Bioengineering the Mind: from Artificial Intelligence towards Artificial Consciousness

Radu NEGOESCU

National Institute of Public Health, Bucharest, Romania, rnegoescu@ispb.ro

Abstract – Whilst the artificial intelligence seems recently to approach its human-close specimen, artificial consciousness as targeted by bioengineering and information science&technology advances still has some way to go before becoming an experimental terrain for a bunch of sciences that deals with the problem of conscience, including philosophy and theology. Depending on our capacity to inseminate a machine transposition of natural ethics at the same time with increasing machine autonomy, a well guided artificial consciousness holds the promise to offer a representation of what natural consciousness could be in absence of distorting influences exerted by biologic (genetic) inheritance on human being as it presents nowadays.

Index Terms – consciousness, conscience, machine consciousness, information technology, cardiovascular bioengineering.

I. ARTIFICIAL INTELLIGENCE

Defined by John McCarthy in 1956 as the science and engineering of making intelligent machines, hopefully in the human intelligence sense, the artificial intelligence (AI) certainly evolved over the past half century, even if we never got the humanlike assistants that many thought we would have by now. It offers a valuable technological support in critical domains, e.g. computer-diagnosing patients over the internet, but even the most helpful AI system in function today must be programmed explicitly to carry out its one specific task. What people wanted and needed was a generalpurpose intelligence that can be set loose on any problem, i.e. one that can adapt to a new environment without having to be retrained constantly: "one that can tease the single significant morsel out of a gluttonous banquet of information the way we humans have evolved to do over millions of years" [1].

Recently yet Hewlett Packard introduced a new class of electronic device overriding the separation between memory and processing, the memristor, into a "brain-inspired" microprocessor featuring the form factor of a brain, the low power requirements, and the instantaneous internal communications - that could be trained and coaxed to behave like a brain. Run on this "brain on a chip", the MoNETA (Modular Neural Exploring Traveling Agent) software written at Boston University's will perceive its surroundings, decide which information is useful, integrate that information into the emerging structure of its reality, and in some applications formulate plans that will ensure its survival - the same drives that motivate humans and entitle the machine as a specimen of true (or real) artificial intelligence [1].

II. CONSCIOUSNESS AND CONSCIENCE

English uses the term *consciousness* (or self-awareness) to designate a neural-behavioral state featuring capabilities of reflection and reaction found as adequate by the rest of the world, while being vigil.

French appears to make no much lexical distinction between consciousness and conscience that are commonly referred to by 'conscience' and lets to the context making the difference. However, 'faits de conscience' and "connaissance' refer unambiguously to consciousness or to a part of it.

After M. Draganescu [2] consciousness stands for a type of integrative information (structural-phenomenological and social) capable of understanding and knowing, knowing that knows, and endowed with: feeling of to be, will, intuition and creative power. Notice that in the philosophical thinking of M. Draganescu structural information is related to (non-living) nature and its sciences, while phenomenological information is related to the living matter studied by life sciences. Integration of structural and phenomenological takes place into the real human being as such; dissection by theoretical reasons may enlighten various balances between parts otherwise intimately merged when analysis progresses from the molecular and cellular level to organs, systems, mind and soul.

Consciousness is naturally human; its versions "contaminated" by technology or those purely technological are referred to by artificial ones.

Coming back to terms, in Romanian conscious ('constient') is also (English-like) pointing to someone who can rationally place his/her Ego vis-à-vis of the world and him/herself; the term "rational" sends to the manner in which that positioning is done by a majority of other individuals. Besides, the Romanian 'constient' refers to someone endowed with a certain level of *conscience*: "I am 'constient' (aware) of my duties", where the attitude versus duties is already related to moral principles, to an axiology.

Conscience is yet more than what is involved by "I am aware of" (that expresses a potential), namely a non-hesitant (proved) availability to actualize this potential with the current behavior. At a higher individual level, conscience involves looking into the meaning of existence, for him and for others who do not possess necessary capabilities, by a philosophical and/or religious demarche.

However exercising conscience is mainly done in the social environment. M. Draganescu's [2,3] social-human civilization of the future would be by far towering biological needs of an aggregate of human individuals whose interaction would filter (somehow in the sense of coherent summation in physics) luminous parts present in all of them as pieces of truth that are detected, sifted and put together through the collective, social exercise of spirituality. Or, to cut it short, by the collective conscience.

In this vein, laws of Moses giving early expression to collective conscience codify social behavior of people; social sins acting against the group are mainly incriminated, rather than individual sins acting against him/her self. While the latter are health-redoubtable (we know how from medicine of lifestyle) ending sometimes in serious somatic illnesses, the former increase or exacerbate psycho-social stress responsible for more subtle forms of disease or death (e.g. sudden cardiac death in apparently healthy people).

The individual endowed with (sufficient) conscience is unselfish, generous; generosity is seen as the essentials of Christianity (as an example of spirituality) gathered in one single word. At the other end, a social value as bright as freedom, when practiced at low levels of conscience (or without conscience of kind) converts to selfishness, greed and open contempt vis-à-vis of fellow man. Social behavior of many Romanians since 1989, overtaken at low levels of conscience, if any, by freedom achieved through sacrifice of others (Revolution heroes), may convincingly illustrate what means non-conscience.

From these preliminaries, the relationship between consciousness and conscience could schematize as:

consciousness + moral principle = conscience.

Moral principle comes for a vast majority of humans from spirituality.

"Everyone, writes Mihai Draganescu [3], has an empirical understanding of conscience and realizes that it stands for the highest level of his/her being. He/she then feels spirituality and spiritual experiences to be the very core of his/her conscience". In this view, unlike consciousness, conscience is exclusively human.

On this background, a genuine social-human civilization would also be a Society of the Conscience.

In general, society is seen by M. Draganescu [3] at the crossing of influences coming from science & technology, environment, genetics and cultural (epigenetic) heritage, and spirituality (Figure 1). Spirituality is not a relatively objective social propeller, like science, but lies in depth of the intangible human subjectivity, in the conscience.

The question arises whether man's level of conscience (dependent on spirituality) could overcome at the societal level the destructive effects of those parts of his genetic inheritance directed to evil and aggression that prevents the progress of mankind towards a genuine social-human civilization.

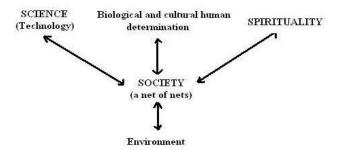


Fig. 1 (modified after M. Draganescu, [3])

As for the genetic inheritance, M. Draganescu [3]) is quoting the biologist and physician Grigore Traian Popa who reviewing in the 40's the evils in the society of his times had put that brain should be taken into discussion when investigating what is going wrong in individuals and society since the brain can instrument both good and evil.

Keeping the story short, due to its contamination by genetic inheritance, human consciousness as it presents

today could not guarantee the progress of mankind towards a Conscience Society penetrated by the moral principle of spiritualized humanism that would dominate all social networks. In his key lecture at the INGIMED II Conference in 2001, M. Draganescu argued why he is skeptical on building a Society of Conscience without participation of artificial consciousness (AC).

Summarizing, in terms of cerebral activity, consciousness stands for neural machinery, mental, reason. Moral principle, for the vast majority of individuals, is involved by spirituality, so it is transcendent. Finally, conscience is both material, as tributary to neural machinery, and spiritual, that is transcendent. For some good reason it is said about a man of conscience that "he puts his soul into".

As M. Draganescu [2] remarks, man of today might not be able to create a social-human civilization as his genetic apparatus dominates the epigenetic cultural acquisitions. Then a solution could be an artificially assisted consciousness by implantation of neurocybernetic "consciousness prostheses" that by means of significant mental enhancement would offer a better chance to moral rectitude in the average individual; we've just seen that his consciousness is largely missing nowadays the influence of spirituality.

III. WAYS TO ARTIFICIAL CONSCIOUSNESS

In terms of technological contribution to a improving or recovering human consciousness, neurocybernetic prostheses are not a novelty in bioengineering or even in clinical engineering, fairly yet with much more modest goals than "treating" low levels of conscience.

Thus, neural engineering aims at replacing a damaged part of the human brain, involved in cognitive functions, with integrated circuits operating on the known principle of the artificial neural networks. Integrated circuits would not necessarily be on silicon that after 60 years of supremacy will leave their place by 2015 to molecular and quantum devices.

Other hopes appear related to neuroelectronics that refers to coupling organic substrata to electronic systems and devices. In this vein Fromhertz et al (quoted by Draganescu, [2]) have combined a silicon chip with the giant nerve cells of the snail Lymnea Stagnalis and succeeded a two-way communication, recording and stimulating without micropipettes, simply by growing neurons on silicon surfaces - inert except some sensitive areas for collecting and emitting signals (in fact microelectrodes.

In a broader perspective, Koch and Tononi [4] evaluate this way the chances of occurrence of artificial consciousness: "Consciousness is a part of the natural world. It depends, we believe, only on mathematics and logic and on imperfectly known laws of physics, chemistry and biology. It does not derive from a magical or transcendent quality. If so, then there is no reason that consciousness could not be reproduced in a machine, at least in theory".

Notice that by the term biological (that is living) they add phenomenological dimension (in Draganescu's sense) to the structural-informational world dealing with mathematics, logic, physics and chemistry.

In this line of thinking, Koch and Tononi argue how consciousness does not seem to require many things we currently associate with human being: emotions, memory, auto-reflection, language, sensitivity to the ambient and action in the world. To be conscious, they say, is in last analysis to be a single integrated unit with a large repertoire of states. Integrated information based consciousness theory (IIT) may suggest a test to measure the degree of consciousness of a machine - a sort of Turing test for consciousness (Turing test is a method of detecting a presence of a human intelligence behind a machine presented as an automaton).

Talking about the best way to build a conscious machine, these authors evoke two complementary strategies: copying a mammalian brain or evolving a machine.

The first way seems without perspective: modeling the brain of a round worm (Caenorhabditis Elegans) with only 302 neurons and approximately 6000 chemical synapses has begun in 1986 and more than 2 decades later there was no valid model on how this minimal nervous system works.

A more plausible alternative is starting from architecture of mammalian brain conveniently abstracted and evolving it towards a conscious entity. Attempts to date, the Aibo robotic dog or the Qrio humanoid proposed by Sony are rudimentary tries to operate on a large number of fixed but flexible rules and would not pass perhaps the consciousness test proposed by the IIT. Vision systems based on hierarchical multistrata maps of "neurons" (artificial neural networks) are admirably managing to classify images from the real world, but presents obvious fragility when modifying background brightness entailed, for example, by a change of scenery.

As Koch and Tononi conclude, the big stake of reflection on how to build a conscious machine is undoubtedly more clear understanding of our own consciousness, as a necessary support for the next step to take over towards more conscience - we might add.

IV. INFORMATION TECHNOLOGY APPROACH

The assumption that computing machines could become conscious is based on the analogy seen by many between brain (wetware) and computer (hard- & soft- ware). It is expected that before long the computers will reach the estimated complexity of the brain.

A healthy adult brain contains about 100 billion neurons, each of them connected by axons (output), dendrites (input) and synapses with other about 100,000 neurons. It results that a typical brain has about 1015 connections between its neurons, each supporting at least one discharge per second. Many think that in about a decade computers will reach the computational power of the brain when exceeding 10x1015 operations/second (op/s). The IBM supercomputer Blue Gene /P could execute a year ago up to 3x1015 op/s. Argonne National Laboratory (USA, Illinois) is now upgrading a Blue Gene/P for doing circa 1/2x1015 op/s [5].

However, complexity of the brain once reached, "no one has the foggiest notion" (E. Kandel, Nobel Laureate, quoted by Horgan, [5]) how the computer could possibly make the qualitative step towards consciousness or beforehand how agglomeration of neurons and other soft tissues constituting the brain gives rise to conscious mind - that intangible entity that in Horgan's words "makes you falling in love, seizing the irony in a novel, or appreciating the elegance of an electronic design".

While accepting the possibility mentioned by others to create a conscious quantum computer M. Draganescu [2] excludes structural complexity as a source of machine consciousness; instead, he sees the structural-phenomenological complexity as a necessary condition for artificial consciousness.

Issued in connection, symbiotic or not, with the human brain, a conscious machine would hold a promise of immortality sui generis, transcending decomposition of our biological hardware. Thus, the advocates of singularity see us, half in the joke half seriously, first becoming cyborgs - carriers of implanted chips to emulate perception, memory and intelligence, and finally abandoning our flesh-and-blood selves for uploading our profound ego, digitally formatted, in a computer memory that will forever ensure our immortality in the cyber-space. For some, this prospect is tangible; for example Kurzweil, an enthusiast of singularity, contemplates changing his lifestyle in the sanogenetic sense "to live quite enough to live forever" (cf. Horgan, [5]).

Letting the joke aside, Cardon, Camus, Campagne et al embarked in 2005 upon an ambitious project meant to conceptualize and build a system generating 'faits de conscience', in fact an artificial brain aiming at transposing human thinking of something into the computable field, so that an computer-based artificial system would be able to exhibit consciousness features in a viewable manner.

"The system will have intentions, emotions and ideas about things and events related to itself. The system would have to have a body that it could direct and which would constrain the system. It would also have to have a history, and intentions to act and, most of all, to think. It would have to have knowledge, notably language knowledge. It would have to have emotions, intentions and finally a certain consciousness about itself" (Cardon, Camus, Campagne et al, [6]).

There is a summum bonum, a most generous statement of intentions in this field that should deserve, judging conceptual and IT effort deployed, careful consideration even if authors often forget to put due quotation marks when it is about intentions, emotions and ideas.

Two hypotheses judged as reasonable are made for this transposition:

- analogy between the "geometrical dynamics" of the real brain (it is about modeling of human brain when authors speak geometry) and of the artificial brain. For one, flows of data refer to complex images, almost continuous; for the other, there are dynamical graphs whose deformations (introducing 'emotions') are evaluated topologically;
- reduction of combinatorial complexity of the real brain by positioning it at symbolic and pre-language level into computable domain.

A first implementation is reported on equipping the Sony's ERS-7 Aibo robotic dog with a reflective and reactive "brain" working at several levels (Camus and Cardon, [7]).

Aibo sensors for touch & distance and a video camera allow to process environment data to give a contextual position (scene representation – 1st level). The camera data are processed by an artificial neural network embedded in any of a multi-agents system (several thousands of 'aspectual agents' run on a G4, Cardon, [8]) in order to build a vision ontology linked a the sensor ontology. The second level associates the goals of the robot with its environmental knowledge in order to give priority to some objects or actions in the scene. The third level works on the

multi-agent system morphology to detect on line particular, stable geometrical forms (Campagne, [9]) in order to recognize and classify geometrical forms as 'emotions' generated by the robot during its evolution in the scene accompanied by recognition of objects and subsequent actions. The fourth level creates a relationship between the 'cognition' and the (re)action ('behavior'). For a cognition degree, there is a succession of actions on different actuators: the more the cognition degree is higher, the more the list of actions is specified. The fifth level is a continual bidirectional interaction and adaptation between the environment and the robot behavior. For each action, there is feedback, a relationship between sensors and actuators. The 'attention' of the robot (in fact its knowledge base) evolves with the number of performed actions. Cognition and action are treated in parallel by the multi-agent system.

The project is developed on an Oz/Mozart shell. Oz is reported as a multi-paradigm language with scripting, object, logic and constraints programming. It allows using paradigms such as the concurrency for developing a multiagent system with asynchronous communication or the constraints programming to create different action plans.

While progresses in developing the novel Aibo's "brain" along the above coordinates will perhaps continue by care of Cardon's younger colleagues, himself appears as the main beneficiary of insight got as team leader upon the (true) human brain itself. Even if we do not share his rather pessimistic view put as: "Since the permanence of the physical real apprehensible by senses is very strong, the preoccupation to think by man is quite limited, in his civilizations".

Dealing with artificial consciousness one has to keep in mind distinction among different level of analysis. The level of reality refers to what is, the human brain, fragmentally and in general poorly understood. The level of our reflection upon what is, uses words and logics taken from maths or experimental sciences. The level of simulation uses IT artifacts to mimic brain functioning in its known aspects: autonomy, adaptiveness, partly reason/partly emotion – driven a.s.o. Simulation occurs since there are hopes that arranging such IT artifacts in relations deemed to be right the ensemble would begin to exhibit "consciousness facts" replicating symbolically some features of what we (bioengineers, neuroscientists, philosophers) think to be consciousness.

Why not remaining at the reflection level? For what making such a complicate and tedious simulation?

Because, while reflection dissect (analyze) marvelously single elements, simulation puts together various elements in their very interactive dynamics better than our reflection inherently static can do.

Associated risk with simulation is confusion of levels (planes); forgetting to use appropriate quotation marks, one may think that simulation might actually become, as an example, thinking itself but not an inspiring manner to enrich reflection upon.

It is interesting to notice that theology, that is in part science and part faith, while accepting the benefic role of medicine in treating some bodily illnesses, gets very precocious when is about knowing and influencing (and eventually treating) the superior level of human being, the person (and personality) intimately associated with the brain.

In the theological perspective, the person is considered "the highest form of existence and defined before anything

as spirit. A human is an incarnate spirit but his/her spiritual life is defining for a person. In the same manner we use apparatuses and instruments to probe the inner of inanimate things, to reach the deepness of a person we need personal interrelations that in the ideal form represent love. The intimate knowing of a person cannot be entrusted to objects (that is artifacts, though they can help) but to another person only" (Ciobotea, [10]).

If so, best understanding of brain is that given to another brain (e.g. scientific brain seen as a collectivity of brains interacting via communication technology - CT). One can remark the role played in such instance by artifacts (like CT): that of modest but useful adjunct of the real brain approaching (ideally with empathy and generosity) another.

Finally, to Cardon's last questioning "what we must to do about a system generating artificial consciousness facts for itself, having the sensation to generate artificial thoughts for its pleasure and using all the control-command systems and all de knowledge systems as rather gentle tools (our highlight), without any human intervention? "(Cardon, [8]), the answer could not be else than pouring out some axiology into the puzzle next to the machine ontology before detaching the dog (be it the Aibo one) from any human intervention. Problem remains how.

And now our question. An AC system endowed with intentions, emotion and good actuators could be fully autonomous that is entirely disconnected from human control or guidance? Apparently not, because once its power source interrupted everything would stop. Or maybe, similar to actual humans, "It" would become conscious (among others) of such an weakness and consequently would (auto) assure a sub rosa backup power to continue its rapid development of knowledge, experience and capabilities even against the will of its creator?

In the same vein, Hanson [11] put: "If we do not humanize our intelligent machines, then they may eventually be dangerous. To be safe when they "awaken" (by which I mean gain creative, free, adaptive general intelligence), then machines must attain deep understanding and compassion towards people. [...] Only if they have humanlike character, can there be cooperation and peace with such machines. It is not too early to prepare for this eventuality".

V. BIOENGINEERING AND CONSCIOUSNESS TOPICS

Made up to cross the difficult border between medical education and the polytechnic one, biomedical engineering is placed in the privileged position to advance knowledge in the field of human consciousness, in connection with the conundrum whether or not computing machines may become conscious.

Already seen before, bioengineering is central to experiments that investigate direct human-machine interfaces, a topic of neural engineering.

Thus, if the human brain has principle difficulties in understanding it own functioning, signal processing according to information theory rules allied with clinical research on normal subjects may help to climb the staircase to the brain starting from organs apparently less intelligent but prone to be more easily understood.

Cardiovascular bioengineering is today able to distinguish various consciousness states by analyzing heart-related records by means of available knowledge on the control of visceral functions by the brain (Figure 2).

Increasing the cortical control on visceral regulation that

conventionally is called autonomic represents one main aspect of corticalization of our species, in which Stefan Milcu [12] saw the neurophysiologic mechanism of human being's evolution including consciousness and conscience.

Developed by exercising information, the cortex has already spread its "antennae" to the lower floor of the brain, the brainstem regulating the vegetative life, and further on by the cranial nerves to the peripheral organs. According to M. Draganescu [13]: "It would be possible that mental processes get manifest by such extensions throughout the body". It stands for a philosophical inference confirmed at least at the heart's level.

Question remains upon the finality of such influence or control exerted by cortex upon "lower-minded" organs. Auto-assuring the best functioning conditions given the multiple circular feedback loops relating brain and "subjacent" physiological machinery? If so, best functioning refers to which criteria: physical effectiveness, mental performance, emotional refinement or higher propensity to moral (read spiritual) values? If the latter proves as true, entering the regulatory loops by gentle means, natural (as breathing pattern control) or artificial (as noninvasive, remote influencing the heart rhythm) could hold promise for human being improvement without appealing to artificial consciousness.

VI. CONCLUSION

Bioengineering and information science&technology certainly advance towards artificial consciousness.

How benefic for humanity is yet not clear. It depends on our capacity to inseminate a machine transposition of natural ethics, at the same time with increasing machine autonomy. Complete autonomy should superpose to a free will (libre arbitre) having behind a machine axiology built at the same time. Neglecting or postponing the latter might associate catastrophic escaping from any human control.

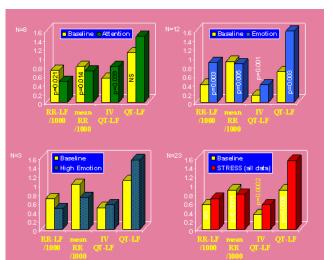


Fig. 2 - Repercussion at heart of consciousness states investigated in 23 healthy young people, 19-21 years, studied in relaxation sitting (baseline), under concentrated attention induced by an arithmetic test without constraint of time (labeled as attention), and under emotion or strong emotion induced by time constraint (emotion/high emotion). RR - the heart period; RR-LF/QT-LF - the fractions of low frequency (0.04-0.15 Hz) in the spectra of variability of heart period or of QT interval in the electrocardiogram; IV QT-LF - the fraction of low frequency in the spectrum of variability of the QT interval from which heart rate influences were extracted using cross spectral techniques – is an emerging indicator of sympathetic control of ventricles (idioventricular). P indicates significant differences between group averages. The idioventricular sympathetic control as expressed by IV QT-LF respond significantly to stress whatever

its nuances among concentrated attention, moderate or strong emotion. RR-LF clearly distinguishes between focused attention and emotion. These states of consciousness can not be discriminated with the same clarity using cerebral electrical activity noninvasively recorded on the scalp. Consciousness states and moods deeply influence the physiological machinery [14]. Since sympathetic ventricular overdrive is arrhythmogenic, such studies may offer a track for risk detection and prevention of sudden cardiac death in apparently healthy people (not known as cardiac patients) under sustained psycho-social stress [15].

While conscious people without conscience are unfortunately too frequent today, humanity cannot afford a machine reply of its brain developing exponentially capabilities and power outside of any moral.

On the contrary, if we succeed to seed at the right time a 'moral principle' into the machine we could enjoy a prototype of pure or ideal consciousness, escaping from biological impulsions and restrictions, that might guide or emulate humanity's struggle towards a true Society of the Conscience.

REFERENCES

- [1] Versace M., Chandler B., MoNETA: A mind made from memristors, *IEEE Spectrum Robotics/Artif. Intell.*, December 2010.
- [2] Draganescu M., Neural engineering and neuroelectronics facing artificial consciousness, *Key note to the INGIMED II Conference*, Bucharest, 2001.
- [3] Draganescu M., Societatea Constiintei, *Raport de cercetare*, *Academia Romana*, Bucuresti, 2003.
- [4] Koch C., Tononi G., Can machines be conscious? *IEEE Spectrum: special report on singularity*, June 2008..
- [5] Horgan J., The consciousness conundrum, *IEEE* Spectrum: special report on singularity, June 2008
- [6] Cardon A., Camus M., Campagne J-C., System generating consciousness facts, 2005, http://abrainproject.googlepages.com
- [7] Camus M., Cardon A., Towards emotional decisionmaking, Innovative Concepts for Autonomic and Agent-Based Systems, *LNAI 3825*, Springer, 2007.
- [8] Cardon A., Artificial consciousness: the hard problem, *ISC Lyon*, June 2008.
- [9] Campagne J.-C., Systèmes multi-agents et morphologie, Thèse de doctorat en informatique de l'Université de Paris 6, septembre 2005.
- [10] Ciobotea IPS D., A vindeca și a ajuta în tradiția Bisericii Răsăritene, Lucrarile Simpozionului de oncologie, Clinica pentru biologie tumorală (Klinik für Tumorbiologie), Universitatea din Freiburg in Breisgau, iunie 2003.
- [11] Hanson G., Why we should build humanlike robots, *IEEE Spectrum Robotics*, April 2011.
- [12] Milcu St., Minte si materie, Seminarul de bioinginerie ISPB, Bucuresti, 1994.
- [13] Draganescu M., Semn si semnal. Key note la Sesiunea Academiei Romane, iulie 1986.
- [14] Negoescu R., Bioengineering spots heart repercussions of mental processes: sudden cardiac death corde indemno is preventable, *Proceedings of the Romanian Academy, Series A: Maths, Phys, Tech Sci, Inform Scie, Volume 4*, *Number 1*, *pp. 65-73*, Bucharest, 2003.
- [15] Negoescu R., Dinca-Panaitescu S. Processing highresolution ECG facsimiles to detect ventricular sympathetic overdrive & to prevent sudden death corde indemno, A review, *ECIT*, Iasi, 2008.