Applied Medical Informatics
Original Research Vol. 35, No. 4/2014, pp: 21-26

Grammar Formalism for ECG Signal Interpretation and Classification

Salah HAMDI*, Asma BEN ABDALLAH, Mohamed Hedi BEDOUI

Laboratory of Technology and Medical Imaging (LTIM), Faculty of Medicine of Monastir (FMM), University of Monastir, Ibn Sina City, 5000, Monastir, Tunisia.
E-mails: hamdisalah@yahoo.fr; asma.benabdallah@cristal.rnu.tn; medhedi.bedoui@fmm.rnu.tn

* Author to whom correspondence should be addressed; Tel.: +216-52-27-19-12;
Fax: +216 73 460 737

Received: 29 October 2014 /Accepted: 18 December 2014/ Published online: 26 December 2014

Abstract
This paper shows that the grammar formalism can be pushed to be applied for the description and the classification of electrocardiograms signals (ECG). We will describe an ECG signal as a sequence of tokens based on specific vocabularies and a set of grammatical rules. *QRS* complexes, *RR* distances, *PR* and *QT* intervals will be calculated. This type of work is intended for medical diagnosis assistance of an ECG signal.

Keywords: ECG (electrocardiogram); *RR* Distance; Classification; Grammar; Lexical Analysis; Syntactic Analysis.

Introduction

An electrocardiogram (ECG) is a graphical representation of the electrical potential which controls the activity of the heart muscle. It is an essential element of patient monitoring and cardiovascular diseases diagnosis. By using three electrodes: one on each wrist around the radial artery and one that is grounded on the upper, left foot, we are able to record these electrical potentials produced by the heart. A wide variety of approaches of ECG signal analysis and classification were used including neural networks, fuzzy sets, Support Vector Machines (SVM) and wavelets. In the context of classification of the ECG signal, more interesting works can be found in the literature. In [1], Ince et al. proposed a technique that uses the wavelet transform to extract morphological information from the ECG data. In [2], Sahambi et al. presented an approach that used a dyadic wavelet to characterize the ECG signal. In [3, 4], a process for detecting ventricular contraction are provided by using wavelet transform and fuzzy neural networks. In [5], a dyadic wavelet transform is used to extract features of the ECG signal. The proposed algorithm detected the QRS complex representing Q, R and s peaks, the T wave and the P wave. Also, the SVM machine has shown efficacy in many application including ECG signals classification [6, 7, 8]. Kampouraki et al. [9] extracted ECG features using statistical methods and SVM machines. On the other hand, a few approaches have been based on the grammar formalism [10, 11]. Grammar is a formalism to describe one language and recognize all the learned words. Kokai et al. [10] presented a learning machine that was able to learn both syntax and semantic rules of an ECG grammar. QRS complex is considered as a negative peak followed by a positive peak followed later by a negative peak. Panagiotis et al. [11] applied a syntactic recognition method to measure the associated parameters of an ECG signal. In this context, our work is warranted: we will describe an ECG signal as a sequence of tokens based on a specific vocabulary and set of grammatical rules. First, the aim of our analyzers is to classify the input signal in the case of an ECG signal or not. Next, if it is an ECG signal our analyzers aims to determine the number of cardiac cycles and the various indicators such as the *QRS* complex, the *RR* distance, *PR* and *QT* intervals for each cardiac cycle.

Material and Methods

Grammar-Based ECG Signal Description

Figure 1 shows one cycle of an ECG signal. Each cycle is characterized by both *P* and *T* waves, *R*, Q and *S* peaks. These elements are eventually separated by resting phases. Mathematically, an ECG signal can be represented by one function *S(t)* as follows:

$$S: \quad \mathbb{R}_+ \rightarrow \mathbb{R}$$
$$\qquad t \rightarrow S(t)$$

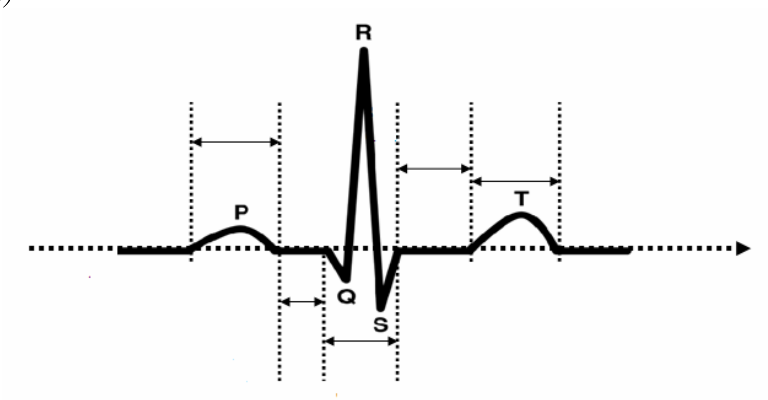

Figure 1. A portion of an ECG signal representing only one cardiac cycle.

Figure 2 summarizes the different steps of our grammatical method. First, the signal is normalized then filtered by a series of low-pass and high-pass filters to eliminate and reduce noise. Second, lexical analysis is used to specify the recognized lexemes and tokens. In this case, one token corresponds to an element of the signal (wave, peak or rest). Third, syntactic analysis arranges the known tokens based on a grammar. The aim of this step is to determine whether one token belongs to the language generated by a grammar or not. If the sequence of tokens respects the language so the input signal is an ECG. Indeed, the *QRS* complex, the *RR* distance, *PR* and *QT* intervals are deducted. Finally, these indicators will allow us to clearly classify the ECG signal in normal or abnormal.

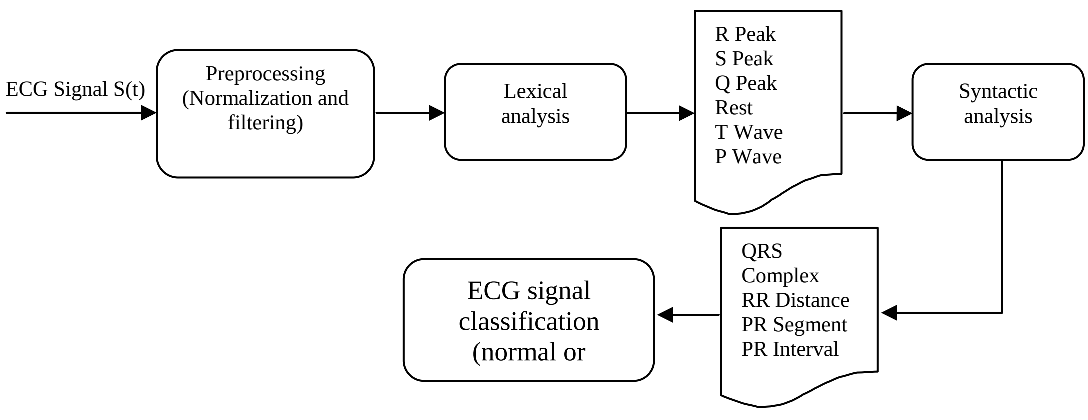

Figure 2. A general view of grammatical analysis and classification of the ECG signal

ECG Preprocessing

In the real case, an ECG signal is too noisy (Figure 3); whence the necessity of preprocessing to reduce noise and facilitate parsing later. The following mathematical equations describe the different steps of the preprocessing phase of the signal *S(t)*:

Step 1: Normalization (equation 1).

Step 2: Low-pass filtering and normalization (equations 2 and 3).

Step 3: High-pass filtering and normalization (equations 4 and 5).

Please note that *Mean* function returns the mean value of the signal, *Max* function returns the highest element in the signal and *Conv* function convolves vectors and returns the convolution and polynomial multiplication. The results are shown in Figures 3 and 4.

$$S(t) = [S(t) - \mathrm{Mean}(S(t))]/\mathrm{Max}(|S(t)|) \tag{1}$$

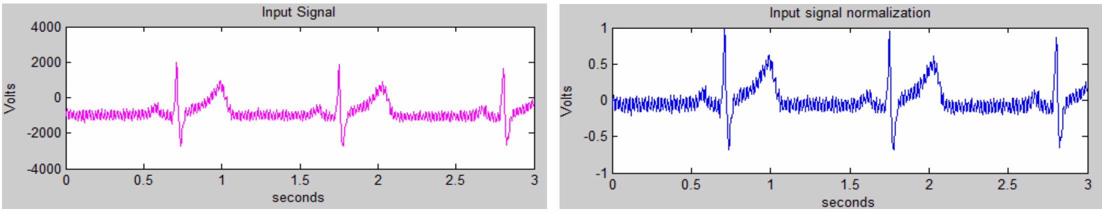

Figure 3. (a) 3 seconds of real ECG signal. **(b)** Normalized ECG signal.

$$S(t) = \mathrm{conv}(S(t), ([10000\text{-}200001], [1\text{-}21], [1000000000000])) \tag{2}$$

$$S(t) = S(t)/\mathrm{Max}(|S(t)|) \tag{3}$$

$$S(t) = \mathrm{conv}(S(t), ([-100000000000000032\text{-}32000000000000001], [-1\ 1], [10000000000000000000000000000000])) \tag{4}$$

$$S(t)=S(t)/\mathrm{Max}(|S(t)|) \tag{5}$$

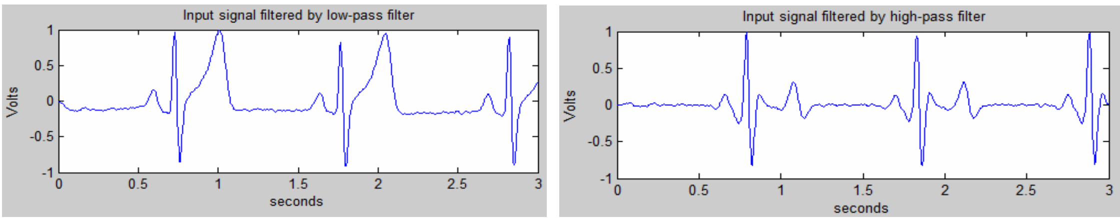

Figure 4. (a) 3 seconds of real ECG signal filtered by low pass filter. **(b)** ECG signal filtered by high-pass filter

ECG Grammatical Analysis

In this section and after normalizing the ECG signal, we will treat signal values that belong to the limited interval [-1, 1]. Indeed, we have come to describe the signal as a sequence of reels almost nil, positive or negative. We are based on the following grammar *G* to describe the signal. *VT* is the terminal vocabulary of the grammar. *VN* is the non terminal vocabulary. *S* is the axiom. *P* is the set of production rules.

The following grammar *G* can describe the ECG and automate the detection of deviations from the norm where the characters ' * 'and ' + ' mean ' zero or more times' and 'One or more times' respectively. The characters ' ? ' means 'zero or one time' and the characters ' | ' means 'or'.

Grammar G = (VT, VN, S, P)

VT = {0,1,2,3,4,5,6,7,8,9, -, ., ε}
VN = {ECG, Col, Palier, Coda, Repos, P, Q, R, S, T, End}
S = {ECG}
P is the set of the following rules:
S → Repos P Col
Col → Q R Palier
Palier → S Coda
Coda → Repos T End

End → ECG | ε // ε indicates the end of the signal
P → {0.[1-9][0-9]*}+ // positive reels
Repos → {{-}?0.0[0-9]*}+ // almost nil reels
Q → {-0.[1-9][0-9]*}+ // negative reels
R → {0.[1-9][0-9]*}+ // positive reels
S → {-0.[1-9][0-9]*}+ // negative reels
T → {0.[1-9][0-9]*}+ // positive reels

Results and Discussion

The proposed grammar-based method was applied on different portions of one real ECG signal using *C* and *Matlab* languages. *Flex* and *Bison* tools [12] were also used to make lexical and syntactic analysis respectively. Each signal portion has duration measured in seconds and represents several cardiac cycles. Every time, we have applied our method on only one portion in order to extract peaks and wave-forms and then measure the signal parameters.

Figure 5 shows an application on a portion of 45 seconds. The various signal indicators such as *RR* distance, *QRS* complex, *PR* interval, *PR* segment and *QT* interval are calculated. Table 1 shows an application of our grammar-based method on several portions of real ECG signal. Every time the number of detected cardiac cycles was measured. Table 1 show that for the all portions no parameter was deductible.

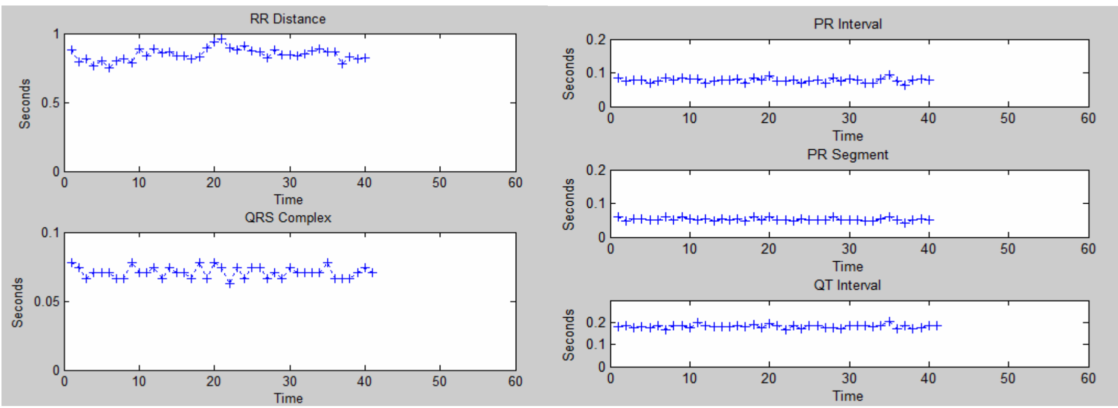

Figure 5. Application on a portion of 45 seconds of real ECG signal and extraction of different signal indicators (RR Distance, QRS complex, PR interval, PR segment, QT interval)

According to the results shown in Figure 5 and Table 1, the *QRS* complex period is less than 0.1 seconds. In addition, the signal parameters are almost stable and the pulse rate (Table 1) is 54 beats / min; thus the input signal is a normal ECG.

We have shown that the grammar formalism can be applied for description and classification of ECG signals. We have described an ECG as sequence of lexemes based on specific vocabularies and grammatical rules. Thus, the *QRS* complex, the *RR* distance, *PR* and *QT* intervals were measured. These indicators have allowed us to clearly classify the ECG signal in normal or abnormal.

Table 1. Application on several portions of one real ECG signal and measurement of the detected cardiac cycles.

Signal duration (seconds)	Number of the detected cardiac cycles	Accuracy rate (%)
1	1	100
10	9	100
20	19	100
30	27	100
40	37	100
50	46	100
60	54	100

In this section, we will make a comparative study with the method of Holsinger [13] for the detection of R peak and the measurement of RR distance. The Holsinger method works as follows:

Calculate the threshold *S* which represents the maximum amplitude of the signal.

The threshold *S* is calculated according to the following equation where the parameter *a* is a constant coefficient <1:

$$S = Max(|S(t)|)*a \tag{6}$$

Find the first point which exceeds the threshold *S* which indicates the *R* peak.

The following figure shows that applying Holsinger technique along a portion of 45 seconds of the real ECG signal, six *R* peaks (about 15%) are franchises. Besides, applying our approach based on grammar, no *R* peak is deductible and the *RR* distances are almost stable.

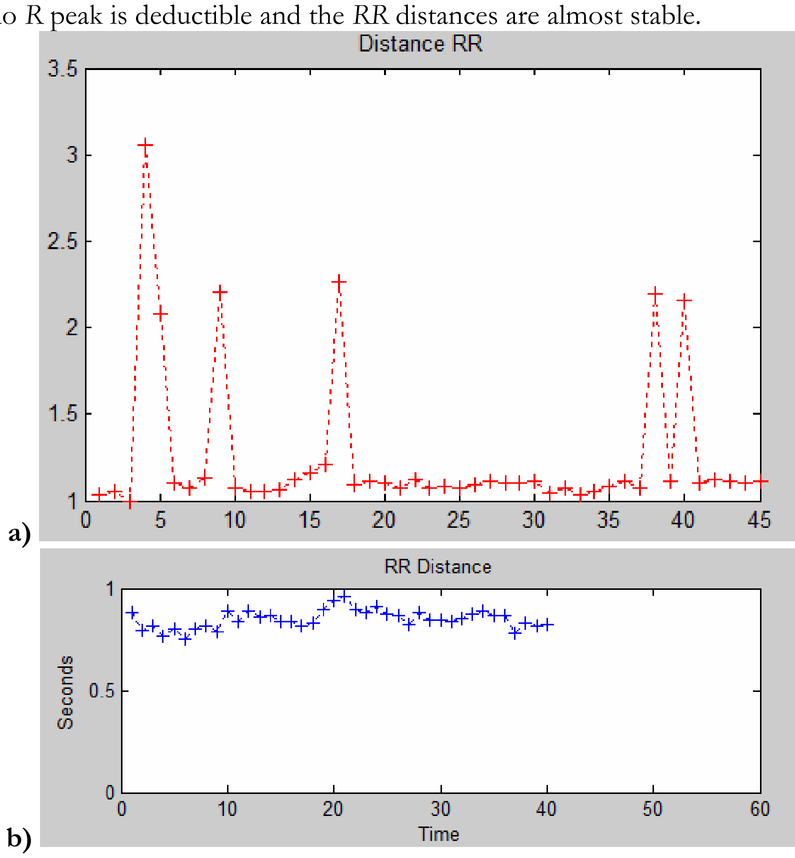

Figure 6. RR distances calculation and a comparison study between the method of Holsinger **(a)** and the method of grammar **(b)** along a portion of 45 seconds

Our future work will focus on the EEG signal and will investigate if the grammar formalism can be applied also for the processing of this signal.

Conclusion

Grammar formalism proved to be useful for ECG signal interpretation and classification. The experiments results have shown that applying our approach on several portions of one real ECG signal, no parameter was franchise and *QRS* complexes, *RR* distances, *PR* and *QT* intervals were calculated. Moreover, a comparative study with the method of Holsinger was established in terms of *R* peaks detection and *RR* distances measurement.

References

1. Ince T, Kiranyaz S, Gabbouj M. A generic and robust system for automated patient-specific classification of ECG signals. IEEE Trans Biomed Eng 2009;56:1415-26.
2. Sahambi JS, Tandon SN, Bhatt RKP. Using wavelet transform for ECG characterization. IEEE Eng Med Biol 1997;16:77-83.
3. Shyu LY, Wu YH, Hu W. Using wavelet transform and fuzzy neural network for VPC detection form the Holter ECG. IEEE Tran Biomed Eng 2004;51:1269-73.
4. Lim JS. Finding features for real time premature ventricular contraction detection using a fuzzy neural network system. IEEE Trans Neural Netw 2009;522-7.
5. Li C, Zheng C, Tai C. Detection of ECG characteristics points using wavelet transforms. IEEE Trans Biomed Eng 1995;42:21-8.
6. Melgani F, Bazi Y. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE Trans Inf Technol Biomed 2008;12:667-77.
7. Ghoggali N, Melgani F, Bazi Y. A multiobjective genetic SVM approach for classification problems with limited training samples. IEEE Trans Geosci Remote Sens 2009;47:1707-18.
8. Khandoker AH, Palaniswami M, Karmakar CK. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings. IEEE Trans Inf Technol Biomed 2009;13:37-48.
9. Kampouraki A, Manis G, Nikou C. Heartbeat time series classification with support vector machines. IEEE Trans Inf Technol Biomed 2009;13:512-18.
10. Kókai G, Csirik J, Gyimóthy T. Learning the syntax and semantic rules of an ECG grammar. Lect Notes Comput Sci 1997;1321:171-18.
11. Trahanias P, Skordalakis E. Syntactic Pattern Recognition of the ECG. IEEE Trans Pattern Anal Mach Intell 1990;12:648-57.
12. John L. Flex & Bison. 1st Ed. USA: O'REILLY, 2009.
13. Holsinger WP, Kempner KM, Miller MH. A QRS preprocessor based on digital differentiation. IEEE Trans Biomed Eng 1971;18:212-17.