ARTICLE

ARTHUR L STEVENS

Control Systems in Engineering, Nature and Society

INTRODUCTION

This article discusses control from an engineering perspective, in particular the ***automatic*** control of dynamical systems. Most people come into contact with control systems every day; for example, a toaster, a home thermostat, an elevator and a motor vehicle cruise control all operate by employing control systems.

In the context discussed here, the purpose of a control system is to modify the behaviour of a dynamical process without any physical modification of the process itself.

One can sometimes change the behaviour of a system by intervening to rebuild it, but control theory is about changing its behaviour in more subtle ways, without destroying and rebuilding the system. In society, revolutionaries often propose that a system must be destroyed in order for it to be rebuilt in a more acceptable form but, by including appropriate control mechanisms, a society can be made adaptable, thereby rendering radical intervention unnecessary.

WHAT IS AN ENGINEERING CONTROL SYSTEM?

Engineered control systems date back to ancient times. Hero of Alexandria created a system to automatically open and close temple doors.[1] When a fire was lit upon the altar, the resulting hot air forced water from a reservoir into a bucket connected to a pulley arrangement which was part of a door-opening mechanism. When the bucket was sufficiently heavy, the doors were pulled open. The doors could be closed by dousing the fire, allowing the air to cool, reducing the counter-pressure and siphoning the water from the bucket back into the reservoir.

In 17th-century Europe, the position of windmills was controlled using geared steering devices to point them into the wind, and Cornelis Drebbel developed an automatic temperature-controlled incubator for hatching chickens.[2] During the industrial revolution, steam engines used flyball governors to regulate speed.[3]

The problem of flight was finally solved by the Wright brothers in 1903 when, in addition to meeting the challenges of aerodynamic lift and power requirements, they added the missing ingredients of stability and control. In modern times, the most comprehensive mathematical study of the interactions between communication and control in machines, organisms and society was undertaken by Norbert Wiener.[4] Wiener called this new scientific discipline "cybernetics".

As technology advances, motor vehicles are using more automatic control systems such as antilock braking and automatic roll-stabilisation systems. In addition, robots that manufacture many types of commercial products, from cars to computer circuit boards, also use control systems.

Control system engineers use block diagrams as a visual aid to describe and analyse control systems and the subsystems that contain them.

The block diagram of a simple dynamical process ***P*** is shown in Figure 1, where ***u*** is the input and ***r*** is the output.

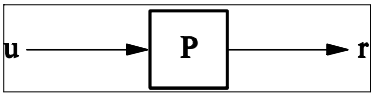

Figure 1: A block diagram representation of a simple process.

The process is dynamic, so the variables ***u*** and ***r*** are real physical quantities that vary with time, such as voltages or sound pressures, and they can therefore be thought of as ***signals***.

The input ***u*** causes the process ***P*** to exhibit some type of behaviour and ***u*** is therefore called the ***actuation*** signal. The response of the process to the actuation signal is represented by the variable ***r***. Strictly speaking, both ***u*** and ***r*** should be written as functions of time – i.e., ***u(t)*** and ***r(t)*** – but to simplify the notation, the time arguments will be omitted.

To describe the behaviour of the process due to a given input, one needs to show the relationship between the input and the output. Typically, a transfer function is used. The transfer function is defined as the ratio of the output to the input:

$$P = r/u \qquad (1)$$

If the transfer function of the process is accurately known, a desired behavioural response ***r*** can be obtained by a suitable choice of the input ***u*** which can be calculated directly using equation (1). If this is the case, the control problem is solved and nothing more needs to be done. Unfortunately, however, there are a number of difficulties with this approach:

- The exact behaviour of ***P*** may not be known or, if ***P*** is a mass-produced item, its behaviour will be affected by manufacturing tolerances.
- The behaviour of ***P*** may be different from that desired, but changing the process may be constrained by the laws of nature or may not be feasible due to cost.
- The behaviour of ***P*** may change with time due to various influences such as mechanical wear, aging and temperature.
- Unwanted signals such as thermal noise may be generated internally by the process.

- The output response is affected by external disturbances acting on the process. For example, wind gusts acting on the dish of a radio telescope cause perturbations in its position.

All of the above effects can be regarded as different manifestations of ***uncertainty*** in the process and its environment. The control system engineer's task is to find some way of obtaining the desired behaviour from the process via a suitable input ***u*** without modification of ***P*** itself. In this case, due to the uncertainty that exists in the actual behaviour of ***r***, it is futile to try and calculate ***u*** from ***P*** to produce a desired ***r***, so an alternative approach is needed.

To determine the actual behaviour ***r*** produced by ***P*** with a given ***u***, a measurement device ***M*** can be used to measure ***r*** directly. The measured value of ***r*** can then be processed by a controlling device ***G***, which changes ***u*** in a way that gives the desired response. The desired response is communicated to the control system in the form of a command signal ***c***. An example of a desired response would be the position of a robotically-controlled cutting torch following a given curve on a sheet of metal.

To force the control system's actual response to match the desired response, the controlling device ***G*** needs to have an indication of how the actual response ***r*** differs from the desired response ***c***.

Figure 2 illustrates a simple way of implementing this idea.

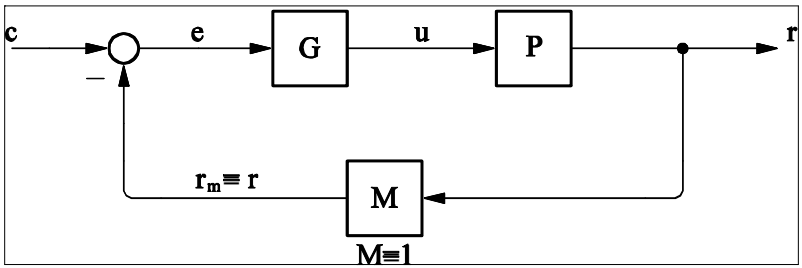

Figure 2: A simple feedback control loop.

The measured value of ***r***, r_m, is subtracted from the command signal ***c*** to produce an error ***e***. The error ***e*** represents the difference between the desired or "commanded" behaviour of the system represented by ***c*** and the actual measured behaviour r_m. In this simple illustrative example, the transfer function of ***M*** is unity, but in real systems it is more complex.

The new system embodies a powerful and important principle, namely ***negative feedback***; the output is measured, fed back and, after subtraction from ***c***, arrives at the process input ***u*** via the controller ***G***. The feedback is termed negative because the signal fed back is ***subtracted*** from the command ***c***.

The error ***e*** is processed by the controller ***G*** which produces the actuation signal ***u***, causing the process ***P*** to produce a corrected value of ***r***. Systems such as those described in Figure 1 (and Figure 2 with ***M=0***), are called "open-loop systems" as there is no measurement of their output behaviour and therefore no feedback loop exists; that is, the feedback loop is open.

The control system engineer's task is to design a suitable controller ***G*** that will take the error ***e*** and produce a suitable actuation signal ***u*** that will change ***r*** in a way that will minimise ***e***. The objective of the controller is, therefore, to continuously monitor ***r***, compare it with ***c*** and adjust ***u*** in a way that causes the response ***r*** to follow the command ***c***.

Ideally, the controller will be able to keep ***e=(c-r)*** at zero, making ***r=c*** and thereby forcing the behaviour of the system to match that of the desired response specified by the signal ***c***. In practice, however, there is always some error present.

An important property of a closed-loop feedback control system is that the quality of its performance is only as good as the quality of its measurements. One cannot control a process better than it can be measured!

At this stage of the discussion, the reader may ask: "If ***c*** is the desired behaviour, why not just use ***c*** directly and dispense with ***P***, ***G*** and ***M***?" To answer this question, consider the following example. A 100-ton radio telescope dish needs to be positioned and moved to follow the motion of a star. Powerful electric motors are used to move the dish in the horizontal (azimuth) and vertical (elevation) directions via a set of gears. In each axis, a power amplifier supplies the required voltage and current to power the motor.

Let ***r*** be the actual elevation of the dish with respect to the earth's horizon and let ***c*** be the desired elevation of the dish as it follows the star.

Figure 3 is a block diagram of the elevation control system.

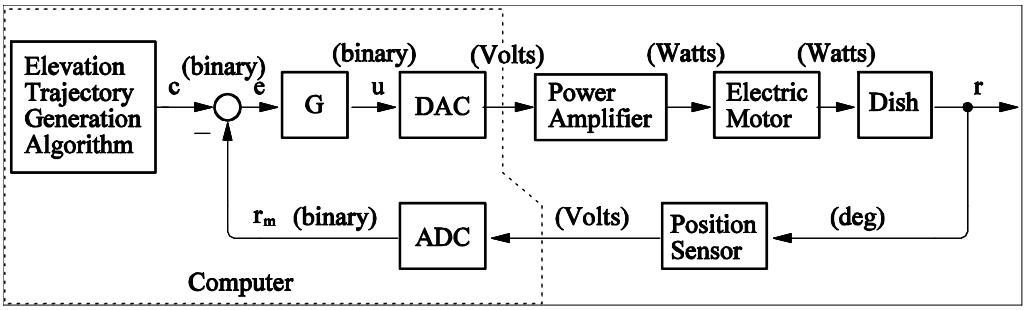

Figure 3: An elevation control loop for a radio telescope dish.

Modern control systems are almost exclusively implemented using computers or embedded microprocessors. Digital computers can only work with binary numbers so, in order to interface with a physical system, the mechanical or electrical analog signals have to be converted to binary quantities using an Analog to Digital Converter (ADC), or from binary quantities to analog ones using a Digital to Analog Converter (DAC).

The position sensor produces a voltage proportional to the angle of elevation. The ADC converts the voltage to a binary number. The elevation trajectory generation algorithm produces a time-based sequence of binary numbers representing the desired elevation at any given time. The digital control algorithm ***G*** calculates the corresponding binary actuation sequence ***u*** from the measured and desired position values. The DAC converts the binary sequence ***u*** to a voltage

signal and the power amplifier amplifies the actuation signal to a high enough power level to move the dish via the electric motor.

If ***G*** is correctly designed, it will strive to force the error ***e*** to zero at all times so that the actual dish position ***r*** will follow the desired position ***c***. Since ***r*** is the actual motion of the heavy dish, it should now be obvious that ***c*** and ***r*** differ substantially in power level and are physically different quantities, and that therefore ***c*** cannot be used directly. In this case, ***c*** is a time-based sequence of binary numbers which causes the output of a 100-ton device to mimic its behaviour.

Another important advantage of a well-designed control system is its ability to ***reject disturbances***. For example, wind gusts acting on the dish will apply forces that will attempt to move it from its desired pointing position. Any change in ***r*** will be measured and corrected by feedback loop action and the unwanted effects of wind forces will be counteracted. The negative feedback property leads to self-correction.

A related property of closed-loop systems is improved ***performance robustness***, which means that the performance of the system tends to become insensitive to any changes in the characteristics of the process ***P***. This allows a high-performance closed-loop system to be constructed using a process with inferior performance. Often an item will be cheaply manufactured with poor tolerances with the assumption that the required performance of the final system can be obtained using feedback control.

AN EVERYDAY CONTROL SYSTEM

A flushing toilet is a simple and widely known example of a feedback control system. After a toilet is flushed, the water tank must be refilled to a suitable level. Imagine how inconvenient it would be if we had to open a tap, wait until the correct level was reached, and then close the tap. Fortunately a simple control system using a valve controlled by a floating ball takes care of this task on our behalf.

Figure 4 is a diagram of the flushing toilet and its simplified block diagram model.

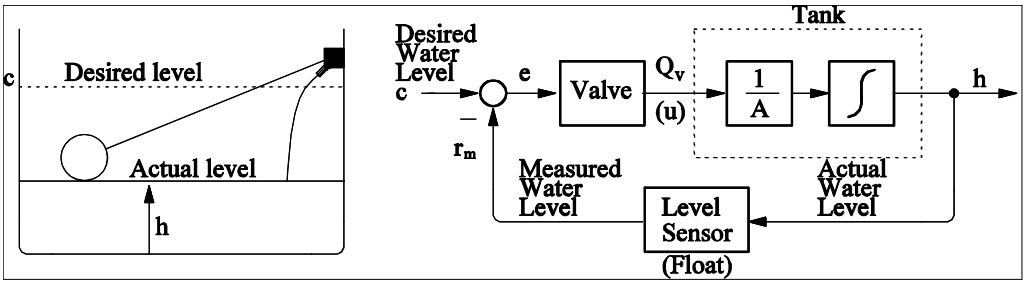

Figure 4: A flushing toilet and its block diagram.

The floating ball is used to measure the water level ***h*** and turn the valve on or off. The rate of rise of the level in the tank is proportional to the flow rate and inversely proportional to the tank area ***A***. The level is given by the integral of the rate of change of level. When the tank is

empty after being flushed, the position of the float near the tank bottom causes the valve to open fully, allowing water to flow into the tank at a rate of Q_v litres per second.

The difference between the desired water level ***c*** and the actual level ***h*** is large and the valve is fully open providing a large actuation signal, that is, a large flow rate. As the water level rises the error decreases, slowing down the flow of water until the error reaches zero, whereupon the valve shuts off and the flow stops ($Q_v=0$). Thus, the negative feedback control system ensures that the actual level reaches the desired level and stays there until the next flush occurs.

THE PROBLEM OF STABILITY

All physical systems are subject to inertia so that they cannot react instantly to sudden changes. Accelerating an object such as a heavy radio telescope dish from rest to a desired speed takes time.

Consider the simple feedback loop model in Figure 2 and imagine that ***r*** is the position of the telescope dish. At any given point in the loop, some time will be needed for a signal to travel from the chosen point around the loop and back again to that point. Thus, information will experience a delay as it travels around the loop. Figure 5 shows the simple feedback system with the loop broken at the input to ***G***.

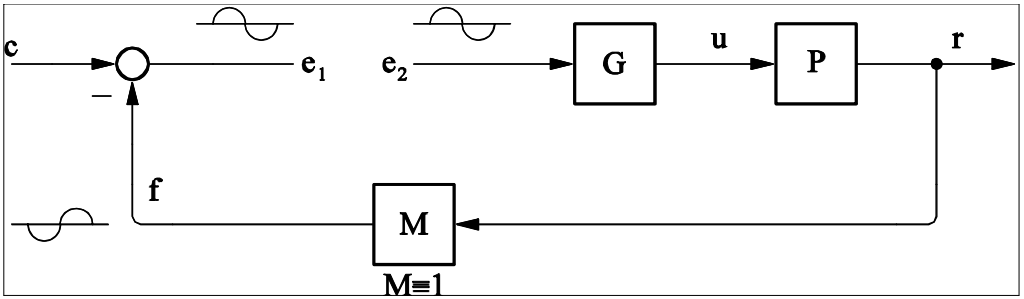

Figure 5: A feedback system with the loop broken.

Now suppose that a sinusoidal signal starts at e_2 and travels around the loop. If the delay around the loop is large enough, a situation could occur where the signal ***f*** at the output of ***M*** has experienced a shift in phase of 180 degrees during its travel around the loop. The signal at ***f*** will be an upside-down version of the original signal at e_2. The negative sign of the subtractor inverts the signal, thereby effectively adding another 180-degree phase shift to the signal so that it is flipped upside down again by the time it reaches e_1.

The returned signal at e_1 and the original signal at e_2 now have the same polarity. In the case where the loop is closed and e_1 is joined to e_2, it should be clear that the signal will circulate around the loop in the form of an oscillation. If there is no amplification of the signal as it travels around the loop, the size of the oscillation will either stay the same or die out. However, if amplification is present, the oscillation will grow and, if left unchecked, could destroy the system.

A negative feedback system can become unstable if amplification is present at the same time that there is a 180-degree phase shift around the loop. The engineer therefore has the responsibility to ensure that the controller ***G*** is designed to correctly apportion the amplification and delay around the loop.

CONTROL SYSTEMS IN NATURE

There are myriads of dynamical systems and processes in nature that make use of automatic control. Plants and animals rely on complex chemical reactions for nourishment and life. In turn, these chemical reactions require the control of parameters such as temperature, pH and enzyme concentrations so that they are kept within certain, often very narrow, ranges.

Movement in the animal kingdom relies heavily on control. Imagine trying to lift a glass and drink from it without continuous control of motion, or trying to visually follow the migration of antelope across the African plains. All of the above activities make use of negative feedback control.

Typical examples of natural control systems are:

- the state of equilibrium between a predator and its prey;
- global weather, which is determined by many complex interacting nonlinear feedback systems;
- human temperature regulation, which is controlled by a feedback process causing the person to either shiver or sweat to keep body temperature fixed to the desired setpoint; and
- the human endocrine system.[5]

CONTROL SYSTEMS IN SOCIETY

In terms of our simple system in Figure 2, a society can be regarded as a feedback control system. The resources, workers, industrial and agricultural capacity represent the process ***P***. The decision-makers and investors are represented by the controller ***G*** and the output measurement by the media ***M***. The role of government as a source of external disturbances is open for debate!

Human learning is error-driven and makes use of negative feedback. A person learns by making mistakes, then taking corrective action to eliminate them. In the classroom situation, a teacher (***G***) provides information and tasks (***u***) to the class (***P***). To measure the performance of the students (***r***), a system of measurement based on examinations (***M***) is used to compare results with the objectives (***c***) to yield a performance error (***e***). The teacher's output is then modified in a direction which reduces the error and, theoretically at least, improves the performance of the students.

The engineer's perspective allows us to understand some social and political phenomena in a new way. The presence or absence of feedback via opinion polls and elections is what makes the difference between democratic and totalitarian systems of government. In a democracy, the control system, although benefiting from feedback control, is nevertheless

vulnerable to propaganda. Distortions of the truth and exaggeration by dishonest or intimidated journalists can lead to societal instability. The feedback signal from the media in this case conveys information opposite to the truth, giving rise to a 180-degree phase shift in the loop. Similarly, exaggeration can be regarded as signal amplification. The two ingredients necessary for instability are now present and chaos is the result.

For example, tenuous information about Saddam Hussein's desire for sophisticated weaponry was exaggerated by the US and UK governments into definite claims (now conclusively known to be false) that Iraq possessed weapons of mass destruction. The resulting control system was therefore ill-conceived.

What benefit can a politician derive from the creation of instability? A system on the verge of instability is very sensitive to perturbations. A practical engineering example is a fighter aircraft. Modern high-performance fighter aircraft are designed to operate on the verge of instability and are difficult to fly without the help of a control system (the autopilot). The benefit is that, because the aircraft is so close to instability during flight, it becomes highly manoeuvrable and is consequently a more agile fighter.

In the case of a political system close to instability, major changes can be more easily put into effect than in a stable system in equilibrium where such changes would normally be resisted by the inertia of the system. Should one wish to introduce legislation that decreases individual freedoms in a society with a long history of respecting such freedoms, then the prior creation of instability would facilitate the changes. One can predict, from this perspective, that a government eager to pass contentious legislation or changing voting patterns would attempt to create and maintain instability – for example by using devices such as "threat levels" which can be manipulated without those affected having the capacity to independently assess whether these signals accurately reflect reality.

Stable feedback systems are important in business and government services. Successful businesses make use of feedback in a closed-loop fashion to improve their marketing and the quality of their products.

Consider the police and those institutions charged with aiding children or the unemployed, as well as the institutions established for the maintenance of physical and mental health. How readily do they seek out and adopt feedback from those they are designed to support? When a police force deems its duty to be discharged by the creation of an emergency number, but has no equivalent dedicated simply to being receptive to the opinions of citizens, then one has a police force destined to get out of touch with the public – one has an open-loop system without any self-correction. The creation of an "us versus them" attitude towards the police is destructive of a sense of community

Given the serious consequences of propagating false information, one may, for instance, feel compelled to set higher standards of corroboration for information emanating from political parties and seek to hold the organs of dissemination to high standards. Imagine if television channels were obliged, by popular demand, to provide actual evidence in every news bulletin that they were not merely acting as a conduit for partisan political propaganda!

Our species has two rival methods for understanding and explaining the working of our universe – science and religion. Figure 6 illustrates "the scientific method" as a feedback control system. The arrow drawn through the block labelled "model" denotes that its characteristics are adjustable and can be altered by the updating block.

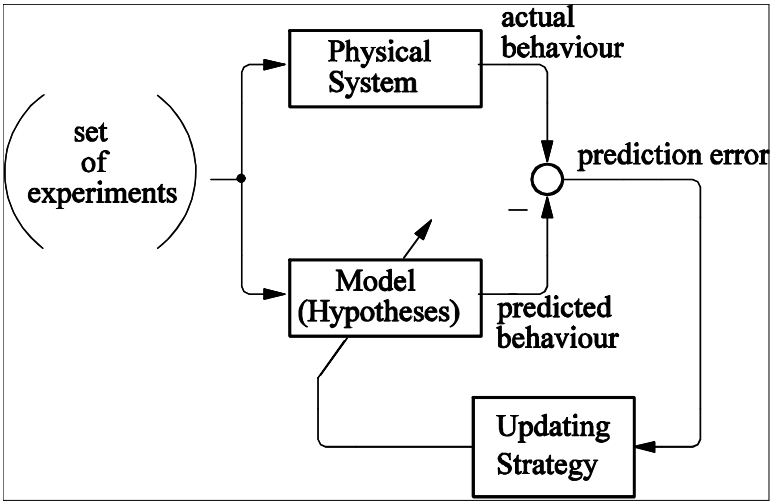

Figure 6: The scientific method as a feedback system.

To develop a useful scientific theory which is able to explain some observed phenomenon, a set of hypotheses are proposed and tested against the actual system under study. To investigate how the physical system works, a set of experiments is performed on both the actual system and the hypothetical system model. The results are then compared.

The experiments used on the two systems will not be of exactly the same form because one cannot, for example, double the temperature of the model – but one can double the value of the term representing temperature in the appropriate equation in the model.

The behaviour predicted by the model is then compared with the behaviour of the actual system and a "prediction error" is formulated. The error is used to try and improve the set of hypotheses in the model in a direction that will lead to a better description of the system's actual observed behaviour. The whole process is repeated until the prediction error is minimised, leading eventually to a reliable and accurate ***theory*** of operation. Thus the scientific method embodies self-correction due to the action of negative feedback. In contrast, dogma (as characterised by explanations based on religion) is an open-loop strategy in that it has no feedback mechanism (***M=0***) and is therefore not self-correcting.

CONCLUSIONS

Feedback control is a fundamental principle of nature and an enabling technology for humankind. The aim of this article has been to demonstrate the pervasiveness of this principle across such diverse fields as engineering, natural science and society.

It is reasonably certain that with the progression of time, future investigations will show that ***every*** complex dynamical system, ranging from biological evolution, psychology and stock market dynamics to the creation of stars and planets in our universe, are all feedback systems. Hopefully, as our awareness of the immense power of this idea is raised, its benefits to society and the environment will ultimately be realised.

ACKNOWLEDGEMENTS

I would like to thank Dr W A Labuschagne of the University of Otago for his insightful comments on an early version of this paper.

1 J J D'Azzo and C Houpis, *Feedback Control System Analysis and Synthesis* (Tokyo: McGraw-Hill Kogakusha, 1966).

2 O Mayr, *The Origins of Feedback Control* (Cambridge, MA: MIT Press, 1970).

3 G C Goodwin, SF Graebe and ME Salgado, *Control System Design* (Upper Saddle River, NJ: Prentice Hall, 2001).

4 N Wiener, *Cybernetics* (Cambridge, MA: MIT Press, 1948).

5 A P Spence and E B Mason, *Human Anatomy and Physiology*, 2nd ed. (Menlo Park, CA: Benjamin Cummins, 1983).

Arthur L "Fred" Stevens is a principal design engineer at STMicroelectronics Inc., involved in the design of electronic systems for hard disk drives. He holds the degrees of BSc (Eng), MSc (Eng) and PhD in Electrical Engineering.