MEASUREMENTS WITH NOISE-INDUCED BIOSIGNALS EXEMPLIFIED FOR ECG RECORDINGS[1]

Peter Jaenecke

Abstract: Biosignals arrive a computer system either as a continuous data stream from a sensory input, or as a section of finite length of such a stream recorded on a storage device. Their analysis rests on measurement results gained from recognized waves like duration and amplitude. It is a characteristic feature of biosignals that they are overlapped with noise and other kinds of distortions. Although the signal quality can be improved up to a certain degree by filtering, always residue distortions will remain. A measurement procedure must therefore accepts perturbations. The paper deals with the consequences which follow from this requirement using the ECG measurement as guideline.

One of the main problem – the definition of the quantities to be measured – is illustrated with experiences made in two historical ECG projects. The deficiencies of these projects are discussed and it is shown that they can be overcome by the interval representation of measurements. In this representation the definition of a quantity does not aim at its true value, rather the goal is to find as accurate as possible a lower and upper bound for it. Definitions for wave parameters are given and their performance is tested by simulations using a noise overlapping in the range of a signal-to-noise ration between 5 and 12 dB. The test results show that it is possible to obtain true lower and upper bounds for the durations and amplitudes of waves with quite simple means.

[1] Formerly version: Correct Definition of ECG Wave Onset and Offset. In: HERMANN K. WOLF & PETER W. MACFARLANE (eds.): Optimization of Computer ECG Processing. Proceedings of the IFIP TC 4 Working Conference on Optimization of Computer ECG Processing, Halifax 1979. North-Holland Publishing Company Amsterdam/New York/Oxford 1980, p. 159-164. That version was supported by Bundesministerium für Forschung und Technologie Projekt DVM 125.

Introduction

A lead of an electrocardiogram (ECG) represents a time-dependent biosignal in which an electrical potential difference is plotted against time (Figure 1). The potential difference is registered with electrodes located at different positions of the body surface, e.g. at the extremities; it originates from the electrical field generated by de- and repolarization processes of the heart muscles. The ECG lead shows typical wave forms which can be assigned to the de-, or repolarization of special heart regions (Figure 1).

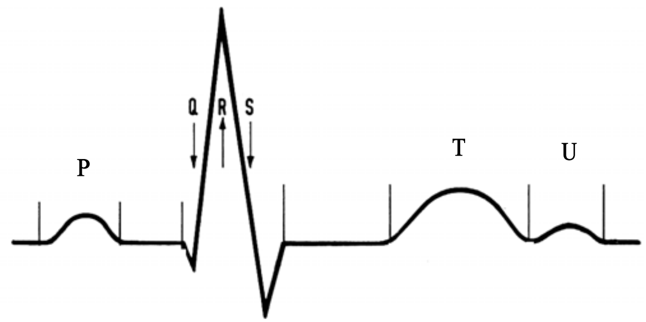


Figure 1: Idealized ECG lead for one heart cycle. The P wave indicates the atrium depolarization, the QRS complex indicates the ventricle depolarization, and the T wave the ventricle repolarization; the U wave is still a subject of debate with respect to its origin and clinical importance.[2]

The computer processing of electrocardiograms starts with a first pattern recognition task for identifying signal deflections. The next step is a measurement procedure in which the deflections' onsets and offsets as well as their amplitudes are determined. Based on these results as next a decision is made, whether a deflection is in fact an ECG wave or not. It follows another pattern recognition task

[2] GUPTA et al. (2005).

for extracting important information from abnormal or striking wave forms (Figure 2). Both the measurement results and the results of the wave form analysis are then the basis for the final interpretation of the ECG lead(s) under study.[3]

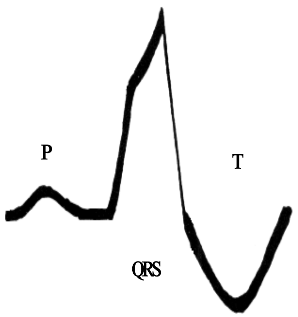


Figure 2: Abnormal QRS complex in case of a left bundle branch block recognizable by its coarse wave form and by its extended duration.

Validation Problems for ECG Systems

Systems for a computerized biosignal analysis have been developed under the assumption that they work faster, more precisely, and more reliably than human observers. But before such systems can be applied to the reality, their results must be proved extreme carefully, especially if – as in case of the ECG analysis – the results are used for recognizing heart diseases so that a misinterpretation may have severe consequences for the persons concerned.

[3] In this short outline of an automated ECG analysis all complications are skipped, e.g., those done by an arrhythmia.

At this point starts a bundle of serious problems around the question: How to define a method with which computer systems can be tested? These problems are not only restricted to the ECG analysis, rather they appear in all systems making decisions on the basis of measurement and pattern recognition.

From the physician's point of view it is correct to say that the quality of an ECG system must be judged by the quality of its diagnoses. Thus, attempts are made to represent the degree of agreement of the computer diagnosis with the clinical one by numerical quality measures like sensitivity, specificity, accuracy, reliability, validity and so on.[4] However, the presence of so many measures for one and the same thing suggests that none of them fulfils its task; none of them provides a unique quality measure. This renders optimization close to impossible, since optimization requires a unique aim.

The reason for the great diversity of concepts seems to be that attempts are made to reduce a multi-dimensional phenomenon to a one-dimensional quantity. But, as is known, this can be done in many ways. It seems that extensions of this theoretical device – like, for example, inclusion of statistical decision theory or information theory, as suggested by RAUTAHARJU et al.[5] – would not be very successful. As in the case of the other approaches this would imply a special weighting of the single components. The strategy, therefore, must be to reduce the multi-dimensionality.

[4] BAILEY et al. (1974); CACERES & HOCHBERG (1970); LAWRENCE (1977); RAUTAHARJU et al. (1976); RAUTAHARJU & SMETS (1979).
[5] RAUTAHARJU et al. (1976).

Scope of the Paper

As mentioned above, automated ECG processing consists of two different parts: measurement and classification. Thus, it suggests itself to treat the accuracy problems of both parts separately. This paper is based on the hypothesis that we cannot consider the quality of diagnosis before clarifying the problems of measurement accuracy. In other words, we first have to set up an accuracy measure for the measurements; thereafter we must look for an accuracy measure for the classification, which should be a function of the measurement accuracy.

In the following we restrict ourselves to measurement problems. We start with a short characteristic of two ECG standardization activities and discuss the problems they remain. As a solution of these problems we introduce the interval representation of measurement results. Some wave parameters are defined in order to illustrate, how this representation can be used for establishing measurable quantities. We test our definitions by simulating measurement procedures with noised signals and summarize in the final section the advantages of the interval approach.

CSE Project

It seems to be self-evident to use for validating purposes a reference library of well documented ECG signals with manually certified results as a standard for the computer systems, assuming that they will work correctly, if they can replicate these results.

This strategy was followed in the so-called CSE project: To establish such a reference library, a large international project was launched to develop "Common Standards for Quantitative Electrocardiography (CSE)".[6] Different techniques are used by the computer programs for measurement and interpretation. Therefore, its main objectives were to reduce the wide variation in wave measurements obtained by ECG computer programs (see Figure 3), and the assessment and improvement of their diagnostic classification. A comprehensive reviewing schemes have been devised for the visual and computer analysis. The task was performed by a board of cardiologists and by programs developed by university research groups and by industry.

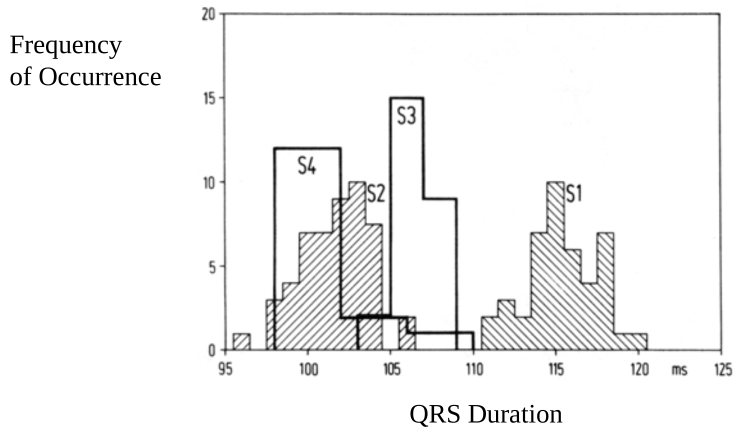


Figure 3: Distributions of the QRS duration for one and the same patient measured with four different commercial ECG processing systems.[7] This is a surprising result because normally the determination of the QRS duration is counted among the easier tasks.

[6] Sponsored by the European Commission in the field of medical and public health research; s.WILLEMS et al. (1990).

[7] From NEUBERT et al. (1980), p. 432.

Standardization Project of the PTB

There was also launched a standardization project at the Physikalisch-Technische Bundesanstalt (PTB), Institute Berlin.[8] Its intentions are very similar to those which initiate the CSE project, however, in opposite to the latter, the theoretical foundation of the test methods is stressed. The approach rests on the assumption that the definition of ECG wave parameters should be based on optimum ECG recordings. Therefore, an appropriate hardware has been developed which represents a standard device for signal sources. It yields real ECG leads in a high resolution quality at a low noise level.[9]

Signals generated with the standard device can be superposed with artificial interferences like noise and base line drifts in order to test the system's stability against those inferences. This procedure allows an error assessment by measuring the wave parameters before and after signal distortion.

It is assumed that the theory seems to be enough advanced to allow the calculation of a model ECG including the electrophysiological processes and their physical appearance. Thus, a special advantage of having available high quality ECG signals is that with them de- and repolarization defects of specific heart muscle regions can be simulated, e.g. for studying the capacitive and resistive current in various tissues and anisotropic conductivity effects.[10] This feature makes it possible to decide whether a small deflection is a small ECG wave or an artifact.

[8] NEUBERT et al. (1980).
[9] TEPPNER et al. (1987), p. 436f.
[10] NEUBERT et al. (1980), p. 434.

Remaining Problems

Both the PTB project with its focus on solving measurement problems, and the deserving CSE project, exemplary for all alike computerized tasks and unique in its effort and in its careful realization, have left some unresolved problems: (i) Defining measurement requirements and establishing a basic library are only necessary, but not also sufficient actions, i.e., they are needed, however they do not guarantee a success; the question arises how to justify them? (ii) Definitions of the quantities are missing which take into account the presence of noise, and with which the program developer can assess the performance of his system's results. (iii) There is a historical burden in the sense that all the clinical knowledge stored in the textbooks was obtained by manual measurements. The computerized analysis, however, provides new chances for operations not manually executable; they open new fields of experience, but up to now less expert knowledge exists about them. Thus, there is the risk that the computer systems are adapted to the traditional knowledge and that their possibilities are left unused at long sight.

We now look at some aspects of these remaining problems in more detail.

The Problem of Finding an Adequate Standard

For measurements we need a standard procedure to provide us with correct results. Only then can we evaluate an algorithm's accuracy in terms of the distance between its results and the supposedly correct ones. But this is a problematical method. To check the accuracy of a standard, another still more accurate standard is needed and so on.

In checking automated ECG measurements it is customary to employ manual measurements by a human observer (e.g. the program developer) as a standard. However, it turned out very quickly that there appear both inter-observer and intra-observer variability.[11] The reason is clear: the determination, e.g. of a wave duration, is a measurement done by a person as measurement device. All real measurements are subject of measurement errors so that differences in the measurement results are a natural consequence.

In the CSE project experienced cardiologists are consulted for doing this task. It is assumed that they do the right. The variability was reduced by an interactive reviewing process. Each of the four rounds of the Delphi type reviewing process led to smaller variances. However, the determination of these quantities depends above all on the visual ability, i.e. no special *medical* knowledge is required to identify a deflection of a curve. If the measuring results of human observers should be used as standard, then the question arise: How exact are such visual measurements?

The Problem of Finite Resolution

It seems to be intuitively clear that, e.g. the onset of a wave is just that point, the curve leaves the base line; and this occurrence should be clearly visible. However, this argument applies only in part, since of what can be seen depends not only on the skill of the human eye, but also on the resolution of the ECG plot.

As investigations at our institute[12] have shown, the P wave duration determined by the physician during routine work can be so inaccurate that it becomes useless in detecting obvious errors in the program. The main reason is the unsatisfactory resolution of the ECG plot (see Figure 4).

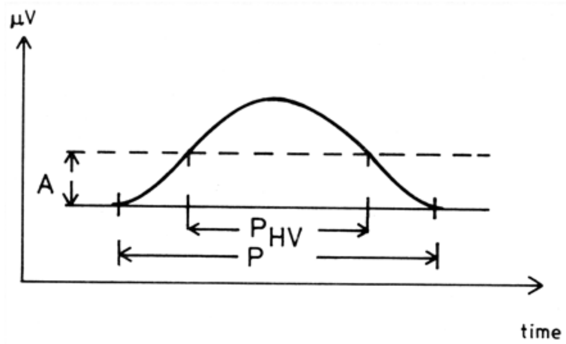


Figure 4: Idealized P wave with a positive amplitude. *P* marks its "true" duration measured at a high resolution. In a real ECG plot the resolution is lower; there can be identified a resolution threshold *A*, i.e. the minimal distance that a curve must depart from the baseline in order to be recognized visually by an observer as a deflection onset or offset. In the figure above a human observer can only see the deflection above the dotted line. Due to this finite resolution, the manually determined duration P_{HV} is only about 50% of the exact duration *P*.

The only way to improve the resolution of an ECG plot is to enlarge the scale. How should the scale be chosen in order to obtain the best possible resolution?

Figure 5 sketches the error of manual measurements as a function of the ECG plot scale for different noise levels *R* (i.e., the mean deflection of the noise signal from the baseline). If *R* is low, the error is determined mainly by the resolution, and the error curve takes an asymptotic course (curve 1). As the noise level becomes higher, an enlargement of scale at first continues to cause a decrease of error, but then the error starts to increase once more. The magnified noise signal renders exact determination of the wave

[11] WILLEMS et al. (1983), p. 51.
[12] Institute for Medical Informatics, Gießen University, Germany.

onset and offset more and more doubtful, and there is an optimal scale which minimizes the error (curve 2). The higher the noise, the more the signal flattens and the optimal scale with minimal error decreases (curve 3). From this it follows that the best resolution for an observer depends on the noise level.

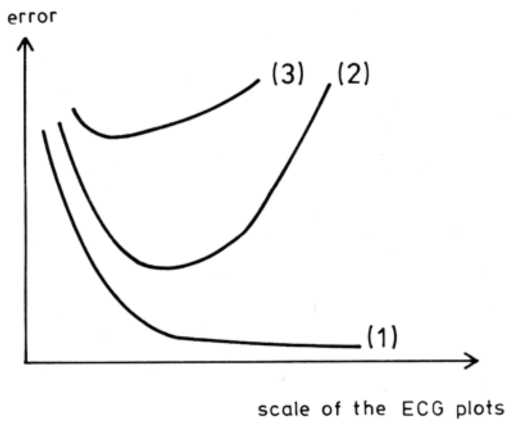


Figure 5: Error of manual measurements as a function of the scale of the ECG plot for different noise levels (see the text for more details).

The Problem of the Historical Burden

The transition to a computerized ECG analysis should not mean that just those procedures are automated which were afore done by hand. Rather, the computerized measurement opens new chances with respect to accuracy and reliability of the measurement results, with respect to a nearly unlimited memory for storing case studies and ECG wave forms, and last but not least with respect to a high number of feature combinations for characterizing diagnostic findings. Thus, we have the situation that the computer systems could provide results for which no application is found in the textbooks.

On the other side, ECG analysis has a long tradition at which a large amount of knowledge was gathered by means of manual analyses. This knowledge however, deposed in numerous textbooks, refers only to the striking wave forms whereas the meaning of the finer wave form variations is unknown. Often the opinion is held that such small variations have no clinical relevance. But this assertion was never proved systematically, it rests only on the experience with the textbooks in which such variations are not mentioned. It may be true that a unobtrusive wave form – considered for itself – has no importance, but this must not be true in combination with other (possibly also unobtrusive) features. There is the problem to combine the traditional knowledge with the results a computer system can extract from ECG leads.

Up to now the potential provided by the computer analysis is left unused to a large extent. In order to utilize it, special methods and new quantities are needed with which the corresponding refined medical knowledge can be acquired. We skip this topic in the following considerations.

The Problem of Different Measurement Practices

Only that can be measured manually what was recognized, and it can be reached only that accuracy which is possible by means of the sensory perception and by means of the enabling instrumentation.

A computer does not depend on these restriction: it is aware of each difference of two numbers, the small the difference may be, i.e. in principle a computer can be more exact than a human observer. From this it follows that visual measurement cannot be a reliable accuracy standard for automated measurements. Moreover, the computer's power of resolution can far exceed that of a human observer, especially if the signal is filtered before wave recognition.

For a computer system it seems to be, therefore, not a good strategy to come as close as possible to the manually measured results, because this would mean that the imperfectness of the manual measurement has to be replicated by the computer system. Doing so, one would not only give away the chance for a better result, but also the computer would be in an inferior position with respect to a human observer.

On the other side manual results are needed for comparing them with the computer results. From it arise a conflict situation: on the one hand needed, on the other hand suitable only to a limited extent. However, there is no other procedure to replace visual measurement. It is an object of this paper to solve this conflict.

The Problem of Missing Methods for the Program Developer

Another problem left unresolved by the standardization projects is that methods are missing with which a program developer can determine the implication of his program modification. He is always confronted with the question: Given are two algorithms, which of them is the better one? The answer to this question requires clear measurement *rules* which include a definition of wave onset and offset. The quality of a system can be tested with a model library having standardized measurement *results*. But in case the computer results are pure, such a library is of little value for finding out a better approach.

Consequences from the remaining problems

What we need are clear definitions of the quantities which should be measured. These definitions must take into account the fact of signal inferences. If such definitions are available then the problem of finding an adequate standard and the problem of different measurement practices are eased, because both the human observers as well as the computer systems must follow them. The problem of finite resolution vanishes, because the signal interferences are incorporated in the definitions, and also the program developers is helped because computerized test methods can be derived from the definitions.

Interval Representation of Measurement Results

The difficulties described above arise from a flaw in the methodology: a quantity which is known only approximately is represented as a number, that is, as a unique mathematical object. Therefore we will change the mathematical representation of a measurement. Henceforth we will describe it not as a number but as an interval

$$\mathbf{M} = \{x \mid a \leq x \leq b\} = [a, b],$$

where a is a lower and b is an upper bound for the true value. It should be noted that any element of **M** can be the true value of the quantity in question. No element is distinguished and the central point of the interval has no special meaning.

Let W be the true value and **M** the measurement interval of some quantity. We call **M** a *true measurement interval,* if $W \in \mathbf{M}$; otherwise **M** is a false measurement interval. It is always possible to obtain true measurement intervals by expanding the interval. But widening the interval will lower its information content. Thus the length $L(\mathbf{M})$ of a true measurement interval **M** provides a measure of the uncertainty – or, the other way round, of the accuracy – of a single measurement.

A necessary condition for a standard is independence of individual peculiarities, which is equivalent to reproducibility. While it is impossible to determine a wave onset (or offset) exactly, it is perfectly feasible to find lower and upper bounds for the true wave onset (or offset) for both the human observers and the computer systems. Doing so will considerably reduce the differences between human observers and computer systems.

Arithmetic operations for intervals have been already defined.[13] Moreover, the set theoretical intersection of intervals may serve to "average" measurement intervals with the following remarkable properties:

Let $\mathbf{M_1}, \ldots, \mathbf{M_n}$ be n ($n \geq 1$) true measurement intervals, $L(\mathbf{M_i})$ the length of the i-th interval, and

$$(1) \qquad \bar{\mathbf{M}} = \bigcap_{i=1}^{n} \mathbf{M_i} \,.$$

Then:

(a) $\bar{\mathbf{M}}$ is also a true measurement interval.

(b) $\bar{\mathbf{M}}$ does not depend explicitly on n.

(c) $L(\bar{\mathbf{M}}) \leq L(\mathbf{M_i})$ for $i = 1, \ldots, n$.

(d) No specific error distribution must be assumed.

(e) $\bar{\mathbf{M}}$ cannot be empty (if $\bar{\mathbf{M}} = \emptyset$, then at least one $\mathbf{M_i}$ must be a false measurement interval, which contradicts our assumption).[14]

Strategic Interval Approach Aspects in Defining a Quantity

We have clarified the mathematical representation of the objects which we are about to define. That is, we have set up the formal concepts to be used in the definitions. The specific features of the interval approach require a new way of thinking in defining wave quantities: In measuring a quantity no longer a single value is searched for as an approximation for the so-called true value of the quantity, rather the aim is now to assess as accurate as possible a lower and upper bound for it.

First of all, these bounds must be introduced in the definitions in such a way that the measurement intervals will be true when the rules of the definition are followed correctly. As mentioned above, true measurement intervals can be always produced by extending the interval length so that there is an infinite number of assessments/ definitions. Because we like to have interval lengths as short as possible, suitable margins of tolerance must be installed in the definitions.

We explicate this strategy with the measurement of a deflection's positive amplitude defined as the maximum value W_{max}. Because of the noise overlapping the measured maximum $W_{max}{}^{m}$ will deviate from the true maximum W_{max}. Assume, that there is a maximum noise amplitude R_b; to ensure that the true value of the amplitude will lie in the interval, we assume a tolerance limit $\pm$ 1.5 R_b so that the interval result is

[13] MOORE (1966); SUNAGA (1958).

[14] JAENECKE (1982), p. 150-160.

$$\mathbf{W_{max}} = [W_{\max}{}^{m} - 1.5R_b,\ W_{\max}{}^{m} + 1.5R_b]\,.$$

It may be argued that the tolerance limit can yield intervals too large for a useful interpretation. Such large intervals corresponds to what is called normally as outliers. But outliers have (in opposite to an averaging with the arithmetical mean) no influence to the result of an interval averaging according equation (1), as long as they include the true value.

Computerized Versus Manual Measurement Rules

It would be the best if the same definitions could be applied to the computerized as well as to the manual measurement. But "smoothing" a wave, e.g., is done visually by see-sawing the eyes, and tolerances are estimated according the rule of thumb. A computer system needs clear instructions for this task which are useless for a human observer. This reveals the fact that a computerized measurement is more sophisticated and therefore also more demanding so that the hope of definitions for the both kinds of measurements must be abandoned.

If manual measurement should provide an accuracy standard, then the rule of thumb must be restricted by fixed operations as counterpart to the computerized operations. By nature they are complicated and cumbersomely to process.

For practical reasons and to solve the above mentioned conflict with the different measurement practices, we allow the manual measurements therefore only a control function, i.e., we assume that the visual ability of a human observer suffices for yielding true measurement intervals and for deciding whether a computer result is correct, whereas the accuracy problems are considered exclusively as a subject for computer systems.

Note that additional conventions are required if visually determined measurement intervals should meet its control function. For the sake of simplicity we skip here statements of particulars.

Definition of Some ECG Quantities

We restricted ourselves to give rules for computerized measurements; the goal is to show, how the interval approach can be applied.

Any clear deviation from the base line is called a 'deflection'. Whether a deflection is really an ECG wave has to be decided by means of minimum wave requirements after the measurement procedure.

We suppose that there are already raw reference points for the begin and end of a deflection. These points may be the output of a wave recognition program; they isolate two different areas of an ECG lead: a deflection and a noise area. A deflection area contains a deflection; a noise area does no contain a deflection, i.e. it situated between two deflection areas. The ECG lead sections shown in the Figure 6 - Figure 9 represent such areas.

Amplitudes are described by x_n, or by $x(n)$, where n (n = 1, 2, 3, ...) indicates a sample index. Any durations will be expressed in number of sample points, i.e., they follow from the difference of two indices. Since the sample rate is known, e.g. 2 ms, the time duration can be calculated directly from such a difference.

Because outliers have no influence on the mean (1), absurd results are trapped by setting them to the worst possible value.

Assessment of the noise parameters

The noise levels, R_b *and* R_a*, are defined as the mean upper and lower deflection of the noise signal enlarged by a margin of safety.*

Let n_1 be the reference point for the end of the preceding deflection and n_2 the reference point for the begin of the following deflection, i.e., it is assumed that between sample n_1 and sample n_2 there is no potential ECG wave. The samples $x_{n_1},...,x_{n_2}$ are subdivided into samples $\xi_1,...,\xi_{n_p}$ having a positive amplitude and into samples $\eta_1,...,\eta_{n_n}$ having a negative one. The means are given by

$$\overline{R}_a = \frac{1}{n_n}\sum_{k=1}^{n_n}\eta_k \quad \text{and} \quad \overline{R}_b = \frac{1}{n_p}\sum_{k=1}^{n_p}\xi_k \,.$$

As margins of safety we use the variances

$$\sigma_a = \frac{1}{n_n}\sum_{k=1}^{n_n}{\eta_k}^2 \quad \text{and} \quad \sigma_b = \frac{1}{n_p}\sum_{k=1}^{n_p}{\xi_k}^2$$

so that the noise levels for the noise area [n_1, n_2] are given as

$$R_a = \overline{R}_a - \sigma_a \quad \text{and} \quad R_b = \overline{R}_b + \sigma_b \,.$$

The signal-to-noise ratio of a deflection area

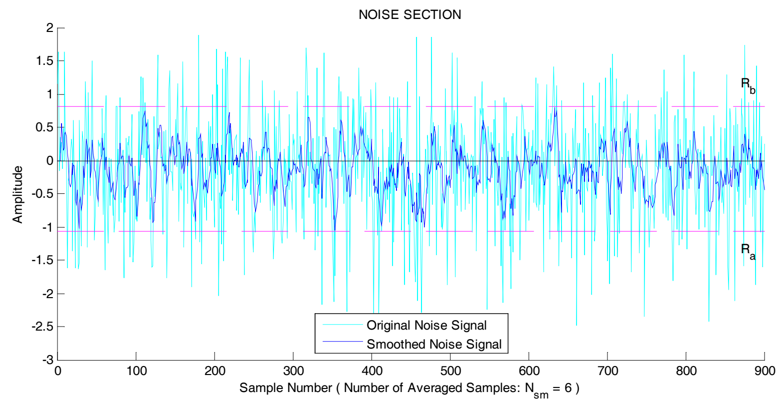
$$(2) \qquad R_{SN} = 10 * \log_{10}\left(\frac{\sum\limits_{n \in deflection\ area} {x_n}^2}{\sum\limits_{n \in preceding\ noise\ area} {x_n}^2}\right)$$

indicates the amount of noise with respect to the ECG signal.

Assessment of the smoothing parameter

The smoothing procedure is defined as the arithmetical mean

$$(3) \qquad \overline{s}(n) = \frac{1}{N_{sm}}\sum_{j=n-N_{sm}+1}^{n} s(j) \quad n = N_{sm}, N_{sm}+1, \ldots \,.$$



It should be noted that this smoothing delays the signal at about N_{sm} samples.

The smoothing parameter, N_{sm}, specifies about how much samples the averaging has to be performed in order to smooth the signal. It is defined as follows:

N_{sm} *is the number of samples about which a noise section has to be averaged in order that the smoothed noise signal lies between the noise levels* R_a *and* R_b (Figure 6).

Figure 6: Smoothed and un-smoothed noise area. R_a and R_b (dotted lines) are the noise levels between which the smoothed noise signal must lie. To reach this requirement, the original noise signal was averaged with N_{sm} = 6 samples where N_{sm} is the smoothing parameter.

Smoothing a signal in a deflection area

Let n_1 be the reference point for the begin of a deflection and n_2 the reference point for its end, i.e., it is assumed that the samples n_1 and n_2 delimit a deflection area. A signal in this area is overlapped in nature by noise; it has to be smoothed, therefore, before the measurement procedure can start, i.e., following equation (2) it has to be averaged about N_{sm} samples according to

$$\bar{s}(n) = 0 \quad n = n_1,\ n_1 + 1,\ \ldots,\ n_1 + N_{sm} - 1;$$

$$\bar{s}(n) = \frac{1}{N_{sm}} \sum_{j=n-N_{sm}}^{n} s(j) \qquad n = n_1 + N_{sm}, N_{sm} + 1, \ldots, n_2;$$

N_{sm} is the smoothing parameter determined from the preceding noise area (see Figure 7 and Figure 8 for an example).

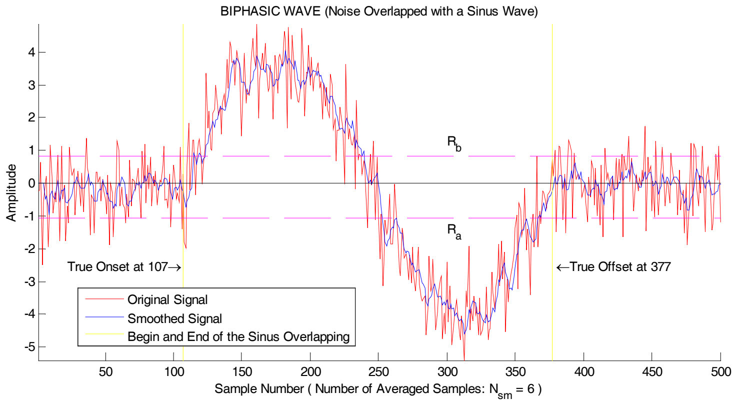
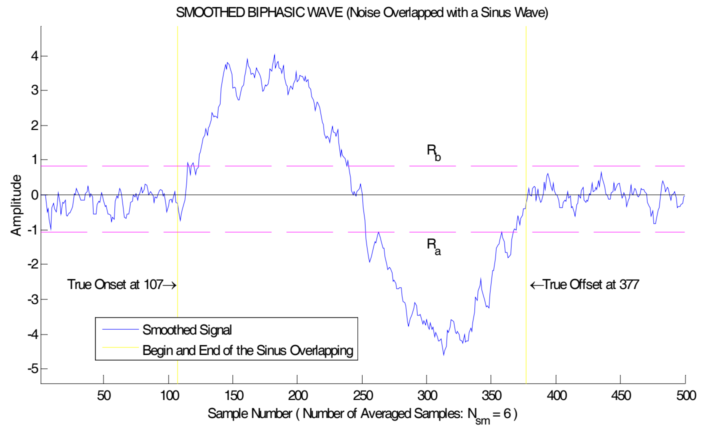


Figure 7: Smoothed and un-smoothed deflection area. The original wave signal was averaged with N_{sm} = 6 samples where N_{sm} is the smoothing parameter determined by means of the noise levels R_a and R_b. The signal was generated by overlapping a noise signal with a sinus wave which should represent a potential biphasic ECG wave. Because the location of the sinus overlapping is known, a true value for its begin and end can be given (see yellow lines).

Figure 8: Smoothed deflection area. It is the same signal as shown in Figure 7.



Assessing a lower and upper bound for deflection onset and offset

Let be $s(n_{max})$ = W_{max} the maximum and $s(n_{min})$ = W_{min} the minimum amplitude of the signal in a deflection area.

(a) $W_{max} > R_b$ (positive deflection)

It assumed that the amplitude W_{max} belongs to a positive deflection. At first, a raw upper bound for the begin of a deflection, and a raw lower one for its end have to be located. The starting point is in each case the maximum position n_{max} of the curve. From this position the curve is followed to the left and to the right side; the idea is to fix the bounds onto the curve's first intersection points with the line R_b. The definitions are:

$n_{begin}{}^{u} / \nu_{begin}{}^{l}$ is the raw upper bound/preliminary lower bound for the begin of a deflection, if

$$x_{n_{begin}{}^{u}} \ge R_b \text{ and } x_{n_{begin}{}^{u}-1} < R_b \,/\, x_{\nu_{begin}{}^{l}} \ge R_b \text{ and } x_{\nu_{begin}{}^{l}-k} < R_b \quad k \ge 2N_{sm},$$

and $n_{end}{}^{l} / \nu_{end}{}^{u}$ is raw lower bound/preliminary upper bound for its end, if

$$x_{n_{end}{}^{l}} \ge R_b \text{ and } x_{n_{end}{}^{l}-1} < R_b \,/\, x_{\nu_{end}{}^{u}} \ge R_b \text{ and } x_{\nu_{end}{}^{u}+k} < R_b \quad k \ge 2N_{sm}.$$

The definition of the outstanding raw lower/upper bound for the begin/end of a deflection is fixed onto the idea (i) to construct two straight lines as a surrogate for the real curve progression, and (ii) to use the lines' intersection points with the base line for assessing the bounds (see also Figure 9):

Given is $W = \begin{cases} W_0 & \text{if } W_0 > R_b \\ W_{\max} & \text{if } W_0 \le R_b \end{cases}$, *where* $W_0 = W_{\max} - 2R_b$. *Let be g(x) a straight line passing the points*

$(\nu_{begin}{}^{l}, R_b), (n_{max}, W),$

then $n_{begin}{}^{l}$ is a raw lower bound for the begin of a deflection, if

$$g(x_{begin}{}^{l}) < 0 \text{ and } g(x_{begin}{}^{l} + 1) \ge 0 .$$

Let be g(x) a straight line passing the points

$(n_{max}, W), (\nu_{end}{}^{l}, R_b),$

then $n_{end}{}^{u}$ is a raw upper bound for the end of a deflection, if

$$g(x_{end}{}^{u} - 1) \ge 0 \text{ and } g(x_{end}{}^{u}) < 0 .$$

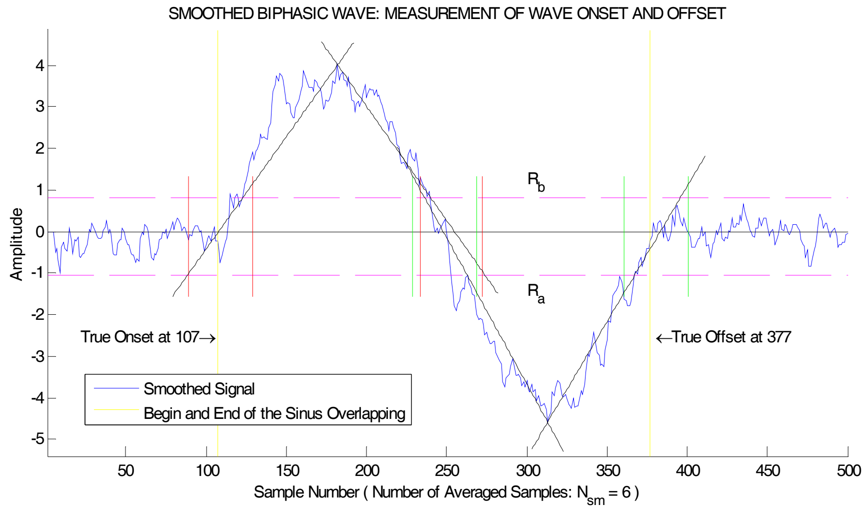


Figure 9: Estimation of the lower and upper bounds for onset and offset of a deflection (see the text for more details).

These values are raw, because no signal interferences are taken into account as yet in terms of margins of safety. The most problematical part is to find a lower bound for the onset and an upper bound for the offset because at these positions a deflection loses itself in the noise interference. This fact was already taken into account in part in constructing the auxiliary straight lines not from the maximum W_{max}, rather from $W_{max} - 2R_b$. Additionally, we add to them the margins of safety $\pm N_{sm}$:

The onset and ofset of a deflection are given by

$$\mathbf{N}_{\mathbf{onset}} = [n_{begin}{}^{l} - 3N_{sm},\ n_{begin}{}^{u}]$$

and

$$\mathbf{N}_{\mathbf{offset}} = [n_{end}{}^{l} - N_{sm},\ n_{end}{}^{u} + 3N_{sm}].$$

(b) $W_{min} < R_a$ (negative deflection)

The onset and offset of a negative deflection can be defined in an analogous fashion in using the minimum $s(n_{min}) = W_{min}$ instead of the maximum $s(n_{max}) = W_{max}$.

(c) $W_{max} < R_b$ or $W_{min} > R_a$ (no deflection)

In both cases the extrema lie within the noise levels; they are therefore not considered as a deflection.

Assessing the amplitudes of a deflection

In the ideal case the amplitude of a deflection is defined by its highest deviation from the base line. Again to this idea a margin of safety has to be added (see Figure 10).

If $W_{max} > R_b$, then

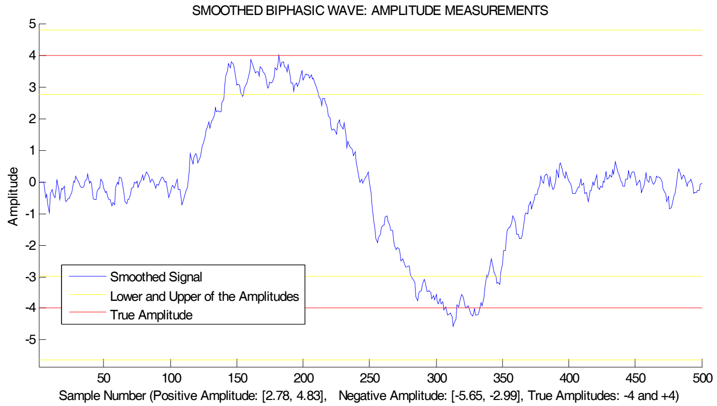
$$\mathbf{A}_{\mathbf{pos}} = [W_{\max} - R_{SN}\ R_b / 4,\ W_{\max} + R_{SN}\ R_b / 5]$$

is the amplitude of a positive deflection; if $W_{min} < R_a$, then

$$\mathbf{A}_{\mathbf{neg}} = [W_{\min} + R_{SN}\ R_a / 6,\ W_{\min} - R_{SN}\ R_a / 4]$$

is the amplitude of a negative one.

Figure 10: Estimation of the amplitudes from a deflection (for more details see text).



Onset and offset of a deflection for a lead group

The above definitions refer to a single ECG lead. However, normally a lead group is simultaneously recorded, i.e., there are three, six, twelve or more ECG leads available showing the same cardiac activity, but recorded from different body surface positions. The electrical field propagates itself in the three-dimensional space, but a single electrode can only capture the electric field strength at one point in this space. As a consequence, one and the same deflection differs in these leads with respect to its amplitude and its onset and offset. It may be that in one lead, e.g. the P wave, is small and has only a short duration, whereas in another lead the same wave is clear formed with distinct onset and offsets.

It seems to suggest itself to use just the latter for measurement purposes. Unfortunately, it is not known which of the lead will be the dominating one for a given deflection. In following the interval strategy, the earliest onset and the latest offset found in the leads of a lead group must be taken as the group onset, respectively offset. The following definition clarifies the terms ‘earliest’ and ‘latest’:

Given are from one and the same deflection the interval onsets

$$\mathbf{N}_{\mathbf{onset}}{}^{1} = [n_{lb\_onset}{}^{1}, n_{ub\_onset}{}^{1}], \ldots, \mathbf{N}_{\mathbf{onset}}{}^{j} = [n_{lb\_onset}{}^{j}, n_{ub\_onset}{}^{j}]$$

and the interval offsets

$$\mathbf{N}_{\mathbf{offset}}{}^{1} = [n_{lb\_offset}{}^{1}, n_{ub\_offset}{}^{1}], \ldots, \mathbf{N}_{\mathbf{offset}}{}^{j} = [n_{lb\_offset}{}^{j}, n_{ub\_offset}{}^{j}]$$

measured for $j > 1$ ECG leads. It is assumed that these intervals are true. Let be

$$n_{lb\_onset}{}^{k} = \min(n_{lb\_onset}{}^{1}, \ldots, n_{lb\_onset}{}^{j})$$

and

$$n_{ub\_offset}{}^{l} = \max(n_{ub\_offset}{}^{1}, \ldots, n_{ub\_offset}{}^{j}) .$$

Then the onset of ECG lead k is the earliest onset, and the offset of lead group l is the latest offset of a deflection. They define the lead group onset and offset as follows:

$$\mathbf{N}_{\mathbf{onset}}{}^{lead\ group} = [n_{lb\_onset}{}^{k}, n_{ub\_onset}{}^{k}],$$
$$\mathbf{N}_{\mathbf{offset}}{}^{lead\ group} = [n_{lb\_offset}{}^{l}, n_{ub\_offset}{}^{l}].$$

This result is used for calculating the duration of a deflection.

Note that k and l are different in general because an ECG lead having the earliest onset of a deflection must not have also its latest offset.

Duration of a deflection and the distance between to deflections

A duration is defined in the ideal case as the difference of two sample indices, where the event belonging two the last index is the later one. In transposing it into the interval formalism the difference of two intervals

$$[a_2, b_2] - [a_1, b_1] = [a_2 - b_1,\ b_2 - a_1]$$

is used:

Let be $\mathbf{N}_{\mathbf{onset}}{}^{lead\ group} = [n_{lb\_on}, n_{ub\_on}]$ *a lead group onset, and* $\mathbf{N}_{\mathbf{offset}}{}^{lead\ group} = [n_{lb\_off}, n_{ub\_off}]$ *a lead group offset of a deflection, then its duration is given by*

$$\mathbf{D} = [n_{lb\_off} - n_{ub\_on},\ n_{ub\_off} - n_{lb\_on}].$$

Let be $\mathbf{N}_{\mathbf{offset\_A}}{}^{lead\ group} = [n_{lb\_off}{}^{A}, n_{ub\_off}{}^{A}]$ *a lead group offset of deflection A, and* $\mathbf{N}_{\mathbf{onset\_B}}{}^{lead\ group} = [n_{lb\_on}{}^{B}, n_{ub\_on}{}^{B}]$ *a lead group onset of a deflection B which is later than deflection B, then the distance between these deflections is given by*

$$\mathbf{D}_{\textbf{between A and B}} = [n_{lb\_on}{}^{B} - n_{ub\_off}{}^{A},\ n_{ub\_on}{}^{B} - n_{lb\_off}{}^{A}].$$

Deciding whether or not a deflection is an ECG wave

The following minimum wave requirements are derived from rather clean signals: A wave has to have at least an amplitude of A_{min} = 20 µV and a duration of d_{min} = 6 ms in order to be recognized reliably.[15] These guidelines have to be alienated into the interval approach as follows:

[15] WILLEMS et al. (1983), p. 201f; (1984) p. 155.

Let **D** *the duration of a deflection and* **A** *its amplitude. Then the deflection will be accepted as ECG wave, if* $d_{min} \in \mathbf{D}$ *and* $A_{min} \in \mathbf{A}$.

Performance Test for the Definitions

The measurement requirements described above are tested in simulating the repeated occurrence of one and the same deflection under different noise levels in a single lead. The true deflection is a sinus wave with a duration of d_{sinus} = 270 samples. It stands for a biphasic ECG wave having the amplitudes A_{max} = +4 and A_{min} = – 4. The noise signal was calculated with a standard pseudo random generator.

As is generally known, in the ideal case the ECG waves replicates themselves in each heart cycle. Therefore, the longer the recording time lasts, the more ECG waves of the same sort are available. This situation conforms to that in physics in which a quantity is measured repeatedly several times under nearly the same circumstances. The result is a series of measurement to which an averaging (e.g. the arithmetical mean) and an error estimation (e.g. the standard deviation) is applied.

We simulate now this procedure with ten different tests characterized by its special signal-to-noise ratio. A test consists in N measurements in which the sinus wave is overlapped in each case with another noise signal of the same signal-to-noise ratio. The quantities to be measured are time duration as well as positive and negative amplitude. The N measurement results of a test are averaged according to equation (1). The results are shown in Figure 11 and Figure 12.

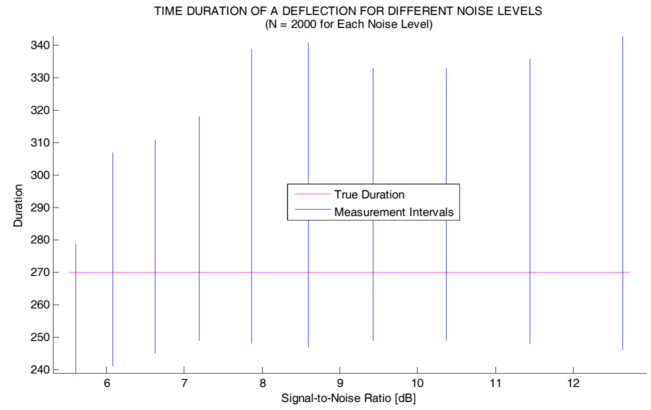


Figure 11: Averaged time duration of a simulated single ECG lead against the signal-to-noise ratio. *N* = 2000 gives the number of measurements performed in a test. The averaging was done by intersecting the *N* durations obtained in each measurement procedure.

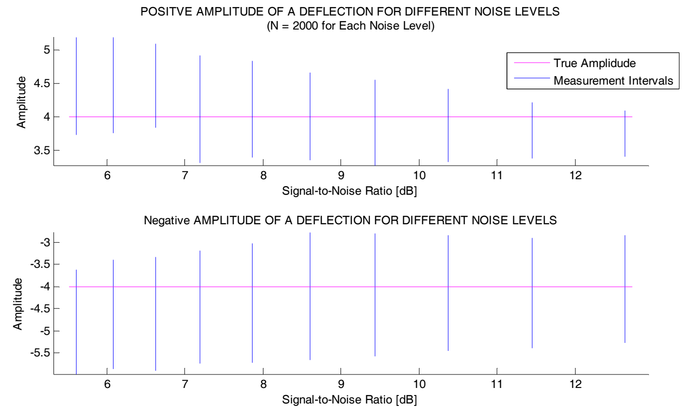


Figure 12: Averaged amplitudes of a simulated single ECG lead against the signal-to-noise ratio. *N* = 2000 gives the number of measurements performed in a test. The averaging was done by intersecting the *N* durations obtained in each measurement procedure.

Discussion of the Test Results

The test results show that it is possible to achieve a good performance in using relatively simple assessments: The measurement intervals remains always true, even though the interval lengths increase as expected with a decreasing signal-to-noise ration.

It may be argued that the simulation is based only on artificial signals so that the results are not very significant. But quite the opposite: The tests are very challenging from different reasons.

(i) The true values of the quantities in question are known so that also the true error is known – a situation which is never given with real ECG leads, but it is very helpful for a program developer because he can identify now the deficiencies in his programs.

(ii) The averaging defined as the intersection of measurement intervals is a very hard approach, since exactly one false interval can cause an empty intersection and thus a useless result. In such a way unexpected situations will encounter not even thought about them.

(iii) The simulation can be done with a nearly unlimited number of runs which would be impossible in using real ECG leads. The large number of runs has a special importance in the interval approach: If the averaging yields an empty interval, then it is proved that at least one interval must be false. Unfortunately, the reverse does not hold: If the mean is non-empty, then it is not sure that it will be also a true mean. The interval length of the mean value defined by equation (1) decreases, or, at the most, remains equal, the more *true* intervals are averaged. That means: if the measurement requirements are yet imperfect then the chance that it produces a false interval increases the more measurement intervals are averaged.

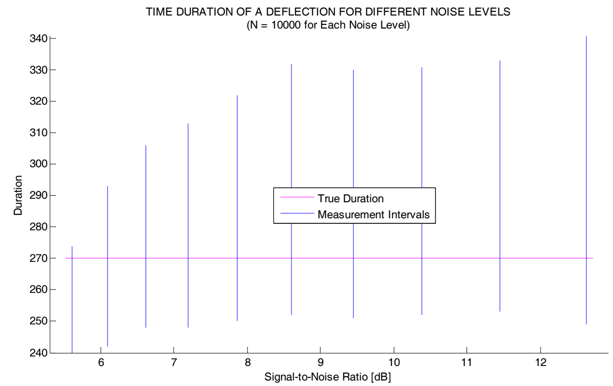


Figure 13: Averaged time duration of a simulated single ECG lead against the signal-to-noise ratio. There is the same situation as shown in Figure 11, however the number of measurements performed in the test is now $N = 10000$.

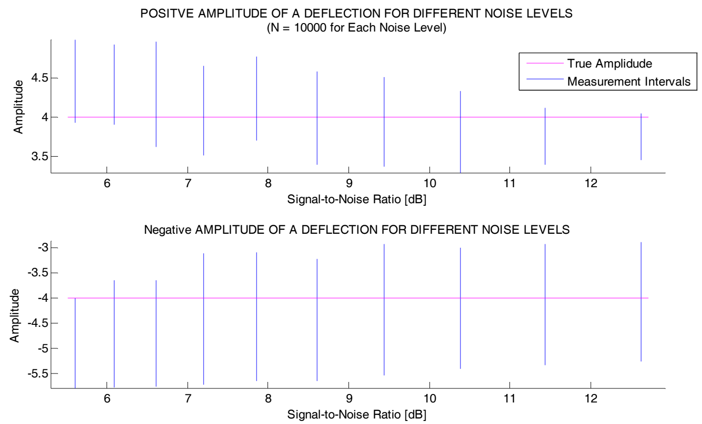


Figure 14: Averaged amplitudes of a simulated single ECG lead against the signal-to-noise ratio. There is the same situation as shown in Figure 12, however the number of measurements performed in the test is now $N = 10000$.

This effect is a great help for a program developer for refining his algorithms; it is demonstrated in Figure 11 – Figure 14. Whereas in case of N = 2000 all time durations are true intervals (Figure 11), there is in case of N = 10000 one false amplitude (Figure 14) indicating that the definitions are not yet ideal. The durations are true intervals for N = 2000 as well as for N = 10000 indicating a more succeeded assessment for them, but because the results are very similar in both case, it can be judge that the intervals can be shortened with a more appropriate assessment.

In processing real ECG leads more interferences has to be observed. One of them is the base line drift done by respiration or other non-cardiac muscle activities. Therefore, the above definitions are not unique, and it would be feasible to conceive of another set of definitions to ensure true measurement intervals also in case of these interferences. The modified definitions have to be tested again by simulation. In the example above only a model for noise was used. In the same manner now additional models for the base line drift and the other interferences must be found with which artificial interferences can be generated for the test procedure.

The program systems should not be applied to real ECG leads until the simulations could be terminated successfully. The tests by simulations are in a certain sense certification procedures based on theoretical assumptions about the signal properties. These non-medical assumptions enter into the definitions; they are necessary for a program developer for his work, but they cannot be derived from the results of a standard ECG library.

However, the latter can be compared now with the results from the simulatively certified programs. This comparison will give information about the quality of both the computer results and those obtained by human observers. It is a chance to learn from each other and to improve the techniques in both worlds.

Conclusion

Up to now there has been no consistent definition of ECG wave onset and offset and the wave recognition algorithms have been based more or less on intuitive ideas about wave onset and offset. Thus the contradictory results reported in the literature are not surprising. For example, DOBROW et al.[16] report good agreement between visual and automated measurement, while on the other hand, great differences between the results of different ECG Systems have been described.[17] It is an old statement from the measurement theory that measurement results are worthless, unless they are accompanied by some error information. We therefore replaced numbers by intervals in representing measurements. It can be shown that true measurement intervals (containing the true value as an element) offer the following advantages:

(1) The discrepancies between visual and automated measurement can be removed. As the error information is contained in the measurement interval itself, the visual measurement is no longer needed as an accuracy standard, but only as a check against false measurement intervals. However, a human observer is generally able to decide whether an onset or offset is contained in some interval or not.

(2) The influence of the individual differences between the human observers will be reduced.

(3) For each measurement interval its length yields information on its accuracy. There are several consequences:

[16] DOBROW et al. (1965).
[17] WILLEMS & PARDAENS (1977); NEUBERT et al. (1980), p. 432.

(i) Resolution and signal perturbation can be taken into account.

(ii) The interval length provides a point of departure for further optimization.

(iii) Signal perturbations will no longer cause false, but just imprecise results. The resulting diagnosis may be uninformative, but not erroneous.

(4) Measurements and diagnoses can be uncoupled. Any diagnostic procedure based on the input of true measurement intervals should be compatible with any measurement procedure yielding true measurement intervals as its output.

The subject matter described in this paper is an old one, but it is still going on. The ECG standardization efforts and the efforts in defining the measurement quantities show, how much preparatory work must be done before an operable system can be established. In the naïve storm and stress period of the artificial intelligence apparently these efforts are underestimated. This may be one reason why a lot of formerly high-praised systems today are felt completely into oblivion.

REFERENCES

BAILEY, J. J., et al., A method for evaluating computer programs for electrocardiographic interpretation, *Circulation* **50** (1974) 73-93.

CACERES, C. A., and HOCHBERG, H.M., Performance of the computer and physician in the analysis of the electrocardiogram, *Amer. Heart. J.* **79** (1970) 439-443.

DOBROW, R. J., et al., Accuracy of electrocardiographic measurements by computer, *Amer. J. Med. Electronics* **4** (1965) 121-126.

GUPTA, MOHIT DAJAL et al. (2005): U wave: An Important Noninvasive Electrocardiogra-phic Diagnostic Marker. *Indian Pacing and Electrophysiology Journal* **5** (1), 2005, p. 63-65 (http://www.ipej.org/0501/gupta.pdf).

JAENECKE, PETER: Grundzüge einer Meßtheorie. *Zeitschrift für Allgemeine Wissenschaftstheorie* XIII/2, 1982, p. 234-279.

LAWRENCE, P. J., *Computers in Cardiology*: The performance of a computer program in separating normal from non-normal electrocardiograms (Rotterdam, 1977).

MOORE, R.E., Interval Analysis (Prentice-Hall, 1966).

NEUBERT, D., TEPPNER, U. & SCHLIEF, R.: Aspects of standardization and validation of automatic ECG interpreting systems. *Acta Imeko* 1982, p. 429-441. Publishing House of the Hungarian Academy of Sciences.

RAUTAHARJU, P. M., Blackburn, H. W., and Warren, J. W., The concepts of electro-cardiographic, vectorcardiographic and polarcardiographic criteria, *J. Electrocardiol.* **9** (1976) 275-281.

RAUTAHARJU, P. M. and SMETS, Ph., Evaluation of Computer-ECG programs. The Strange case of the golden standard, *Comput. Biomed. Res.* **12** (1979) 39-46.

SUNAGA, T., Theory of an interval algebra and its application to numerical analysis, *RAAG Memoirs* **2** (1958) 29-46.

TEPPNER, U., LOBODZINSKY, S., NEUBERT, D., LAKS, M.M. (1987): A technique to evaluate the performance of computerized ECG analysis systems. *J. Electrocardiology* **20**, Suppl. p. 68-72.

WILLEMS, J. L., and PARDAENS, J., *Computers in Cardiology*: Differences in measurement results obtained by four different ECG computer programs (Rotterdam, 1977).

WILLEMS, J. L. et al. (1983): Common standards for quantitative electrocardiography: 3rd Progress Report. Leuven 1983.

WILLEMS, J. L. et al. (1984): Common standards for quantitative electrocardiography: 4th Progress Report. Leuven 1984.

WILLEMS, J. L., ARNAUD, P., VAN BEMMEL, J. H., DEGANI, R., MACFARLANE, P. W. & ZYWIETZ, CHR.: Common standards for quantitative electrocardiography: goals and main results. CSE Working Party. *Methods Inf. Med.* **29** (4), p. 263-271.