AIN SHAMS UNIVERSITY

# ELECTROCARDIOGRAM DATA COMPRESSION: A COMPUTER-BASED COMPARATIVE STUDY

A THESIS

SUBMITTED TO THE FACULTY OF ENGINEERING

ELECTRONICS AND COMPUTER ENGINEERING DEPARTMENT

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

*MASTER OF SCIENCE*

IN

ELECTRICAL ENGINEERING

BY

HAZEM MAHMOUD ABBAS

SUPERVISED BY

PROF.DR. HUSSEIN I. SHAHEIN
DR. HANI K. MAHDI

1988

بسم الله الرحمن الرحيم

بسم الله الرحمن الرحيم

# APPROVAL SHEET

The Thesis on

*ELECTROCARDIOGRAM DATA COMPRESSION :*
*A COMPUTER BASED COMPARATIVE STUDY*

by

Eng. Hazem Mahmoud Abbas

approved by

Prof. Y. A. El-Hakiem
(Alexandria University) ....................

Prof. M. E. Rasmi
(Cairo University) ....................

Prof. H. I. Shahein
(Ain Shams University) ....................

*To my*

*family*

## ACKNOWLEDGEMENT

*I wish to express my indebtedness to Prof. H. I. Shahein for his valued advices, guidance, and supervision throughout the various stages of this work.*

# ABSTRACT

In this thesis, an ECG data compression technique (**CUSAPA**) is developed. This technique is based on combining the features of the well-known cubic spline approximation and the scan-along polygonal approximation. The S-Q segment is compressed using the cubic spline function while the rest of the signal (QRS complex) is processed with the polygonal approximation.

**CUSAPA** has produced as large **CR** as those obtained with other offline techniques (if not better) but with improved visuall acceptance. Also QRS, P, and T segments with fidelities never reached by the 10 other techniques considered in this study. Moreover, **CUSAPA** achieves superior 50-Hz noise removal when compared with other techniques. With some modifications to an existing QRS detector, the QRS complex is filtered out with large degree of accuracy.

This QRS detector extracts the QRS complex without precise determination of the onset or end points. Nor the locations or the amplitudes of the peak points are calculated. With the application of thresholding criteria and line segmentation the attributes of both the terminal and peak points are computed. Thus, a proper QRS detection is resulted.

Additionally, a syntactic algorithm based on the attribute grammar notation is proposed and applied to the S-Q segment to return back the best initial knots for which the optimal ones are calculated using some **IMSL** library routines. The syntactic algorithm has the feature of avoiding locating false knots due to the effect of the existence of 50-Hz noise in the original signal.

An extensive simulation and comparative analysis for the proposed **(CUSAPA)** algorithm and the other well-known techniques were carried out and implemented on an **IBM\XT** machine using an ECG database consisting of nine signals belonging to seven normal and abnormal cases. These techniques are classified into three classes:

1. *Data handling class*: This class enlists the following techniques : Turning-point **(TP)**, Amplitude zone time epoch coding **(AZTEC)**, and Coordinate reduction time encoding system **(CORTES)**.
2. *Redundancy reduction class*: This group contains three techniques : Zero-order predictors **(ZOP)**, first-order interpolator **(FOI)** and Scan-along polygonal approximation algorithm **(SAPA)**.
3. *Transform coding class*: Some of the orthogonal transforms are considered such as Fast Fourier Transform **(FFT)**, Fast Walsh Transform **(FWT)**, Walsh-Hadamard Transform **(WHT)** and Discrete cosine Transform **(DCT)**.

The comparative study is based on different performance evaluation parameters: the compression ratio (CR), the percentage root-mean square difference (PRD), the visual acceptability of the reconstructed signal, the processing time per sample (TPS), the noise removal capability, and the error in QRS complex detection.

The results of this comparative analysis may justify the application of a certain data compression technique to a specific abnormal case.

# List of Symbols and Abbreviations

| *Notation* | *Explaination* |
|---|---|
| **Chapter 1** | |
| ECG | ElectroCardioGraph |
| A/D | Analog to Digital |
| S-A | Sinoatrial |
| A-V | AtrioVentricular |
| LA | Left Arm |
| RA | Right Arm |
| LL | Left Leg |
| VCG | VectorCardioGraph |
| I, II, III | Unipolar Leads |
| aVL, aVR, aVF | Bipolar Leads |
| $V_1$-$V_6$ | Chest Leads |
| X, Y, Z | Frank Leads |
| **PVC** | Premature Ventricular Contraction |
| **VTACH** | Ventricular Tachycardia |
| **AFIB** | Atrial Fibrillation |
| **AFLUT** | Atrial Flutter |
| **VFIB** | Ventricular Fibrillation |
| **PHB** | Partial Heart Block |
| **CHAPTER 2** | |
| *f* | function |
| **AZTEC** | Amplitude Zone Time Epoch Coding |

| | |
|---|---|
| **CORTES** | Coordinate Reduction Time Encoding System |
| **TP** | Turning Point |
| XO X1 X2 | 3-points **TP** Configuration |
| $V_i$ | *ith* ECG data sample |
| $V_o$ | Initial ECG data sample |
| $V_{mx}$ | Maximum value of ECG data samples |
| $V_{mn}$ | Minimum value of ECG data samples |
| $V_{th}$ | Aperture threshold |
| $L_i$ | **AZTEC** line |
| $Y(L_i), V_1$ | **AZTEC** line amplitude |
| $T(L_i), T_1$ | **AZTEC** line duration |
| ZOI | Zero-order prediction |
| S | **AZTEC** slope |
| $V_e$ | Maximum duration of **AZTEC** plateau |
| PNT | Plateau duration threshold |
| $V_{sl}$ | **AZTEC** slope value |
| $T_{sl}$ | **AZTEC** slope duration |
| P(i) | *ith* output point of curve-fitting processor |
| $S_0, S_1, S_2$ | Coefficients of the curve-fitting parabola |
| **E** | Least-squares error |
| x(i) | *ith* input sample to the curve-fitting processor |
| Z | Shift operator |
| $V_{ln}$ | Plateau length threshold triggering saving of TP data |

| | |
|---|---|
| P1,P2 | CORTES array pointers |
| F1-F4 | CORTES flags |
| ZOP | Zero-order prediction |
| FOP | First-order prediction |
| $\Delta^n$ | *nth*-order prediction coefficient |
| K | ZOP aperture width |
| FOI | First-order interpolator |
| SAPA | Scan-along polygonal approximation |
| ECG(i) | *ith* ECG original data sample |
| NECG(i) | *ith* ECG reconstructed data sample |
| $\epsilon$ | SAPA allowable error |
| s | The last saved SAPA sample |
| $g(c,\pm\epsilon)$ | Two lines joining ECG(s) to ECG(c)$\pm\epsilon$ |
| $m_1$ | The current smallest value of $g(c,\epsilon)$ |
| $m_2$ | The current largest value of $g(c,-\epsilon)$ |
| g(c+i,0) | straight line between (s,g(s)) and ((c+i),g(c+i)) |
| KLT | Karhunen-Loeve transform |
| FFT | Fast Fourier transform |
| DCT | Discrete Cosine transform |
| FWT | Fast Walsh transform |
| WHT | Walsh-Hadamard transform |
| T(u) | Transform of original sample f(x) |
| g(x,u) | Forward transformation kernel |
| h(x,u) | Inverse transformation kernel |

| | |
|---|---|
| $N$ | Number of data samples |
| $\mathbf{X}$ | Vector of sampled ECG data |
| $\mathbf{\Phi}$ | Transform matrix |
| $\mathbf{Y}$ | Transform vector |
| $\mathbf{C}_x$ | Covariance matrix |
| $\lambda$ | Eigenvalue |
| $\mathbf{I}$ | Unit matrix |
| $M$ | Number of retained coefficients |
| $e^2$ | Mean-square error |
| **DFT** | Discrete Fourier transform |
| $W_{km}$ | **DFT** orthogonal transform |
| $F_{even}, F_{odd}$ | Even and odd components of **DFT** |
| $m(i)$ | Number of multiplications needed for i-point **DFT** |
| $a(i)$ | Number of additions needed for i-point **DFT** |
| $b_k(z)$ | *kth* bit in the binary representation of z |
| $W_{even}, W_{odd}$ | Even and odd components of **FWT** |
| $H_n$ | Hadamard matrix of *nth* order |
| $r, c$ | Hadamard element at *rth* row and *cth* column |

**CHAPTER 3**

| | |
|---|---|
| **CR** | Compression ratio |
| **PRD** | Percentage roor-mean square difference |
| $RMS_o$ | Root-mean square of the original signal |
| $RMS_e$ | Root-mean square of the error signal |

| | |
|---|---|
| TPS | Processing time per sample |
| $ECG'(i)$ | *ith* output of the low-pass filter |
| 2m+1 | Duration of Q, R, or S waves |
| $ECG_2(i)$ | *ith* output of the band-pass filter |
| $ECG_3(i)$ | *ith* output of the high-pass filter |
| F | Weighting coefficient |
| BRD | Number of data points of QRS complex |
| $Start_j$ | Starting point of the the *jth* QRS complex |
| $Final_j$ | Final point of the the *jth* QRS complex |
| Nqrs | Number of QRS complexes in ECG |
| Nwaves | Number of waves in a QRS complex |
| Loc | Wave number whose length equals 2m+1 |
| DECG | First derivative of ECG |
| $(LSign, Lstart, LDur)_i$ | Sign, starting point, and duration of the *ith* line segment |
| Dirc | Direction of Loc*th* wave |
| Ploc,Pval | Time coordinate and amplitude of the peak point |

**CHPATER 4**

| | |
|---|---|
| **CUSAPA** | Cubic splines / **SAPA** combined algorithm |
| pp(t) | Piecewise polynomial |
| $P_i(t)$ | *nth* order polynomial |

| | |
|---|---|
| $a_{ij}$ | Polynomial coefficients |
| $S(t)$ | Spline function |
| $a, b$ | First-degree spline coefficients |
| $S'(t)$ | First derivative of a cubic spline function |
| $S''(t)$ | Second derivative of a cubic spline function |
| $a, b, c, d$ | Cubic spline function coefficients |
| $t_j$ | Time coordinate of *jth* knot |
| $F(\mathbf{c}, t)$ | Approximation function |
| $\mathbf{c}$ | Vector of parameters of approximation |
| $D(t)$ | Data function |
| $L_\infty$ | Chebyshev norm |
| $L_1$ | Least deviation norm |
| $L_2$ | Least-squares norm |
| $f_i(t)$ | *ith* order approximating function |
| LN, LP, LZ | Large negative, large positive, large zero |
| SN, SP, SZ | Small negative, small positive, small zero |
| $Sg_K$ | *kth* line segment |
| $L_i$ | *ith* **SAPA** segment |
| $M_i$ | *ith* spline segment |
| **OCR** | Overall **CR** |
| n | Knots representing spline segment |
| NKD | number of knots descriptor |
| K | Number of subsegments of the syntactic structure |