Pursuit for the Knowledge of a Cardiology Expert – a Hidden Poll Methodology

Piotr Augustyniak, *Member, IEEE*

***Abstract*— The human-computer interaction became a routine and intuitive task of a cardiologist. The ECG interpretation software may be modified to trace the interaction history and to provide the information useful for the investigation of the expert preferences. Because the consciousness of the investigated expert is not engaged as much as in the disclosed questionnaire, the response is not biased making the measurement more objective. The paper demonstrates metrological backgrounds and the use of a hidden poll in the investigation of the priority of ECG diagnostic parameters in context of various cardiac diseases. This knowledge, difficult to be derived otherwise, is applied in the disease-oriented assessment of the diagnostic quality.**

I. Introduction

All measurement techniques in technical, economic and social sciences assume the less-possible influence of data acquisition process on the observed phenomena [1]. Technical measurements express this idea as non-energetic information transmission. For this reason, the investigation of the medical expert knowledge based on the use of his willing expression only may not be accepted as the objective measure. Although assumed to not 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 [2], [3]. The behavior is thus not spontaneous as it is naturally and the auto-observation usually implies subconscious auto-restriction. In result the memorized facts are not complete and altered.

Verbalization is necessary to express the memorized knowledge with a limited set of tokens belonging to a specified dictionary [4]. Such dictionary depends on a language used, but is also influenced by subjective preferences of the speaker. Therefore the output of the 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 *in vivo*. From this point, however, the originality of our approach is that the medical experts are proposed to be subjects in our experiment. This innovative idea will be developed throughout this paper.

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 [5]. 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. Because the auto-adaptivity doesn't guarantee the improvement of diagnostic quality without a medical background, we have to investigate the human performance of visual signal inspection and data management in order to extract the rules and correctly implement them into automatic procedure for interpretive software management in remote recorder.

Fig. 1. Status-specific prioritized lists of diagnostic parameters in a wearable ECG surveillance system.

II. Materials and Methods

A. Comparing human and software result

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 [7]. This approach has two drawbacks:

Manuscript received April 2, 2008. Scientific work supported by the AGH-University of Science and Technology grant No 10.10.120.783

Author is the head of Multidisciplinary School of engineering In Biomedicine (MSIB), AGH University of Science and Technology; 30, Mickiewicza Ave., 30-059 Krakow Poland, august@agh.edu.pl

- 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 and not for a particular person, because of not considering the intra-patient variations.

B. Human expert as the subject of experiment

Human expert - taking the cardiology only as an example - 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 him 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 [8].

In order to approach the conditions of an objective measurement, we should restrict the information delivered to participants about experiments backgrounds. The experts were informed that they will be asked for the ECG interpretation and they will be observed at work in some manner. During and after the experiment, the measurement technique remains undisclosed. The experimental setup reproduced as close as possible the natural working environment of the expert.

The knowledge exploration technique is based on the expert-computer interaction. It doesn't require any additional equipment apart from the computer already used to visualize the trace for interpretation. 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.

C. Usual interface with hidden poll functionality

The experiment was performed with an agreement of the software manufacturer verifying a new prototype (Cardioteka©, Aspel) in selected cardiology expert offices. Thanks to modification of tested software we applied a hidden poll functionality in the standard interface of report content selection. It consists in replacement of default proposal of a final report contents by a random order and pre-selection of items (fig. 2). 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.

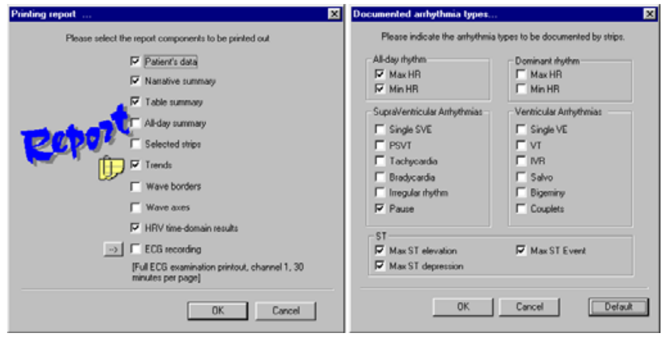

Fig.2. Example selection screen for the choice on the report contents. Subsequent displays differ by items order and initial selection state

Thanks to prompting the expert to actively select or deselect report components we collected data for experimental studies concerning doctors' preferences. An artificial restriction of resources was applied in order to avoid all-inclusive selections. 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.

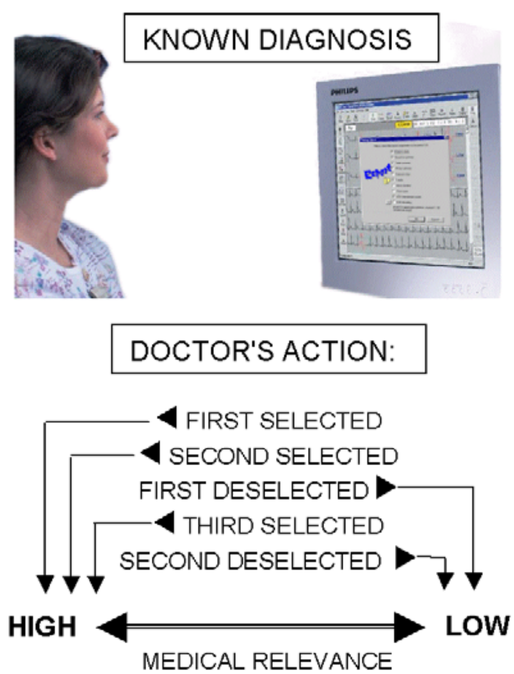

Fig.3. The principle of quantitative doctor's choice measurement leading to disease-specific hierarchy of diagnostic parameters

D. Experiment setup

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 statistically justified conclusions may be drawn from.

E. Processing of the poll results

The statistical processing of the gathered data was aimed at revealing the knowledge about doctor 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 untouched 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 $|L - L_C|$:

$$F = \frac{\sum C : L_C = L}{\sum (C \cdot |L - L_C|) : L_C \neq L} \quad (1)$$

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

$$W_p = \frac{F}{L} \quad (2)$$

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

$$\sum_p W_p = 1 \quad (3)$$

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

III. Results

Table 1. presents the excerpt from the matrix of result attributing to each of the 12 pathologies a hierarchical list of most relevant diagnostic parameters. The table presents normalized values of the weighting coefficients, although due to the lack of completeness the values in rows doesn't sum to the unity.

TABLE I
DISEASE-SPECIFIC HIERARCHY OF ECG DIAGNOSTIC PARAMETERS
(FULL MATRIX CONTAINS 22 COLUMNS WITH PARAMETERS)

disease	heart rate	dominant trigger	PQ duration	QRS axis
normal sinus rhythm	0.15	0.25	0.12	0.15
sinus tachycardia	0.55	0.25	0.10	0.13
sinus bradycardia	0.57	0.23	0.10	0.13
probable AV block	0.23	0.17	0.27	0.18
ventricular escape beats	0.27	0.61	0.07	0.05
atrial fibrillation	0.35	0.08	n. a.	0.06
AV conduction defect	0.19	0.13	0.39	0.21
myocardial infarction	0.15	0.12	0.14	0.28
atrial enlargements	0.12	0.02	0.15	0.02
ventricular hypertrophies	0.03	0.07	0.05	0.31
left bundle branch block	0.11	0.18	0.21	0.34
right bundle branch block	0.08	0.17	0.21	0.28

Significant variations of medical relevance may be observed in the resulting table. In tachycardia, bradycardia and fibrillation the heart rate seems to be the most important parameter. In case of ventricular hypertrophies its importance drops dramatically giving priority to other parameters, particularly to QRS axis positions. The dominant trigger source shows its primary importance in detection of ventricular escape beats, and also medium importance for distinguishing sinus rhythms from other rhythm types. The PQ duration or duration stability, representing the stimulus propagation during an atrioventricular depolarization sequence is a factor of primary relevance in case of AV conduction defects detection, and may be considered as auxiliary determinant in atrioventricular blocks or blocks of His bundle branches. The electrical axis of QRS complex represents the ventricular wave front propagation and thus it is much more important in detection of myocardium-related diseases (hypertrophy, infarct) or His bundle problems, than in detection of stimulus triggering or conduction abnormalities.

Table 1. presents four parameters showing primary relevance throughout the tested examination set. Other 18 parameters has a statistically minor contribution to the report contents, however, in very specific case, they take the priority over those specified above.

IV. Discussion

Our studies confirm the common, but poorly justified belief, that for the human expert some diagnostic results are more important than others. In course of the experiment, the parameters hierarchy was established, but also the relevance factor has been measured and quantitatively expressed. 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.

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 management of diagnostic parameters by the experienced human. 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.

V. Applications

The formulation of disease-specific hierarchical lists of ECG diagnostic parameters is new knowledge derived experimentally in a result of the objective measurement. It may be useful in many aspect of medical informatics, human behavior assessment in cardiology and beyond.

This particular experiment yielded new sort of information necessary for the correct setup of an auto-adaptive wearable ECG recorder. The primary requisite of such systems is software adaptation aimed at avoiding unnecessary data processing and transmission. The auto-adaptation consists in continuous optimization of the distributed processing chain justified by the adequacy and quality of the patient's description. The medical reports collected by the central server from mobile patient recorders-interpreters are irregular in frequency and contents, and therefore the report is continuously checked for compliance with patient status-specific expectations.

Another interesting application is the medically justified assessment of signal and data quality in compression and retrieval systems. As far, most of the distortion estimators neglect the medical content of the signal assuming every sample had the same importance. This assumption is convenient but not true, however the quantitative distribution of local signal importance may hardly be found in the literature. Assuming the diagnostic parameters may be retrieved from the signal in a fully automated process with a reasonable reliability (which is much closer to the reality than the previous assumption), two ECG signals may be justified as similar or not on a background of the similarity of the diagnostic outcome. Here the patient status-dependence plays the principal role, since it defines weighting coefficients for the quality estimators of respective diagnostic parameters for their contribution to a global estimate of similarity.

As described in the experimental part of our research, the hidden poll methodology and behavioral studies of qualifying parameters as relevant or not relevant for a particular medical cases may also be a background of human performance assessment. This was not intended as a target application, and needs further support in aspect of medical correctness and compliance with professional guidelines. Particular action of the cardiologist under test in response to a purposely prepared set of examination traces may be assessed automatically in an objective way and yield a quantitative estimate of the human expertise.

References

[1] S.K. Kimothi, "The Uncertainty of Measurements: Physical and Chemical Metrology: Impact and Analysis", ASQ Quality Press, Milwaukee, Wisconsin 2002.

[2] C. Izawa, "Human learning and memory: Advances in theory and application". Mahwah, NJ: Lawrence Erlbaum Associates. 2005

[3] J. H. Mace, (Ed.) "Involuntary Memory". Malden, MA: Blackwell Publishing, 2007

[4] TJ Lloyd-Jones, C. Brown, S. Clarke "Verbal overshadowing of perceptual discrimination." Psychon Bull Rev. 2006 Apr;13(2):269-74.

[5] P. Augustyniak " The use of selected diagnostic parameters as a feedback modifying the ECG interpretation. Proc. Computers in Cardiology; 33, 2006, pp. 825-8,

[6] IEC 60601-2-51. "Medical electrical equipment: Particular requirements for the safety, including essential performance, of ambulatory electrocardiographic systems." 2003,

[7] The CSE Working Party "Recomendations for measurement standards in quantitative electrocardiography". Eur. Heart J. 6, 1985, 815-825.

[8] A. Capron, " Human Experimentation: Basic Issues."[In:] "The Encyclopedia of Bioethics" vol. II. The Free Press, Glencoe, IL. 1978,