"Controllare" necesse est

L. Keviczky

Institute of Computer Science and Automation
MTA-BME Control Engineering Research Group
Hungarian Academy of Sciences
H-1111 Budapest, Kende u 13-17, HUNGARY

1. Introduction

It is still a memorable event for me when I had my first public appearance at the academy as the youngest director of an institute. Listening the problems of my colleagues and leaders, I used the following classical quotation in Latin in my comment: "Navigare necesse est". Everybody has to make his/her duties, has to continue his/her research work, the efficiency is the sum of particular successes, and I undertook my part. Since this sentence was often cited later by our Secretary General, my predecessor, that is why I used the artificial version of the quotation as the title of this article what seems to be a logical characterization of the subject to be presented. (Unfortunately there is no precise translation of the English word "control" in the Latin vocabulary, therefore relying on the on-line ethimological dictionaries I took the courage to use the word "controllare" what comes rather from the Italian language, but it can be interpreted.) I would like to tell something for everybody in all area, and I would not be happy if my colleagues working on the field of liberal arts would turn over the pages, I hope to write them, too something interesting.

Actually our modern, accustomed comfortable life cannot be imagined without automatic instruments, equipment and processes. In many cases it seems to be quite obvious since we all have up-to-date or less up-to-date equipments with pushing buttons, thus we switch **on or off**, or set/adjust something where the technology behind is invisible. That is why it cannot be expected from anyone to think of how many control loops, processes start when, for example, a radio, TV, video, DVD or any device is working, think of the very simple equipment like the iron, the hot water tank or the single gas heater. This fact is more valid for subject areas like the functioning of living creatures, what kinds of inner control loops keep them in life, what phenomena control the relation of demand and supply or the global environmental processes of the globe. In the devices of the show business a couple of dozen, in the airplanes a couple of hundred, in large industrial complexes a couple of thousand control loops operate ensuring the required services, quality and stability without causing risk in the environment.

We used to learn even in our early age that there exist devices which cannot work without control: such equipment is the two-wheel bike. Without powering and steering we can fall down providing the direct painful experience of instability.

In the case of the umbrella or rod balanced on the fingertip even nobody is thinking of the very similar problem of keeping the airplane in the air and this thought comes back again at adulthood traveling, especially if someone is afraid of traveling by plane (like me). Only very few persons try to balance two umbrellas or rods on each other or ride a one wheel bike. Only jugglers having special capabilities can do that. Such capabilities are required to control the less stable flying objects like helicopters. The solution of the above mentioned technical-technological tasks can be easily picked up during conventional control engineering courses and can be solved by computers routinely.

2. Basic notations and concepts

Our subject area is mostly called *systems and control theory* (technique) in all over the world. The common language uses the word *automation*, what can be better understood, but its meaning is much more limited than the previous one. Next mostly control procedures of technical systems will be discussed but the aim is to bring the subject close to the representatives of all subject areas. The control of industrial production has significant importance in ensuring better product quality, minimizing energy consumption, increasing safety and decreasing environmental pollution. In these production processes material and energy conversion processes are going on. The control has to ensure to start the process, to maintain it appropriately and to stop it. In a thermal power station, e.g., the chemical energy of the coal is transformed to thermal energy. The heat is used to produce steam. The steam drives the steam turbine, producing mechanical turnover rate energy. The steam turbine turns the swivel of the phase generator in the magnetic sphere of the stator what is resulting in electricity (electric current). These processes have to be kept under prescribed condition. During the above mentioned energy production processes the electrical energy has to be produced with prescribed voltage level and frequency under varying daily load. The processes must be stopped in a safety way. The maintenance of the process usually means to keep certain physical quantities on constant value or to change them according to given principle. Such physical quantities may be: the temperature of the medium or its pressure, the compound of a material, the rev of a machine, the angle position of a spindle, the level of a liquid tank, etc. Thus the control is what provides intervention into the process in order to start it, to maintain it and to ensure the required runoff, or to stop it.

The control is based on the information obtained from the process or from its environment by observation, sensing and measurement. To measure the different physical quantities measuring instruments are required. The decision on how to intervene the process is based on the control goal and the information obtained from the process and its environment. It is quite common that during the control processes of high energy are usually influenced by effects of low energy. Control engineering is dealing with principles of control systems, with investigation methods of control operations, with the design and implementation of control systems.

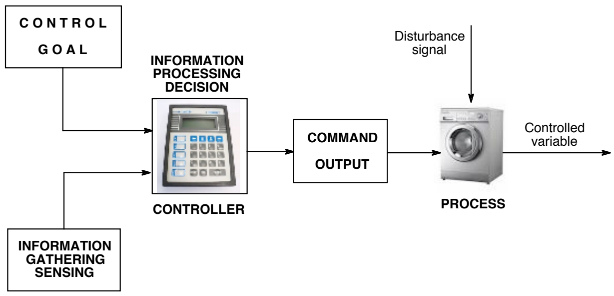

Figure 1. The operational scheme of the control system

The control process consists of the following operations (Fig.1):

- Sensing: getting information on the process to be controlled and on its environment
- Decision: processing the collected information and making decision based on the control goal: determining the control algorithm if it is performed by computer
- Action: command for the intervention, sending the command: intervening the process according to our goal

The case when the information not directly comes from sensing the controlled signal is called *guidance* or open-loop control. An example for this situation is the washing machine controlled according to a time schedule to perform the different consecutive tasks like washing, rinsing and centrifuging. The output (the cleanness of the cloth) is not measured. If the heating of the room is set directly based on the outer temperature (and not on the measured room temperature) then it is also called guiding.

The case when the information directly comes from sensing the controlled signal is called *control* or feedback or closed-loop control. Fig. 2 gives the operation scheme of a control loop. The basic feature of the control, i.e., the *feedback* comes from the comparison of the reference signal and the sensed controlled signal. (Several schemes exist for the general control problem, but all have a common feature what is the feedback.) Remark that the negative feedback is known in several subject areas from the cognitive psychology through the learning and recognizing processes of the basic medical-biological, environmental and physical-chemical systems.

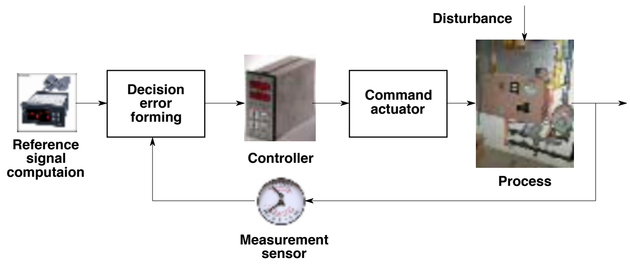

Figure 2. The block scheme of closed-loop control

3. Comparison of guiding and control

If the relationship between the control signal and the controlled attribute is exactly known and reliable information is available on all elements of the control loop and on disturbances, then the guidance can provide satisfactory operation. However, if this information is false, the guidance will not be satisfactory. Guidance provides low price solution, since it does not use costly sensors to measure the controlled quantity, but applies a priori information or information obtained from outer quantities to making decision. In case of guidance stability problems do not occur.

Closed-loop control is more expensive than guidance. The controlled attribute is measured and the intervention is based on the difference between the reference signal and the sensed output. The control can perform reference signal tracking and disturbance-rejection. Since the controlled attribute is influenced by the disturbance, the control prevents the effect of the unforeseen disturbances and compensates the effect of the parameter uncertainties coming from the uncertainty of the process model. The control eliminates the deviation of the output from the required value independently what kind of effect caused the difference. Due to the negative feedback, however, stability problems, unwanted oscillations in the system may occur. The system can be stabilized by the proper design of the controller.

4. A short history

The application of negative feedback cannot be considered a new phenomenon, it has been already applied from the antique Greeks. Looking back to the development of control engineering some tendencies can be observed. The application of negative feedback can be connected to the solution of engineering tasks. The development of control engineering is closely connected to the practical cases what expected certain solution in the different era of mankind history. Certain eras of the history had significant influence on the development of control engineering:

- Antique Greek and Arabic culture (BC ~300 – AC ~1200),
- Era of the industrial revolution (the years of 1770, but it started around 1600)
- Start of telecommunication (1910-1945)
- Appearance of the computer, start of space research (1957-)
- Fusion of control, computer technique and communication, appearance of high level intelligence in the decisions (2000-)

Considering the above eras it can be stated the mankind was looking for its place in space and time then tried to make the life and the environment more comfortable, what was supported by industrial development. Then by the help of communication it founded its place and status in the society then tries to make connection with the whole world.

Even the antique Greeks used different automata. One of the first control systems was the water clock of KTESIBIOS in Alexandria (BC 270). The construction used a float to sense the level of the tank and to keep it on a constant value. If the level of the tank decreased a valve was opened and the tank was re-filled. The constant water level ensured that the quantity of the outlet water from the tank was constant. The outlet water was put into a second tank. This tank was filled proportionally to the time. PHILON (BC 250) in Bezant also used a float to control the oil level of the oil lamp. HERON in Alexandria (AC 1st century) applied a similar instrument for level control, proportioning the wine and opening the doors of churches, etc.

The Arabic engineers used several float control instruments around AC 800 and 1200. They also discovered the switching-on and off controllers.

By the construction of the mechanical clock mechanism the float water clocks sank into oblivion. The principle of control found newer application during the industrial revolution.

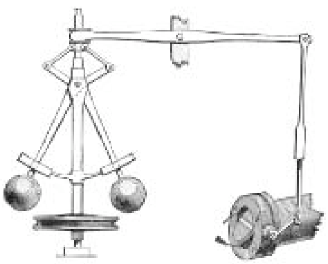

Figure 3. Tumble control

In the era of the industrial revolution several automatic equipment were elaborated. In these equipment automatic level, temperature, pressure and velocity control tasks were solved. Even from the years 1600 different control applications appeared (rpm control of wind mills, temperature control of furnaces (Cornelis DREBBEL), pressure control (PAPIN), etc.) The beginning of the industrial revolution is marked by the discovery of the steam engine (SAVERY and NEWCOMEN, ~1700). The radial force controller of James WATT can be considered the first industrial application what is used to control the rpm of the steam engine (Fig. 3.) The position of the radial force sensor depends on the rpm of the steam engine. The sensor sets the position of the steam piston via a motive lever arm by influencing the quantity of the run in steam, i.e., the rpm of the steam engine. (The interesting thing is that more than further 100 years passed until MAXWELL could give the exact description of the system by differential equations system.)

After the industrial revolution the big step in the development of control engineering was the mathematical description of the control loops what made possible to investigate the behavior of the control systems more seriously and exactly.

A new chapter of control engineering has started by the discovery of the phone and by the application of the feedback operational amplifiers to compensate the attenuation appearing in the information transfer.

During the Second World War several precision control systems were elaborated, e.g., automatic airplane control systems, radar antenna setting systems, the control devices of submarines, etc. These techniques were also applied later in the industrial production. The proliferation of the computers opened a new era in the development of the control systems. The computer appears not only as an outer equipment by which the design can be performed more easily and thoroughly, but it is embedded as a part of the control loop in the real-time applications. The process and the process control computer are connected via peripheries and the control signal computed by software in each sampling time is put to the input of the process. Thus the computer is an integral organic part of the control loop. New software and hardware elements came out to ensure such level of operation.

Besides the continuous control systems, the computer process control systems gained more and more ground when the process and the process control computer are connected via A/D (analogous/digital) and D/A (digital/analogous) converters. The essential functions of the control are performed in real-time by the computer, repeating them in each sampling time. In industrial process control systems mostly distributed control systems are applied where the control systems distributed in space are cooperating in a coordinated way.

In the majority of the most modern airplanes (e.g., the Airbus series) not only the (electro) mechanical equipment forward the commands of the robot pilot, but redundant, multiplied computer nets (fly-by-wire) are operating. This kind of operation slowly also appears in the most modern cars. The automation of the transportation on the highways is nowadays an interesting task and some partial results have already been achieved.

Industrial robots performing precision tasks have also appeared. The robot is an automated machine controlled by a computer. In many cases the robots try to imitate human behaviors, e.g., the robot manipulator imitates the movement of the hand. The moving robots are usually equipped with certain intelligence, like the recognition of the obstacle during the movement and their avoidance.

Space research means newer challenge for the control systems. Putting the space shuttles, artificial space objects to a trajectory and targeting require precise adapting to the environment, learning control systems, where the safe operation is also very important.

The subject area of control engineering deals with construction, analysis and synthesis of control systems. The classical era of control engineering (till 1960) provided the basic principles of the operation, analysis and synthesis of control based on negative feedback.

The modern era of control engineering (1960-1980) concentrated on state-space description of control systems and on the design methods based on the above description.

In the years 1970-1980 learning and parameter adaptive control systems were developed which can operate optimally even in varying circumstances. This situation forced to develop the identification and estimation procedures of the process models.

In the last decade the design of robust, less sensitive for the parameter changing, reliable control came to the front. The control of nonlinear systems, the intelligent learning systems recognizing the changes of the environment and capable to adopt them, the control systems operating in network opened new directions in control engineering.

Control engineering is still developing dynamically nowadays. The new equipment and techniques raise new theoretical questions and enable new applications. By the application of advanced sensors (high precision cameras, microscopes, the most different material behaviors like compounds or movement parameters like velocity, acceleration, position, etc.) or by their combination (sensor, detector arrays) the possibility of the decision put on the top of the classical control system became our everyday practice. The automatic system of the Mars-robot cannot be imagined without high level intelligence: it senses the position, the environment, its facilities and solves the control task under the requirements of its mission.

5. Interdisciplinary approach

Control engineering is an interdisciplinary science area. The operation of the process has to be understood what requires physical, chemical, biological, etc., knowledge. Mathematical skills are necessary to the modeling, analysis and synthesis. The investigation of the control loop requires knowledge on the signals, systems and on the behavior of the systems having negative feedback. During the design reasonable considerations and essential constrains have to be taken into account. The design has to consider economic efficiency, safety, environmental protection points of view. A complex control task needs the coordinated works of different experts. Since the practical realization is done more and more frequently by computer, therefore the role of the experts on informatics has significant importance.

The interdisciplinary character means that several science areas use the concept of control theory and can explain their significant phenomena. The control of the human body, i.e. the control of the blood pressure, blood sugar, some hormone characteristics, is a complex multivariable control loop what is hard to set, similarly to the technical systems. Who would think that the control of a logistic storehouse, the supply and demand control of a market and the level control of a tank are analogous problems?

The way of thinking of system theory is going hands in hands with control theory, according to which the processes are dynamic having inputs, outputs and inner state variables, provides good support, basically rational approach and methodology for several science subjects.

6. Constrains of the control

In spite of the very optimistic picture unfortunately not all kind of problems can be solved by control. The process to be controlled is often sluggish (inactive), inert, dynamically slow and the task is to speed it up. This can be done only if some extra energy (over-excitation) is injected to the process for shorter or longer time in a range computed from the closed control loop. Usually this is prevented from some practical reasons. If someone has seen a giant closing valve of the Friendship Crude-oil Pipeline in Hungary then it becomes obvious that it cannot be closed in some seconds and it cannot be switched on and off within a short time. This is an extreme example but everyone can imagine that, depending on the character of the task, the volume of the above mentioned extra energy is limited due to the intervening equipment. This constrain, of course, determines how fast acceleration can be made. The maneuvering possibilities of the fighting airplanes are not usually constrained by technical limits but by the g-load of the human body what can be tolerated without loss of consciousness.

In certain cases the solution does not depend on us but on the process to be controlled. It is interesting that not the stabilization or suitable control of the unstable processes (see the case of the balanced rod, i.e., the inverted pendulum) is the real problem because the stabilization is a routine task. Other dynamic features of the process might have vital importance and they cannot be modified: these features are called invariant properties and their effect cannot be eliminated. From these two are presented here.

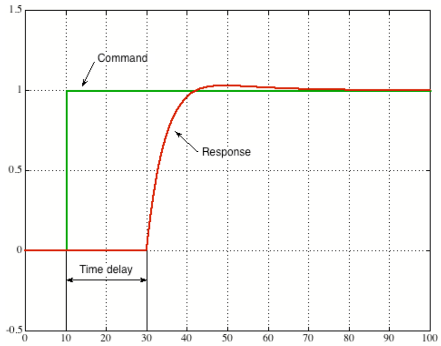

Figure 4. The response of the time delay process

One of the basic invariant properties is the so-called time delay or dead time what can be derived from the transportation time of the technology, from the spreading velocity of the signal or from the evaluation time of the complex measurements. The time delay is an invariant process property what cannot be compensated, eliminated, only its effect can be attenuated. The main problem is that the process does not react, does not give any answer to the command during the time delay, and in order to operate the control properly the first reaction has to be waited, the first answer has to appear, thus the process must be slowed down and the closed loop must be made more cautious. Most probably everybody has experienced that if, by taking a shower, the hot water tap is far from the shower rose and the required water temperature can be set only by cautious twisting. Several time delay processes are known, e.g., the higher education belongs to them: the result can be seen only after some years depending on the function of the education time, therefore the system and control engineering experts escape from reformers having blazing eyes who would like to perform certain kind of reforms every year, because they know that this is the most sure method of introducing instability (chaos). My first professor in control engineering, the vice-president of the Academy that time, Frigyes CSÁKI resigned from the rector position when such a reform was forced what was opposite to his persuasion. But I also could mention the life-stock farming (animal husbandry), where the period of pregnancy is some months when the result can be seen only after this. In spite of this several monthly or quarterly stimulating strategy is forced what is against the theory. It can be stated that in frequency higher than what corresponds to the time delay, the control cannot operate. The quality of the control does not depend on us but it does on the process. The determination of the control algorithm needs thorough expertise.

If the time delay increases in great extent finally it prevents to apply closed control loop. This is the case, for example, for the Mars robot where the remote control practically cannot be applied. (In the case of the Moon it would be performed with some quality loss, see the regarding historical steps.) The only solution is to extend the automatic equipment with much higher level of intelligence and make the decision locally.

The other class of the processes is called non-minimum phase ones which have also invariant properties. This complex name can be easily explained if the step response of the process is presented.

The response of this system first starts in an opposite direction (not in the direction required by the command) then it changes the direction and reaches its steady state (Fig. 5). If there is not enough "patience" then the consequence of the intervention can be catastrophic: unstability. (Unfortunately this happened in the case of Chernobyl where the time constant, the dynamics of the change was too slow and in the case of switched-off control system the staff making experiments could not state whether the initial special part of the transient was over or not and they made a wrong decision.) The similarity to the time delay processes can be recognized. (The time delay processes are usually modeled by non-minimum phase approach. The experts working with electronic equipment know well the all-pass filter which belongs to this model class.) In the present technologies about 10% of the processes have this property where the quality of the control does not depend on us but it does on the process. The design of the controller is a difficult task.

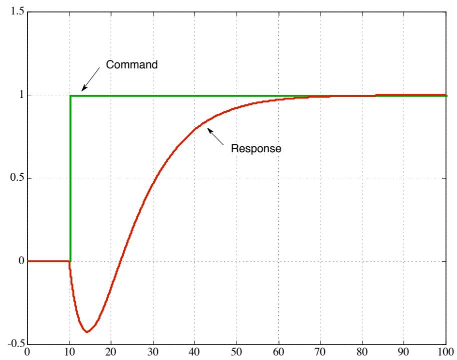

Figure 5. The response of a non-minimum phase process

By the widely applied digital computers it can be stated that almost all kinds of control algorithms can be performed. These so-called sampled (or discrete-time) systems practically substitute the continuous processes with their discrete-time models. In the case of such models almost 90% of the models represent non-minimum phase properties. Therefore the design of these computer controlled systems needs thorough theoretical knowledge.

It is worth to mention that the optimal control tries to apply the inverse of the process dynamic behavior in a certain extent. Unfortunately the inverse of the mentioned time delay and the non-minimum phase processes cannot be realized or they are unstable. To find the nearest inverse what provides optimal control under realizable conditions is always a mathematical challenge.

7. Complex problems

Naturally the above problems increase if the goal is to control more than one variable (*multivariable systems*). In a paper for wide public it is not easy to illustrate the difficulties arising when the dimension of the variables and the parameters of the design are increasing significantly. Here we refer only to the fact that even in the case of a dozen intervention and variables to be controlled the decision whether a certain intervention can influence the prescribed property cannot be made by human capabilities. In complex industrial processes like in the oil refinery, the number of variables to be controlled can reach several hundreds. The Hungarian origin Rudolf KÁLMÁN made a vital and pioneer activity to elaborate the basic elements of this kind of investigations, introducing the concept of controllability and observability. Here the vectors take the role of the scalar variables and the parameters appear in matrices. Even for linear systems the elaboration of the different multivariable process models and the feasible control solutions cannot be discussed without deep mathematical knowledge.

It has not been mentioned before that the briefly discussed solutions above use the linear approach of the processes to design the control system. The real systems, however, usually have complex nonlinear properties, so the above solutions can be used only in a given region when the linear approach can be considered valid. Due to the manifold character of the nonlinear systems it is not easy to develop a general theory, but more and more applications can be recognized. Those who have a car with board computer for petrol dosage can be sure that a real nonlinear controller solves this very useful and essential task. Namely the quantity of the required fuel depends on the service data through a very complex nonlinear expression, what cannot be linearized or is not worth to linearize. The models and the control algorithms of nonlinear systems also need high level mathematical expertise.

In general the design of the controllers requires the knowledge of the process model with acceptable accuracy. This model can be obtained in many different ways, but it can also be computed from the observed input and output signals of the process (by estimation, identification, etc.), what is used at the next step of control design. This procedure is called learning, adaptive control. This subject was very trendy in the years of 1970s and 1980s then some vital recognition, conclusion could be made. There is an antagonistic conflict between the information need of the learning (what requires big changes in the signals of the control loop) and the conventional need of the control (after fast transient the signals are not changing). The real goal of the learning control is to develop an optimal controller by learning the continuously changing process parameters. Here comes the next antagonistic conflict. By learning the new parameters the old ones have to be forgotten. This property can be tuned by the forgetting factor. If the forgetting is too fast then the algorithm can handle fast parameter changing but the risk is high: if new information does not come then the algorithm forgets the learned model and the designed controller becomes useless. If the forgetting is too slow then the adaptive model cannot track the changes of the process. Thus the learning algorithm has to be prepared to forget only if it gets new information otherwise keep the knowledge. It is very difficult to provide the capabilities summarized above with mathematically correct methods. The industrial practice could not accept these contradictory requirements and high standard theoretical background: " We cannot put an expert beside all adaptive controllers who could tell what is the problem" – it is complained. In certain technologies these problems can be solved by the so-called self-tuning compact controllers. (The first inventor was K.J. ÅSTRÖM.) This means that pushing a button the prescribed excitation of the process starts, the learning process runs its course and the controller sets its own optimal parameters. It is far from the continuous learning adaptive control, because this self-tuning process is very rarely required.

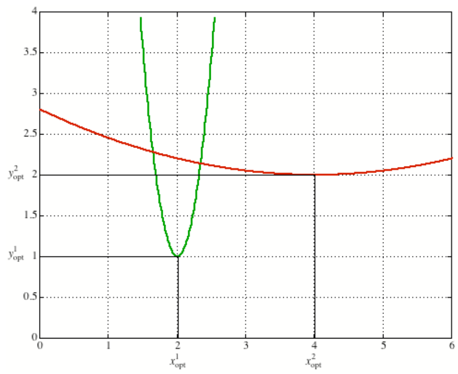

Figure 6. Different shapes around the optimum points

It is rewarding to mention the situation regarding the concept of the optimality during the past decades. Optimality means the minimum of a certain quality property what is determined by mathematical expression. The controller what ensures this minimum is called optimal controller. In the previous decades tremendous theoretical efforts were made to calculate the optimum or to design optimal controllers. Similarly to the learning capabilities the need of practice elaborated the concept of the "useful" optimality. Namely in the practice the optimum $\left(x^2_{\text{opt}}, y^2_{\text{opt}}\right) = (4,2)$ is better (see Fig. 6) than the smaller optimum $\left(x^1_{\text{opt}}, y^1_{\text{opt}}\right) = (2,1)$. However, in the vicinity of the second optimum the curve (or in multi-dimension the surface) is significantly flatter, insensitive: by the technical terminology it is more *robust*. The real optimum of a certain quality property is important only in a very minor percentage of the practical cases what is often very sensitive, which means that even for a little change we can get far from the optimum value. Such worse optimum when it can be kept in a very wide region and insensitive for the changes of the conditions is more appreciated. Modern control theory uses the concept of the norm of the signals and operators as a criterion for optimality.

The optimality can be interpreted also for the structure of multivariable complex control systems. In the early months of the change of the regime in Hungary, in my lecture on the Academy I made the following remark: "According to system and control theory the centralized control is optimal". There was a deep silence then seeing the scared eyes I continued: "The above statement is valid only if the center has all the necessary information.

This case, however, never happens in the practice. This was forgotten by the leaders of the former unsuccessful political systems (though they put their best to get all the information in the center). The distributed control system may not be optimal from theoretical point of view but it is more reliable and robust, where the control is based on information centers in the vicinity of certain suitable centers. Of course these centers need coordination and cooperation. This is the found of the democracy! The similarity can be recognized even in the social sciences. But do not forget that the complexity of the mankind, social systems is much higher than that of the technical systems, but the analogy can be observed". (The audience took breath.)

8. Conclusions

The control technologies are usually embedded and hidden, the majority of the users knows nothing about their operation. The professionals often complain about the hidden character because it needs great effort to convince the decision makers, research fund providers, etc. about the importance of the field. The deeper thinkers like to mention the technologies C^3 (or CCC or 3xC) where the different C's are based on English expressions like Control, Computation and Communication. The modern technical systems really realize the high level combination of the three fields therefore it is difficult to judge which subject area is more complex, this is determined by the task itself. The computer aided digital technique shots all three fields therefore sometime it is extended with I, which means the programmed intelligence.

In the most modern control systems, the goal of the control is determined by complex algorithms and we cannot speak anymore about the classical comparison of the measured and prescribed values. It is expected from the controller built in the car to keep it on the road, to recognize the track of the road under extreme weather conditions, to take over the steering for a while if the driver is unable to do that, and to sense the traffic situations on the highways keeping safe distance in the front and back and to keep the car in such a traffic situation, where at the same time the consumption, tracking, safety and lane violation conditions are optimized. The technology of the control still remains hidden, and by the development of the methods it will be more unseen by the customers. This is the demand of the people, they do not want to think about what happens if the control stops. Therefore our responsibility is high, the comfort, safety, opportunity of the future and the quality of our life depends on us.

Therefore:

"*Controllare*" *necesse est.*

László Keviczky