
SNR = 10 \log \left( \frac{\sum_{n=1}^{N} (x(n) - \bar{x})^2}{\sum_{n=1}^{N} (x(n) - \bar{x}(n))^2} \right)
\]

where \( \bar{x} \) is the average value of the original signal. The relation between the SNR and the PRD is: \( SNR = -20 \log PRD \).