Exploring the Knowledge of Human Expert Beyond His Willing Expression

Piotr Augustyniak[1]

AGH University of Science and Technology, 30 Mickiewicza Ave. 30-059 Krakow Poland august@agh.edu.pl

Summary. The paper discusses the alternative method of medical experts participation in technical inventions for medicine. Blind tests and various statistic-based correlations of human and automatic interpretation results are commonly used today. Our paper postulates a deeper insight into the expert performance in order to better understanding and simulating his reasoning in the software. The benefit is twofold: the measurement is objective and the closer simulation of human reasoning yields better performance in case of unexpected input. Although the area of application is the very broad intersection of medicine and technology, we focus on the automatic ECG interpretation, and propose the agile software featuring a human-like behavior. Two examples of experiments aimed at extraction of some aspects of ECG interpretation knowledge are also included in the presentation.

1 Introduction

1.1 Rising a need for new knowledge

Information technology not only supports the medical practice of today, but also makes new challenges and opportunities stimulating the progress in medicine. The Holter ECG recorders or Computed Tomography may be recalled as first-hand examples of such inventions. In AGH-UST Biocybernetic Lab. we designed and prototyped a wearable ECG recorder-interpreter designed for a wireless cardiology-based surveillance network. Unlike the currently marketed systems, it continuously adapts the ECG signal interpretation process to several prioritized criteria of medical and technical nature. The process is designed as distributed and is performed partially by separated thread on the supervising server (network node) and partially by the agile software of the remote recorder [1]. Important novelty is also the use of digital wireless link in a bi-directional mode for patient and device status reporting but also for management and control of the remote software, requests for adaptation of report contents and data priority and reloading of software libraries as necessary. Such adaptive system yields unprecedented personalization and diagnosis-oriented processing and thus better simulates the seamless presence of a cardiologist. Until today the prototype brings rather scientific challenges, revealing new unexploited areas present in clinical practice but not covered by the standards, recommendations or guidelines [12].

1.2 Current validation rules and their limitations

The standards of ECG interpretation quality assessment [6] require the values of the diagnostic results to fall within a specified tolerance range around the value believed to be true. Such "true" reference is usually estimate by averaging the response of independent human or software experts. This approach has two drawbacks:

- it is based on the similarity of results and not of reasoning making even good automatic interpreters useless in case of unexpected input,
- it is optimal for a hypothetic "average" patient a not for a particular person, because of not considering the intra-patient variations.

1.3 Objective measurement of behavior

All measurement techniques in technical, economic and social sciences assume the less-possible extent of influence of data acquisition process on the observed phenomena [8]. Technical measurements express this idea as non-energetic information transmission. For this reason, the investigation of the medical expert knowledge using his willing expression may not be accepted as the objective measure. Although assumed not to be biased intentionally, it is influenced by several mental factors. Two principal are: memorization and verbalization.

Memorization uses the short-term memory and a part of human attention to capture his own behavior. The behavior is thus not spontaneous as it were naturally and the auto-observation usually implies subconscious auto-restriction. In result the memorized facts are altered and not complete.

Verbalization is necessary to express the memorized knowledge with a limited set of tokens belonging to a specified vocabulary. Such dictionary depends on the language used, but is also influenced by subjective preferences of the speaker. Therefore the output of an interview with an expert concerning his own reasoning may not be considered more seriously as discrete, incomplete and inaccurate impressions.

Fortunately, the interview as a research methodology has many alternatives, among of them the experiment. From this point of view, however, the originality of our approach is that medical experts are proposed to be subjects in our experiment. The presentation of this innovative idea will be developed throughout this paper.

2 Methodology

2.1 Human expert as experiment subject

Lets take the analogy form the medical diagnostic process. It usually starts with the interview, but commonly needs supplementary tests providing objective measurements of various diagnostic parameters. The patient, assuming he is a highly cooperative proprietor of the information, is not able to estimate several important facts about himself without specialized sensors and measurement methodology. In the scheme presented above, we postulate to replace the patient by the medical expert at work and to use specialized interdisciplinary technologies aimed at extracting the milestones and foundations of decision-making path. We are conscious, that similarly to the health information, this is a potentially very sensitive area and employ all ethical guidelines to the management of the experimental results. As a principle, we don't make any judgment about the correctness of the results and we keep all demographical data of the experts involved in a separate database.

In order to fulfill the requisite of an objective measurement to the maximum, we should not inform experts about their participation in the experiments. That may be feasible with the video surveillance of people on the street, but not in case of medical task performance. The experts willingly accept the participation in the experiment, as they were informed that they will be observed at work in several manners. Some of them were clearly visible, some others remain undisclosed. The applied measurement techniques should be selected with regard of the efficiency, but also the experimental setup should reproduce as close as possible the natural working environment of the expert.

2.2 Knowledge exploration techniques

The most common technique for the exploration of human behavior involves image acquisition and sequences analysis. As the interpretation of ECG signals has no kinetic background, we do not apply the video sequence analysis, however some sessions were video-recorded and the records were found helpful for validation of the session flow, identifying obstacles and interfering events. For the reason of commodity, we also do not apply complicated techniques of brain monitoring. The fMRI [4], although revealing the physiologically evoked parts of the brain is not able to reproduce the reasoning, and does not allow to simulate the usual working environment. The EEG may probably be used in human in course of the visual ECG interpretation process, but, here again long preparation, relatively poor repeatability and no representation of the cognitive process motivated us to focus on more appropriate methods.

The principal method was thus based on the fact, that the ECG interpretation process may be considered as fully visual [2]. By fixing the trace, we may consider the eye gaze trajectory as representative to the reasoning process. In fact, the observation of the trace (or any object) by the human consists of two mental processes running in parallel and interchanging information in restricted time windows [7]. Due to the very limited visual field, after the image is acquired in the human retina, the interpretation starts and as soon as acquisition completes, the research for a new focus point begins in order to provide the interpretation with a complementary image. The sight is a principal sense in human and thus eyetracking techniques are commonly used as objective indicators of visual cues (advertising) or reading and linguistic skills (education).

The second knowledge exploration technique is based on the expert-computer interaction. Assuming the computer is already used to visualize the trace for interpretation and the expert interacts with the system with use of a standard mouse, it doesn't require any additional device. After the interpretation is completed, the expert had to express his preferences in a field of selection marks. Unlike in the standard software, these fields order and the initial selection state were controlled by a random generator what prompts the expert to search for most important items at first and to revert the selection as necessary.

3 Pursuit of the human eye in course of the visual ECG interpretation

3.1 Perceptual model concept

Perceptual models have been recently recognized as valuable tool enriching the visual interaction of human with sophisticated devices [5], [9]. As a perceptual model of a biosignal record we understand a result of statistical processing of scanpaths, analyzed as polygonal curves in context of displayed visual information. The gaze order and fixation time correspond to the seeking sequence and to the amount of data gathered visually by the observer and thus they represent the diagnostic importance of particular regions in the scene [3], [10]. In the ECG, subsequent events in the cardiac cycle are represented by P, QRS and T waves positions, therefore the wave start- and endpoints were selected as reference time points for the analysis of human foveation sequence aiming at estimating the local density of medical data. Assuming the observer is properly engaged in the trace inspection, the gaze is controlled instinctively and the eyeglobe movements objectively represent the information gathering sequence. The analysis of experts' eyeglobe trajectories captured during the manual interpretation not only reveals regions of particular importance in the signal trace, but also represents the human reasoning involved in the interpretation process. Apart from main interest of our research, the prospective area of applications for eyetrack features captured during the visual inspection of biosignals include:

- objective assessment of cardiologist interpretation skills,

- teaching of the visual interpretation using the guided repetition of expert's scanpath,

3.2 Eye tracking method

The infrared reflection-based eyetracker OBER-2 capturing two-dimensional trace of each eye at 750 Hz during the ECG presentation lasting for 8s was used in visual experiments. The device provides the angular resolution of 0.02 deg and uses time-differential method for the sidelight discrimination. This angle corresponds to a time interval of 30 ms on a standard ECG chart plot (25mm/s) viewed from a typical reading distance (40 cm). The position of both eyes was recorded simultaneously and a custom-developed software detects the dominant eye which trace is used to determine the electrocardiogram conspicuity. Figure 1 displays the physical principle of the differential infrared reflection-based eyetrack acquisition.

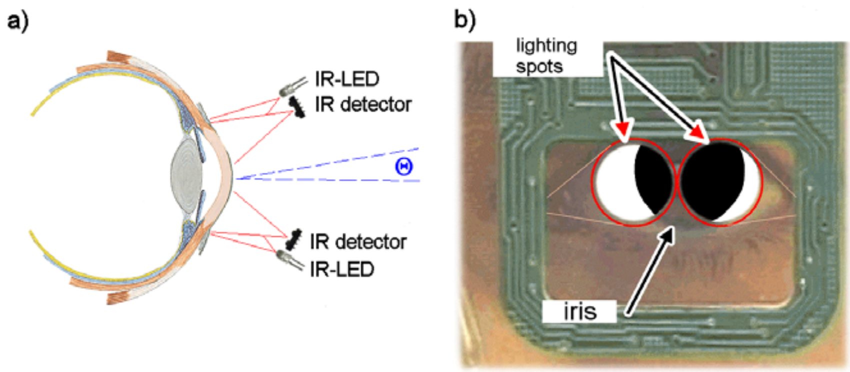

Fig. 1. a) Physical principle and b) technical details of the infrared reflection-based eyetracker OBER-2

3.3 Experiment setup and participants

The total of 17 experts (12 ± 4 years of experience) volunteers accepted the invitation to the laboratory for the visual experiment. All observers were asked to complete the statistical questionnaire on their ECG experience and possible eyesight defects before attempting to the visual task. The ECG traces were randomly selected for interpretation from CSE recordings [13] and were presented as bitmaps on a 17 inch CRT monitor. The display simulated a conventional 12-leads paper recording (fig. 2). The reading distance was controlled with use of a chin support set 40 cm apart from the display center. Each

ECG trace presentation was interlaced with the fixation point in the center of the display. The reference wave borders, although not displayed, provided the cardio-physiological context for the scanpaths analysis. The horizontal axe of the scanpath is projected on the temporal progress of heart cycle, represented by positions of wave borders. Piecewise integration of scanpath time allows to estimate for each cardiac component the amount of information it contributes to the final diagnosis.

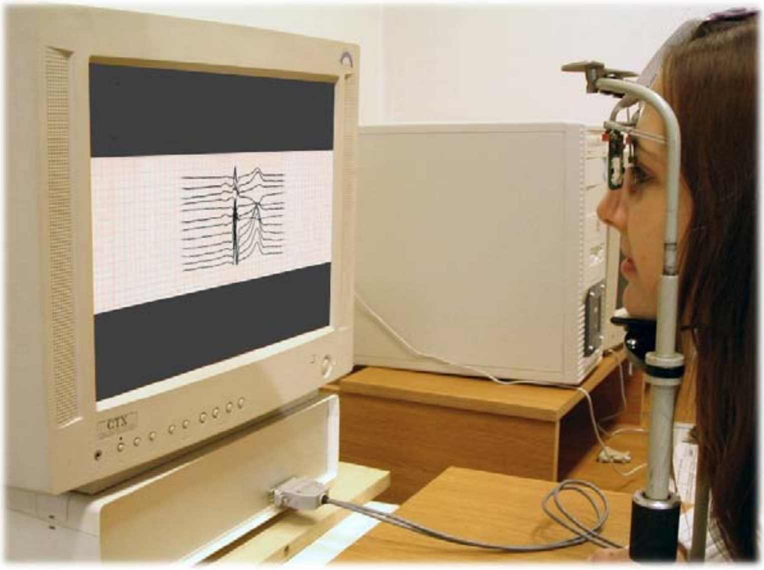

Fig. 2. The expert volunteer performing a visual task on ECG interpretation

3.4 Scanpath signal processing

Each visual experiment provides a four-column matrix representing raw eye-globe coordinates at the evenly spaced time points [11]. Prior to the ECG traces investigation, the calibration rectangle is displayed and the observer is asked to gaze at its corners. The gaze points corresponding to the corners are identified in the eyetrack and help in calculating the display-relative coordinates from the A/D converter output. The further signal processing routines were developed in Matlab with regard to the aims of visual experiments. Main stages of this calculation are performed on the pre-detected dominant eye trace and include:

- the initial idle time t_i and the interpretation task completion time t_e were detected in the scanpath,

$$\underset{t<t_i}{\forall} \sqrt{\Delta x_t^2 + \Delta y_t^2} < \varepsilon \tag{1}$$

$$\underset{t>t_e}{\forall} y_t > m \tag{2}$$

where ϵ is the noise level and m is the maximum vertical screen coordinate

- using a set of reference wave borders $S^i_{\min}$, $S^i_{\max}$ provided in the CSE database, each foveation point P in the scanpath was qualified as corresponding with the particular ECG sections i,

$$P \subset S^i : S^i_{\min} \leq p_x < S^i_{\max} \tag{3}$$

- the number and duration D^i of foveation points was integrated separately for each ECG section i in all ECG displays,

$$D^i = \sum\nolimits_{t=t_i}^{t_e} P(t) \subset S^i \tag{4}$$

- the contribution of each section's conspicuity was referred to the total observation time.

$$C^i = \frac{D^i}{t_e - t_i} \tag{5}$$

The intrinsic variability of waves' length does not influence the result, since the foveation points are referred to ECG fiducial points and not directly to the ECG time. Apart from the waves conspicuity statistics, the processing reveals the perceptual strategy related to main stages of the ECG interpretation process. The principle of strategy description is the identification of:

- most attractive points coordinates,
- their gaze order in context of the ECG time and displayed ECG leads.

These parameters were chosen as most representative to the global density of diagnostic information distribution in the heart cycle and to the information priority required by a diagnostic decision scheme followed intuitively during the manual ECG interpretation by the human expert.

3.5 Results of the human eye pursuit

The statistical parameters of all visual experiment results are summarized in table 1. Figure 3 displays an example of eyeglobe trajectory over a 12-lead ECG plot together with the corresponding bar graph of attention density.

The results in table 1 prove the common belief about irregular distribution of medical data in the electrocardiogram. However, main novelty here is the quantitative assessment of expert's attention density and its variations in the heart cycle. As much as 38 percent of information in the signal is represented in the QRS complex attracting the experts' gaze to this relatively short (105 ± 23 ms) section. Despite the considerable length of the baseline (278 ± 115 ms), only 14 percent of gaze points fall to this section confirming its little diagnostic

Table 1. Results of ECG inspection scanpaths analysis

Parameter	Unit	Observers Experts
idle time	ms	73 ± 55
interpretation time	s	5.5 ± 1.5
P wave foveation	%	23 ± 12
PQ section foveation	%	7 ± 5
QRS wave foveation	%	38 ± 15
T wave foveation	%	18 ± 10
TP section foveation	%	14 ± 5
max. attention density	s/s	21.0
min. attention density	s/s	1.9

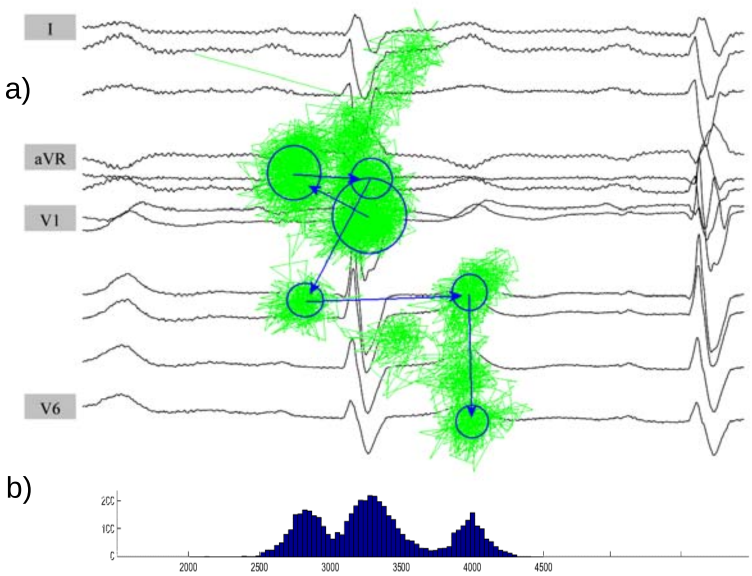

Fig. 3. a) The example of expert's eyeglobe trajectory over a 12-lead ECG plot (CSE-Mo-001); the circle diameter represents foveation time, b) corresponding bar graph of the attention density

significance. Comparing minimum and maximum attention density values we found interesting that this value varies over 10 times, what represents high expert concentration on most informative part of the image. The second group of result were derived by the analysis of perceptual strategy. Figure 3a displays an example of the strategy over a 12-lead ECG plot and table 2 summarizes the corresponding strategy description parameters. For studies on perceptual strategy repeatability we selected electrocardiogram images investigated by at least two observers. By comparing the positions and gaze order of five most important foveation points in the scanpaths we found several different strategies applied by the experts. Having no means to assess them, we only rank them by frequency and observe, that the similarity between two experts may be expected with the probability of 37%. This result prove the proper representation of ECG interpretation process in the visual strategy.

Table 2. Quantitative description of the most frequent perceptual strategy

Parameter	Unit	Observers Experts
relative foveation time for the main focus point	%	31 ± 12
number of foveation points		6.1 ± 1.7
foveation points distance	deg.	5.7 ± 2.4
scanpath length to the last foveation point	deg.	34.7 ± 5.1
scanpath duration to the last foveation point	s	3.6 ± 1.3

The scanpath, however, is very sensitive to the voluntary observer cooperation during visual tasks. Poor cooperation or misunderstanding of visual task rules was the main reason for exclusion of 18% of records from the s-canpaths statistics. The scanpath statistics and perceptual strategies revealed many differences between cardiology experts concerning the ECG inspection methods. However, all the statistical parameters indicate a very precise and consistent way of information search by experts. Moreover, high variation of first foveation points focus time and distance suggest the hierarchical information gathering reflecting the parallel decisive process.

4 Pursuit of the human choice in course of diagnostic result selection

4.1 Expert's choice as indicator of medical data relevance

Following the requisites of objective measurement, the pursuit of human choice for diagnostic results priority was made with use of a hidden poll. In order to avoid all-inclusive selections, an artificial restriction of resources was applied. It is based on the expected data stream value attributed to each diagnostic parameter. The total available data volume was set as equal to a half of the data volume of all parameters. In such environment, the doctor has to allocate the space first for the most relevant data, and simultaneously exclude the data he or she considers useless. The aim of this investigation was to record and analyze the expert's behavior in order to extract the knowledge about the relative relevance of ECG diagnostic parameters in most frequent diseases. Such hierarchy yields promising advantages in systems with patient-specific adaptation of the interpretation processing.

4.2 Usual interface with hidden poll functionality

The manufacturer of ECG interpretation software performs a standard technical validation procedure which is followed by clinical usability verification in selected cardiology expert offices. This last step is very important as it is done by medics, able to demonstrate the software usefullness in live conditions. A trivial modification of the commercial ECG interpretation software (Cardioteka ©, Aspel) was a background for experimental studies concerning doctors' choice about the report contents. The default proposal of a final report contents was replaced by a random pre-selection (fig. 4).

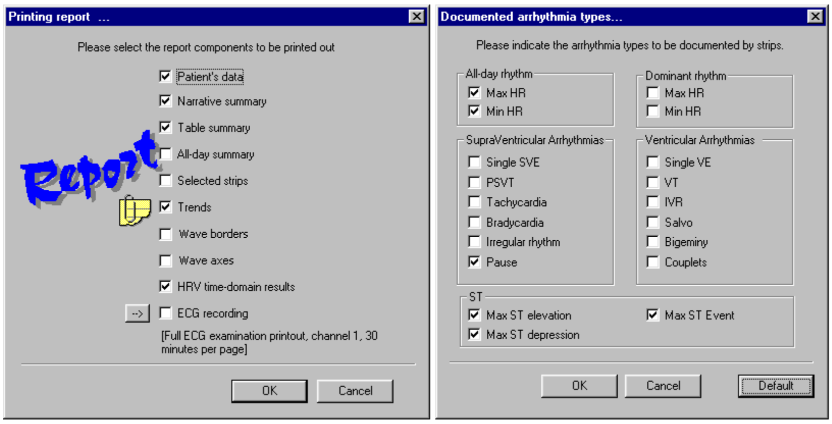

Fig. 4. Example selection screen for the choice on the report contents. Subsequent displays differ by items order and initial selection state

Once the interpretation is completed, all available report items appear together on the screen and the doctor had to select (deselect) results he or she wish to include in (exclude from) the report contents. The order of selections made and chosen items are memorized with the diagnostic outcome. The survey included 1730 ECG analysis cases and allowed to pursuit the cardiologists' preferences in 12 most frequently observed diseases (normal sinus rhythm, sinus tachycardia, sinus bradycardia, probable AV block, ventricular escape beats, atrial fibrillation, AV conduction defect, myocardial infarction, atrial enlargements, ventricular hypertrophies, left bundle branch block, right bundle branch block). The observations count for these pathologies ranged from 16 to 323 cases. For other 17 pathologies, the occurrence frequency was below 16 in the available population and thus no statisticaly-justified conclusions may be drawn from.

4.3 Statistical processing

Finally, the statistical processing of the gathered data was aimed at revealing the knowledge about doctors' preferences. Main steps of the calculations were the following:

- Inclusion or exclusion of a parameter to/from a diagnostic report increases or decreases its relevance accordingly to the expert action. First inclusion is the most relevant, first exclusion is the least relevant etc. Items remaining untoutched by the expert are not considered for the hierarchy statistics.
- For each considered disease, the disease-specific hierarchy list was build of diagnostic parameters p, ordered by their frequency F of occurrence at a given position L_C relative to all occurrences at other positions L weighted by their distances $|L - L_C|$:

$$F = \frac{\sum C : L_C = L}{\sum (C \cdot |L - L_C|) : L_C \neq L} \tag{6}$$

- The diagnostic relevance is represented by the weighting coefficient W_p including the rank L and the frequency F:

$$W_p = \frac{F}{L} \tag{7}$$

- Finally, the weighting coefficients were normalized so as they sum to the unity

$$\sum_p W_p = 1 \tag{8}$$

This operation yields a disease-specific vector of weighting coefficients representing the medical relevance of ECG diagnostic parameters (tab. 3)

Table 3. Hierarchy matrix of diagnostic parameters in patient state (excerpt); The normalized weighting parameters W_p are displayed

diagnostic parameter	patient status space			
	normal sinus rhythm (111)	persistent supraventricular tachycardia (163)	ST suppression heart muscle ishemiae (508)	.....
heart rate (HR) [1/min]	0,15	0,35	0,22	.....
dominant trigger rate [%]	0,25	0,21	0,07	.....
PQ section [ms]	0,12	0,17	0,03	.....
.....	.....	.....	.....	.....

4.4 Results of the pursuit of human-made relevance indication

These studies confirm the common, but poorly justified belief, that for the human expert some diagnostic results are more important than others. The relevance of particular medical parameters highly depends on the known status of the patient. Moreover, assuming that several common diseases may be reliably diagnosed in a fully automated process, our studies result in attributing each disease with a hierarchical list of most adequate diagnostic parameters. That list is very useful as a background of disease-dependent report modification in a distributed cardiac surveillance system. It may also be applicable to a medically-justified estimation of cardiac data quality.

5 Conclusions

The paper presents the idea of behavioral observation and measurements in human, applied to the extraction of the cardiology expert knowledge being a background of the visual signal interpretation and management of parameters. The methodology of undisclosed observation is not new, it is commonly accepted in sociology and medical sciences. It fulfills very well the requisite of objective or unbiased measurement.

Although very advantageous, the method involves ethical issues, and probably such methods should be used under a supervision of ethical commissions similar to other experiments *in vivo*. The human under test being a proprietor of the knowledge and of the performance, has only a limited influence on the information he or she provides to the analytic system. On the other hand, medicine is at the leading edge in the usage of similar experiments in animals and in human and therefore high level of understanding from doctors participants should hopefully be expected.

6 Acknowledgment

Scientific work supported by the AGH-University of Science and Technology grant No 10.10.120.783.

References

1. Augustyniak P. (2006) The use of selected diagnostic parameters as a feedback modifying the ECG interpretation. Proc. Computers in Cardiology; 33, 825-8,
2. Augustyniak P., Tadeusiewicz R. (2006) Assessment of electrocardiogram visual interpretation strategy based on scanpath analysis, Physiol. Meas. 27, 597-608
3. Boccignone G (2001) An Information-theoretic Approach to Active Vision Proc. 11th International Conference on Image Analysis and Processing, 2001

4. Buxton RB. (2002) An Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques, Cambridge Univ Press
5. Dick AO. (1980) Instrument scanning and controlling: using eye movement data to understand pilot behavior and strategies" NASA CR 3306
6. IEC (2001) 60601-2-47. Medical electrical equipment: Particular requirements for the safety, including essential performance, of ambulatory electrocardiographic systems.
7. Kasprowski P., Ober J. (2004) Eye Movements in Biometrics. ECCV Workshop BioAW, 248-258
8. Kimothi, SK. (2002) The Uncertainty of Measurements: Physical and Chemical Metrology: Impact and Analysis (online)
9. Ober JK., Ober JJ., Malawski M., Skibniewski W., Przedpelska-Ober E., Hryniewiecki J., (2002) Monitoring Pilots Eye Movements during the Combat Flight-The White Box, Biocybernetics and Biomedical Engineering, vol. 22, (2-3), 241-264
10. Pelz JB., Canosa R. (2001) Oculomotor behavior and perceptual strategies in complex tasks, Vision Research, vol. 41, 3587-3596,
11. Salvucci DD., Anderson JR. (2000) Automated Eye-Movement Protocol Analysis" Human-Computer Interaction vol. 28,
12. (1997) The standard action in Holter examination. Polish Society of Noninvasive Electrocardiography.
13. Willems JL. (1990) Common Standards for Quantitative Electrocardiography 10-th CSE Progress Report, Leuven: ACCO publ.,