Towards a Formal Model for ECG Data Analysis and Decision Making

Oswaldo Cuen Tellez, **Ines F. Vega Lopez**

Universidad Autonoma de Sinaloa, Culiacan, Sinaloa, Mexico
o.cuen11@info.uas.edu.mx, ifvega@uas.edu.mx

Abstract

ECG sequences databases lack of a formal way to model analysis and query operations. In the pursuit of such a formal model, the definition of relevant primitive operators is necessary. We identify the convolution operator of sequences as one of those primitive operators. We take two QRS detection algorithms and propose a new way to express them formally using the convolution operator of sequences.

Keywords: ECG Data, ECG Formal Data Model, Convolution, QRS Detection

1. Introduction

Analysis of Electrocardiogram (ECG) signals is a widely used procedure for Cardiovascular diseases (CVD) diagnosis. CVDs are the first cause of deaths worldwide, contributing to nearly one third of global mortality. For this reason, the World Health Organization (WHO) states that the development of methods for monitoring trends of morbidity, mortality and risk factors is a research priority for the prevention and control of CVD [1].

ECG signals carry information about the heart condition of a patient. Therefore, ECG signals are a logical choice of data to be monitored in global health database systems. A large volume of ECG signals is generated everyday by medical centers and physicians. Although vendor-supplied ECG management systems exist, they do not support transferring complete stored data to other databases and they are limited in terms of analysis and querying, supporting only searching and viewing the ECG data of a specific patient [2]. In many medical centers, due to the lack of tools and resources for analysis, ECG data is being just accumulated or, even destroyed and lost, and along with them, valuable health-related information. This makes it necessary to generate knowledge and to develop tools for proper storage and analysis of ECG signals in databases.

ECG signal databases with large enough number of records of a broad spectrum of cardiopathies would represent a valuable tool for medical research, acting as a validation platform of ECG analysis algorithms

and as a diagnosis aid for physicians [3, 4]. The creation and curation of the aforementioned databases requires a proper representation of ECG signals. Digital ECG signals are basically ordered equally spaced discrete-time samples with quantized amplitude, that is, a sequence of values or *time series*. Time series have been studied by the database community for many years, nevertheless, a formal data model equivalent to what the relational model [5] is for conventional data is yet to be proposed. An ECG formal data model must capture the nature of ECGs represented as time series and it has to be able to express the ECG processing and analysis algorithms for decision making.

The goal of this paper is to lay the ground for such formal model by defining a primitive operator. Expressions using this operator can represent commonly used algorithms for the preprocessing of ECG signals before classification and decision support stages. The benefit of using this formal expression is that we can represent, in the same context, different ECG preprocessing approaches, facilitating the task of comparison and analysis of algorithms.

The rest of this paper is organized as follows. Section 2 gives the preliminaries of ECG analysis algorithms and digital signal processing. Section 3 describes two QRS detection algorithms and presents its representation using the convolution operator. Section 4 summarizes the contribution of this paper and gives an outlook to future work.

2. Background

The foundation that provides mathematical soundness to database systems is a formal data model. A formal data model is an abstract, self-contained, logical definition of the data structure and data operators that together make up the abstract machine with which users interact [6].

The data structure of ECG analysis, a time series, can be formally represented as a sequence of real values. A sequence can be seen as an ordered n-tuple. A sequence x is formally written as $x = \langle x[0], x[1], \ldots, x[n] \rangle$, with $x[i] \in \mathbb{R}$.

In this paper, we refer to $x[i]$ just as the "ith sample of the sequence x", in contrast to most of the signal

processing literature, where it usually represents the entire sequence too.

ECG data analysis algorithms take sequences as input. Each algorithm can be seen as a combination of processing blocks or operations over sequences. Our task is to identify the set of primitive operations that is included in the ECG signal analysis algorithms.

2.1. QRS Complex Detection

Since the shape and timing information of an ECG signal, particularly intervals and amplitudes can be used to diagnose CVD [7], several algorithms have been proposed to automatically delineate waves, peaks, and boundaries. We distinguish two main groups of algorithms for this purpose, namely, QRS detection algorithms and wave delineation algorithms [8]. The automatic detection of QRS complexes (and central R-peaks) is the essential first step in any automatic ECG analysis [9].

The QRS complex is the most distinguishable waveform of the ECG signal due to its high amplitude (see Figure 1). It usually serves as a reference point to detect other elements of the cardiac cycle like P and T waves. Kohler et al. [10] presented an in-depth revision and comparison of QRS detection algorithms. This revision suggests the need of reliable methods for noisy and pathological signals as well as more accurate algorithms for research applications. Approaches based on mathematical morphology [11, 12], wavelets [13, 14] and matched filters [15] have recently been proposed.

Since the typical frequency components of a QRS complex range from about 10 Hz to about 25 Hz, almost all QRS detection methods use a filtering stage prior to the actual detection. This filtering stage attenuates the other signal components and artifacts, such as the P wave, the T wave, baseline drift (noise due to respiratory signal and body movements), and uncoupling noise [10]. Therefore, digital filtering is a common task in the ECG analysis. Consequently, it is an evident first operator to be defined as part of the formal data model.

A digital filter is basically an algorithm that transforms an input sequence x to an output sequence y. We can express the relationship between x and y as $y = T(x)$, where T denotes the transformation on x to produce y. In Section 2.2 we describe the digital filter equation and its properties.

Besides filtering out undesired bands of frequency, a digital filter can also be used as a computational means of performing other functions such as integration, differentiation, and estimation [16].

A block diagram for the algorithm structure that matches most QRS detectors is shown in Figure 2. There are two main stages in this structure. A preprocessing, or feature enhancement stage, which includes linear and nonlinear filtering and a decision stage including peak detection and decision logic. On the preprocessing stage, the original ECG signal is transformed to accentuate the QRS complexes. At the same time, noise and artifacts are eliminated. The decision stage is where the QRS complexes are detected, usually through amplitude thresholds for peak detection and decision logic. Often a postprocessing stage is used for the exact determination of the temporal location of each QRS complex [10].

Nonlinear filtering consist of transforming a signal with operations such as rectification through squaring. Nonlinear transformations are not always used in the QRS detection process. In fact, some detection algorithms feed the decision rule stage with the linear filtered signal without any intermediate process.

2.2. Digital Filters

Linear digital filters are an important class of digital signal processing systems. They are modeled as Linear Time-Invariant (LTI) systems. This means that these systems are governed by the principle of superposition and that any delay of the input sequence causes a corresponding shift in the output sequence. An LTI system is completely characterized by a sequence, h, representing the impulse response of the system. The output sequence, y, given an input sequence x, can be obtained through a computation known as the convolution sum. If causality (the output depends only on past and current inputs) is assumed, each ith sample of the output is computed by Equation 1.

$$y[i] = \sum_{k=0}^{\infty} x[k]h[i - k]. \qquad (1)$$

We will use the notation $y = x \star h$ to refer to the convolution of sequences x and h.

The convolution operator in the domain of LTI systems satisfies the following properties. These properties hold for every input sequence [17].

Commutative Law.

$$x \star h = h \star x$$

The convolution operator satifies the commutative law. In other words, x can be regarded as the impulse response of the system and h as the excitation or input signal. We can think of x and h as two elements of the same kind.

Associative Law.

$$(x \star h_1) \star h_2 = x \star (h_1 \star h_2)$$

When x, representing the input sequence of a filtering process, is filtered in cascade by two filters characterized by h_1, and h_2, the output is equivalent to the convolution of x with the output of $h_1 \star h_2$.
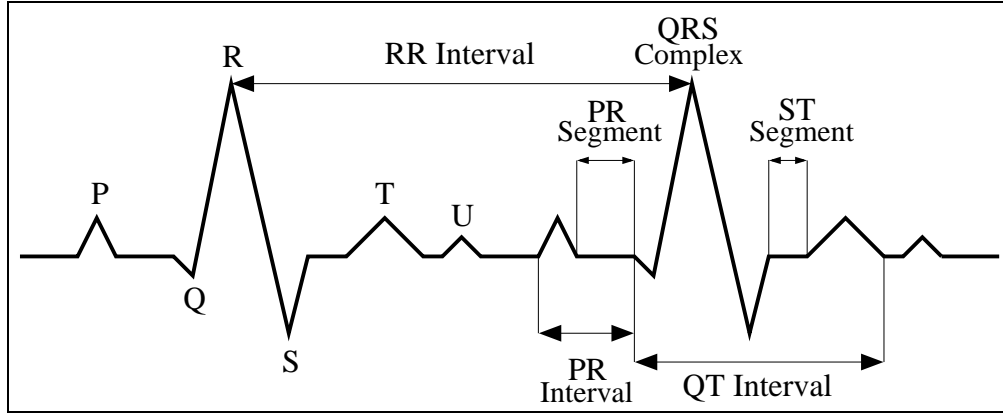
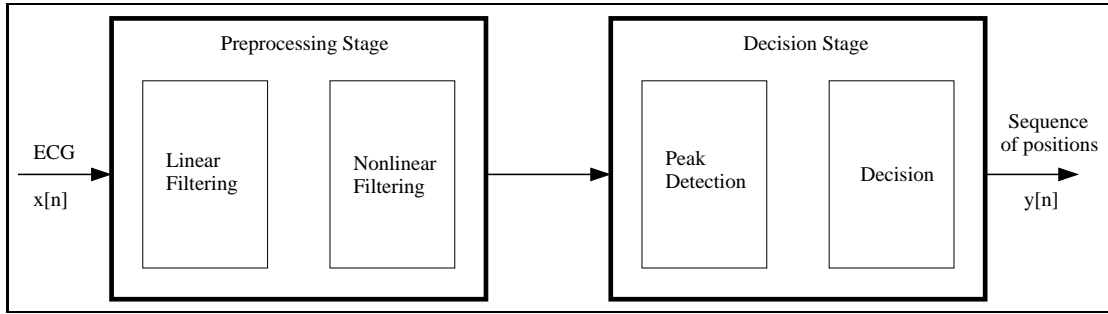Fig. 1: ECG elements: P, Q, R, S, T and U waves, PR and ST segments and PR and QT intervals.

Fig. 2: Algorithm structure of a QRS detector. Adapted from [10].

Distributive Law.

$$x \star (h_1 + h_2) = x \star h_1 + x \star h_2$$

or

$$h \star (x_1 + x_2) = h \star x_1 + h \star x_2$$

If we have two filters characterized by h_1, and h_2, and both are excited by a sequence x, the sum of outputs is identical to the result of a convolution process between the sequence x and sequence $h_1 + h_2$. In the same way, two sequences x_1 and x_2, both filtered by the same sequence h, the sum of outputs is identical to the result of a convolution process between the sequence h and sequence $x_1 + x_2$.

2.3. Wavelet Transform

A wavelet is a real or complex valued function $\psi(t)$ that describes a small wave that grows and decay in a limited time period. Opposite to a big wave like the function $sin(t)$ that keeps oscillating up and down over any instant t, with $0 \leq t < \infty$.

A wavelet function must satify the *admissibility condition* in order to be used for analysis (e.g. edge detection, singularity analysis, etc.). A wavelet $\psi(x)$ is said to be admissible if its Fourier transform, namely,

$$\Psi(\omega) = \int_{-\infty}^{\infty} \psi(t) e^{-i\omega t} \, dt,$$

is such that

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} \, d\omega < \infty.$$

The admissibility condition implies that $\Psi(0) = \int_{-\infty}^{\infty} \psi(t) \, dt = 0$, this means that the average value of the wavelet in the time domain must be zero, and therefore it must be oscillatory. The admissibility condition allows the reconstruction of a function $f(t)$ from its continuous wavelet transform.

The *continuous wavelet transform* is a linear operation that decomposes a signal into basis functions. The basis functions are obtained from a prototype wavelet by means of dilations and contractions (scaling) as well as time shifts.

A function $f(t)$ is scaled when $f(t) \rightarrow f(at)$ where $a > 0$, then it is contracted if $a > 1$ and expanded if $a < 1$ [18]. The dilation of a prototype wavelet is obtained by $\psi_a(t) = \frac{1}{\sqrt{a}} \psi(\frac{t}{a})$. The factor $\frac{1}{\sqrt{a}}$ is used for energy normalization [19].

The *continuous wavelet transform* of a function $f(t)$ at scale a and position τ is given by Equation 2.

$$W f(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t - \tau}{a} \right) \, \mathrm{d}t, \qquad (2)$$

where $*$ denotes the complex conjugation. a and τ are the scale and position parameters respectively. Both parameters vary continuously over the real numbers. If the wavelet prototype is contracted by a small scale value, the wavelet transform can provide information of the finer details of the function $f(t)$. In the same way if the wavelet prototype is expanded, a global view of $f(t)$ can be obtained.

The wavelet transform can be seen as a pass band filtering process, represented as a convolution [20]. The continuous wavelet transform can be rewritten using the convolution operator [21]. Equation 3 computes the wavelet trasform with dilated band-pass filters using the convolution operator $\star$.

$$W f(a, \tau) = f \star \hat{\psi}_a(\tau) \qquad (3)$$

with $\hat{\psi}_a(t) = \frac{1}{\sqrt{a}} \psi^* (\frac{-t}{a})$.

2.3.1. Discrete Wavelet Transform

In the discrete time domain, such as in ECG data, a sequence f is obtained from a sampling process. In addition to its discrete nature, f is finite, $0 \leq i < N$, where N is the number of samples (i.e., elements) in the sequence. The convolution must therefore be modified to take into account the border effects at $i = 0$ and $i = N - 1$, because values outside the boundary of the input signal will be required. The techniques used to ease this border disstortion consist of extending the signal beyond its boundary. The signal is usually padded by zeros, periodicity or reflection [22]. For discrete time signals, the wavelet transform can only be calculated at scales $1 < a < N$. If the scale parameter is the set of integer powers of 2, i.e., $a = 2^j$ where $j \in \mathbb{Z}$, the wavelet transform is called a dyatic wavelet transform. A discrete wavelet scaled by 2^j is defined by Equation 4.

$$\psi_{2^j}[i] = \frac{1}{\sqrt{2^j}} \psi \left[\frac{i}{2^j} \right], \qquad (4)$$

The dyatic wavelet transform in discrete time domain is given by Equation 5.

$$W f[2^j, i] = \frac{1}{\sqrt{2^j}} \sum_{m=0}^{N-1} f[m] \psi^* \left[\frac{m - i}{2^j} \right]. \qquad (5)$$

This discrete wavelet transform, just as digital filtering, can be expressed by the convolution operator by

$$W f[2^j, i] = f \star \hat{\psi}_{2^j}[i],$$

with

$$\hat{\psi}_{2^j}[i] = \psi_{2^j}^*[-i].$$

3. The Convolution Operator for ECG Data Analysis

The process of filtering, as it was mentioned in previous section, is expressed in the time domain, through convolution. Convolution can be seen as a binary operation between sequences, where the result is also a sequence. Here, we define the convolution as an operator for ECG analysis and it will be denoted as $\star_s$.

It is worth noting that convolution can be defined using primitive operations such as addition, time shift, reflection and multiplication [23]. Nevertheless, for ECG analysis purposes, we are going to treat convolution as a primitive operation on sequences.

3.1. QRS Detection Based on Digital Filters

One of the most popular QRS detection method is the proposed by Pan and Tompkins [24], which presents an algorithmic structure similar to the one described in Section 2. The preprocessing stage of this algorithm starts with a band-pass filter. This filter is implemented through a low-pass filter (LPF) and a high-pass filter (HPF) in cascade, filtering out all frequencies outside of the range 5-12 Hz. The filtered signal is then differentiated, squared and averaged. Differentiation provides slope information while the squaring process makes all sequence's values positive and performs nonlinear amplification emphasizing the higher frequencies [24]. Time averaging or moving window integration consists in adding the last C samples and dividing by C. Figure 3 shows a simplified flow diagram of the Pan-Tompkins QRS complex detector. This algorithm uses two sets of thresholds, one for the moving window integration sequence and another one for the filtered ECG sequence. This improves the reliability of QRS complex detection compared to using one sequence alone [24]. In order to be identified as a QRS complex, a peak must be recognized as such on both sequences.

Low and high linear filtering, differentiation and time averaging can be represented by the convolution operator of sequences, where each of these processes is characterized by a sequence, denoted as L, H, D, and A, respectively. For the particular case of the Pan-Tompkins method, the sequences that characterizes the processes are depicted in Figure 4.

Equation 6 is the formal representation of the preprocessing stage of the Pan-Tompkins algorithm where y is the output and E represents the original ECG sequence.

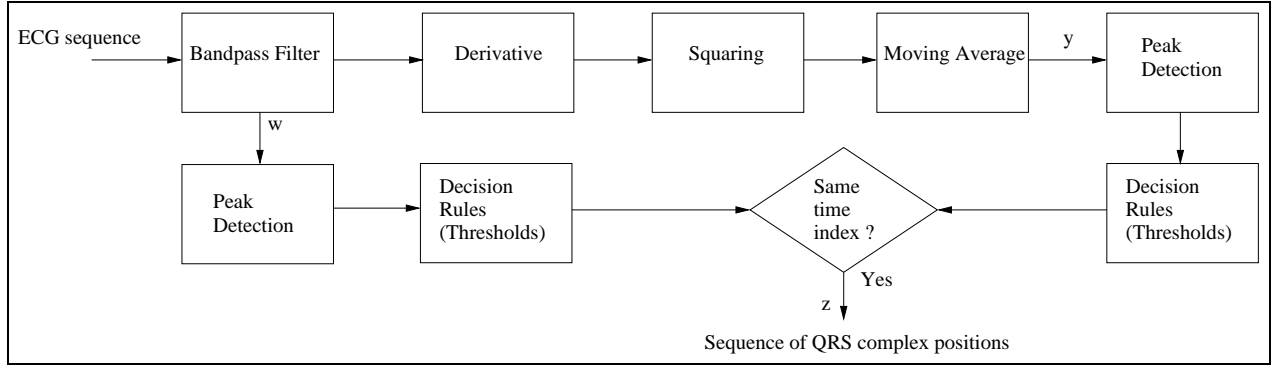$$y = (E \star_s L \star_s H \star_s D)^2 \star_s A, \qquad (6)$$

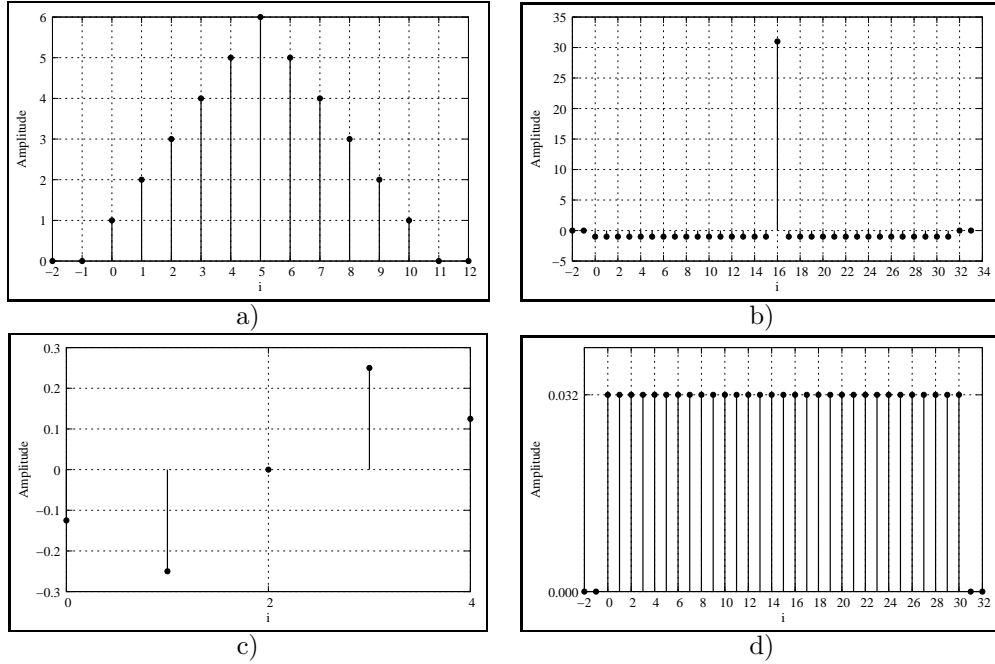Fig. 3: Flow diagram of the Pan-Tompkins algorithm.

Fig. 4: Sequences that characterize a) Low pass filter, b) High pass filter, c) Derivative filter and d) Average filter.

and the filtered sequence w can be formally expressed as

$$w = E \star_s L \star_s H.$$

The process of peak detection consists in determining when the signal changes direction within a predefined time interval. In the decision process, each peak, a local maximum, is classified as a signal peak or noise using two sets of thresholds with two threshold levels in each of the sets. These threshold levels are continuously modified and adapted to the most recent characteristics of the signal.

Most of the decision stages of QRS detection algorithms are rather heuristic and dependent on the results of the preprocessing stage [10]. Therefore, we represent peak detection and decision process as

a predicate. Predicates are functions that return Boolean values. Expression 7 represents the decision stage.

$$Decision_{QRS}(y, w) = z[i], \tag{7}$$

for $0 \leq i \leq length(E) - 1$.

In Table 1 a) is the definition of $z[i]$, where $P(\cdot)$ is a predicate function returning true whenever the argument is classified as a QRS complex and false otherwise.

3.2. Wavelet Based Algorithm for QRS Detection

The wavelet method proposed by Sahambi, Tandon and Bhatt [19] for QRS detection is based on modulus maxima of the wavelet transform. The modulus

$z[i]$	$P(y[i])$ and $P(w[i])$
1	$True$
0	$False$

$z[i]$	$P(Wf_{2^j}[i])$
1	$True$
0	$False$

a)   b)

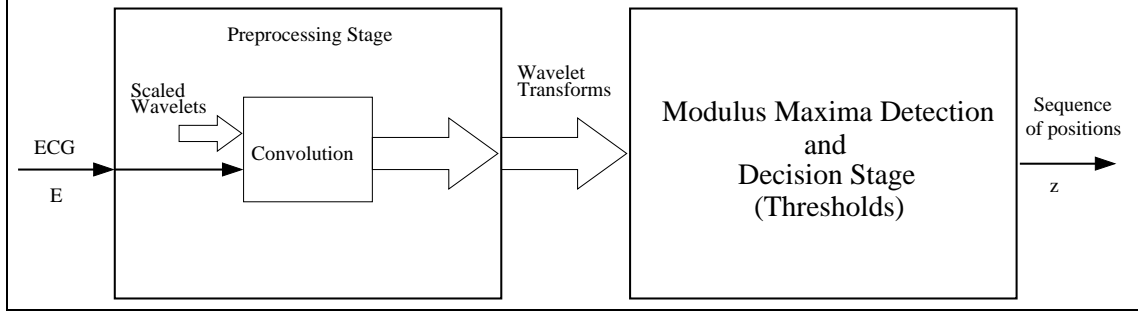Table 1: Definition of $z[i]$ for a) Pan-Tompkins algorithm, and b) Sahambi, Tandon and Bhatt algorithm.

Fig. 5: Algorithm structure of the Sahambi, Tandon and Bhatt algorithm.

maxima is defined as any point $Wf[2^j, i_0]$ such that $\left|Wf[2^j, i]\right| < \left|Wf[2^j, i_0]\right|$ when i belongs to either the left or the right neighborhood of i_0. Moreover, when i belongs to other side of n_0, modulus maxima satisfies $\left|Wf[2^j, i]\right| \leq \left|Wf[2^j, i_0]\right|$. The QRS produces a maxima and minima with a zero crossing in between. Therefore, decision rules are applied to the wavelet transforms to detect the QRS complexes. The scales used are 2^1 to 2^4. These are the scales where most of the QRS complex energy lies [19]. The algorithm structure is depicted in Figure 5.

Expression 8 represents the wavelet transform sequence, Wf_{2^j}, with scale 2^j, of a ECG sequence E using the operator of sequence convolution.

$$Wf_{2^j} = E \star_s \hat{\psi}_{2^j}, \tag{8}$$

where $\hat{\psi}_{2^j}$ is a sequence characterizing the wavelet transform of scale 2^j.

We can used Expression 9 to represent the decision stage in the same way we used it for Pan-Tompkins algorithm.

$$Decision_{QRS}(Wf_{2^j}) = z[i], \tag{9}$$

for $0 \leq i \leq length(E) - 1$, and $1 \leq j \leq 4$.

The definition of $z[i]$ is in Table 1 b), where $P(\cdot)$ is a predicate function returning true whenever the argument is classified as a QRS complex and false otherwise.

4. Conclusions

In this paper, we define a primitive operator of sequences for ECG analysis algorithms. We represent the ECG digital signals as sequences of real values and identify fundamental operations on sequences in analysis algorithms. Linear filtering and wavelet transforms are common techniques for preprocessing ECG signals before classification and decision support algorithms can be applied. Both of these techniques can be expressed by the one common operation proposed in this paper, namely, the convolution.

We use the convolution operator as a primitive operation to express the preprocessing stage of two ECG analysis algorithms. With this formal way of representation of algorithms, we are able to express, in the same context, different approaches of ECG processing algorithms, simplifying the task of comparison and analysis. The main contribution of this paper is the definition of the convolution as a primitive that lays the ground for the creation of a formal model for ECG signal analysis for decision making. For future work, we plan to define a robust set of primitive operators capable of expressing any algorithm for the analysis of ECG signals.

References

[1] World Health Organization. *A Prioritized Research Agenda for Prevention and Control of*

Noncommunicable Diseases. Nonserial Publication Series. World Health Organization, 2011.

[2] M. Y. Park, D. Yoon, N. K. Choi, J. Lee, K. Lee, H. S. Lim, B. J. Park, J. H. Kim, and R. W. Park. Construction of An Open-Access QT Database for Detecting the Proarrhythmia Potential of Marketed Drugs: ECG-ViEW, 2012.

[3] R Gamon and A Cooper. Applying An ECG Database to Aid Decision-Making in the A & E. *Accid Emerg Nurs*, 10(2):62–4, 2002.

[4] C. C. Chiang, W. C. Tzeng, H. C. Cheng, C. T. Lin, Y. C. Yang, S. F. Liang, and S. B Lim. Construction and Application of an Electronic ECG Management System. *Journal of Information Technology and Applications*, 2(3):135–140, 2007.

[5] E. F. Codd. A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*, 13(6):377–387, 1970.

[6] C. Date. *SQL and Relational Theory: How to Write Accurate SQL Code.* Theory in practice. O'Reilly Media, Incorporated, 2009.

[7] F. Rincon, J. Recas, N. Khaled, and D. Atienza. Development and Evaluation of Multilead Wavelet-Based ECG Delineation Algorithms for Embedded Wireless Sensor Nodes. *Information Technology in Biomedicine, IEEE Transactions on*, 15(6):854–863, 2011.

[8] J.P. Martinez, R. Almeida, S. Olmos, A.P. Rocha, and P. Laguna. A Wavelet-Based ECG Delineator: Evaluation on Standard Databases. *Biomedical Engineering, IEEE Transactions on*, 51(4):570–581, 2004.

[9] C. Meyer, J.F. Gavela, and M. Harris. Combining Algorithms in Automatic Detection of QRS Complexes in ECG Signals. *Information Technology in Biomedicine, IEEE Transactions on*, 10(3):468–475, 2006.

[10] B.-U. Kohler, C. Hennig, and R. Orglmeister. The Principles of Software QRS Detection. *Engineering in Medicine and Biology Magazine, IEEE*, 21(1):42–57, 2002.

[11] F. Zhang and Y. Lian. Qrs Detection Based on Multiscale Mathematical Morphology for Wearable ECG Devices in Body Area Networks. *Biomedical Circuits and Systems, IEEE Transactions on*, 3(4):220–228, 2009.

[12] F. Zhang and Y. Lian. QRS Detection Based on Morphological Filter and Energy Envelope for Applications in Body Sensor Networks. *Journal of Signal Processing Systems*, 64:187–194, 2011.

[13] M.-S. Kim, Y.-C. Cho, S.-T. Seo, C.-S. Son, and Y.-N. Kim. A New Method of ECG Feature Detection Based on Combined Wavelet Transform for U-Health Service. *Biomedical Engineering Letters*, 1:108–115, 2011.

[14] B. Abibullaev and H.-D. Seo. A New QRS Detection Method Using Wavelets and Artificial Neural Networks. *Journal of Medical Systems*, 35:683–691, 2011.

[15] Y. Li, H. Yan, F. Hong, and J. Song. A New Approach of QRS Complex Detection Based on Matched Filtering and Triangle Character Analysis. *Australasian Physical & Engineering Sciences in Medicine*, 35:341–356, 2012.

[16] S.T. Karris. *Signals and Systems: With MATLAB Computing and Simulink Modeling.* Orchard Publications, 2008.

[17] J.G. Proakis and D.K. Manolakis. *Digital Signal Processing.* Prentice Hall, 4th edition, 2006.

[18] O. Rioul and M. Vetterli. Wavelets and Signal Processing. *Signal Processing Magazine, IEEE*, 8(4):14 –38, 1991.

[19] J.S. Sahambi, S.N. Tandon, and R.K.P. Bhatt. Using wavelet transforms for ECG Characterization. An On-Line Digital Signal Processing System. *Engineering in Medicine and Biology Magazine, IEEE*, 16(1):77–83, 1997.

[20] S. Mallat. Zero-Crossings of A Wavelet Transform. *Information Theory, IEEE Transactions on*, 37(4):1019 –1033, July 1991.

[21] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way.* Academic Press, 3rd edition, 2008.

[22] G.A. Strang and T.A. Nguyen. *Wavelts and Filter Banks.* Wellesley-Cambridge Press, 1996.

[23] L.A. D'Alotto, C.R. Giardina, and H. Luo. *A Unified Signal Algebra Approach to Two-Dimensional Parallel Digital Signal Processing.* Monographs and Textbooks in Pure and Applied Mathematics, 210. Marcel Dekker, 1998.

[24] J. Pan and W.J. Tompkins. A Real-Time QRS Detection Algorithm. *Biomedical Engineering, IEEE Transactions on*, BME-32(3):230–236, 1985.