Biological Systems Thinking for Control Engineering Design

D. J. Murray-Smith

Artificial neural networks and genetic algorithms are often quoted in discussions about the contribution of biological systems thinking to engineering design. This paper reviews work on the neuromuscular system, a field in which biological systems thinking could make specific contributions to the development and design of automatic control systems for mechatronics and robotics applications. The paper suggests some specific areas in which a better understanding of this biological control system could be expected to contribute to control engineering design methods in the future. Particular emphasis is given to the nonlinear nature of elements within the neuromuscular system and to processes of neural signal processing, sensing and system adaptivity. Aspects of the biological system that are of particular significance for engineering control systems include sensor fusion, sensor redundancy and parallelism, together with advanced forms of signal processing for adaptive and learning control.

Keywords: biology, control systems, design, neuro-muscular system.

1 Introduction

For many years engineering researchers have been interested in exploring ways in which knowledge about biology can contribute to engineering design. An early stimulus to research efforts in that area came from the much-publicised and innovative work of Norbert Wiener, including his introduction of the word "cybernetics", meaning the study of control principles in man and machine. Later milestones include the publication in the 1960s of the proceedings of an AGARD conference [1] on the theme of "bionics", which included some fascinating contributions that showed just how much biologically-inspired engineering research was under way by then, both in Europe and in North America. More recent examples where biology has provided important links that have led to widely-used design tools, include evolutionary methods of optimisation (such as genetic algorithms and genetic programming) and to artificial neural networks. In both of these areas biological analogies and biological thinking have contributed to the tools as they are used at present, but it is clear that the biological context of much of the original research is now largely irrelevant to those who routinely use the techniques in solving engineering problems. Also, although biological thinking did contribute significantly to their development, it is important to point out that the most widely used forms of artificial neural networks have only some features that reflect properties of networks of real biological neurons. Similarly, in the evolutionary computing field, although the algorithms are firmly based on the principles of survival of the fittest, only some aspects of evolutionary biology have been applied within the forms of genetic algorithm that are in general use today.

In considering living systems from an engineering perspective it is clear that there are some fundamental similarities and differences between natural systems and humanengineered systems. Both types of system show behaviour that is based on the same physical, chemical and thermodynamic laws and principles. On the other hand, natural systems sample a design space through the processes of evolution and optimisation that involves small changes to an existing system. They are also strongly integrated at a variety of levels including the molecular, cellular, organ and eco-system levels. Human-engineered sub-systems are never, at the current level of our technology, integrated as closely as biological systems and these engineering sub-systems can often function independently. Indeed structure and function are related at all scales and levels in biological systems in a way that is seldom approached in present-day man-made systems.

2 Biological control systems

Man-made control systems still lag far behind natural systems in terms of their performance. This is especially clear in mobile robots where human performance in complex tasks, such as tennis playing or riding a standard bicycle, goes beyond the capabilities of any present-day man-made system. Present-day manipulator robots also show a level of performance in terms of control system capabilities that is far below that of the equivalent biological systems. A manipulator robot has a dynamic response that is highly dependent on the load being carried by the arm. Human arms on the other hand show dynamic responses that are largely independent of the loading conditions and have characteristics that strongly suggest adaptive properties in the underlying control system. The question arises of how these adaptive properties are provided from the basic elements of the neuromuscular system and whether some of these properties could be translated into equivalent characteristics in robots.

3 The neuromuscular control system

The biological system that controls posture and movement in mammals is highly complex, with significantly nonlinear and adaptive properties. Although the system presents a considerable challenge to biologists and to engineers, it has for many years attracted much attention because of the potential benefits in terms of the treatment of diseases that can affect our ability to control posture and move effectively.

The neuromuscular system has also been taken into account in modelling the human operator for control tasks where the human-in-the-loop has a critically important role, such as in flying high-performance military aircraft. Much use has been made of relatively simple neuromuscular system sub-models within pilot models used for aircraft handling qualities and flight control systems studies.

Interest shown by engineers in the structure and function of the neuromuscular system has also been very considerable because they have recognised the potential benefits for engineering design. A better understanding of the way in which the central nervous system controls complex movements and the role of reflex systems in the regulation of posture could well lead to the design of better control systems for manipulator robots or walking robots.

It is generally accepted that the neuromuscular control system is organised in a hierarchical fashion. The muscular and skeletal subsystems and the associated sensory elements (neural receptors) lie at the lowest level and are directly involved as elements in the feedback loops that provide reflex actions through the spinal cord. These reflex actions are influenced by a network of interneurones which is thought to provide the coordination and pattern generation elements for the system for control tasks involving more than one muscle. The brain stem, cerebellum, motor cortex and other similar areas of the central nervous system are involved in the upper levels of this hierarchy. Neural communication pathways, both ascending and descending, interconnect elements at the various different levels. Fig. 1 shows a simplified schematic diagram of the system.

Most physiologists and engineers involved in neuromuscular system modelling and experimental investigations have adopted a modular approach in which attempts are made to understand the function and structure of the many different elements of the system. The basic actuator is the anatomical muscle. A single anatomical muscle may contain hundreds of active contractile elements, known as motor units, which are connected in parallel to a common tendon. Each motor unit itself consists of a single motoneurone and a group of about 250 muscle cells that can be activated simultaneously by the motoneurone. Each muscle fibre resembles a form of active nonlinear spring in which the force developed is dependent upon both the external loading and the neural input. At a microscopic level the fibres contain many smaller elements known as fibrils, and each fibril is itself packed with microfilaments which are of two types, thin filaments of actin and thick filaments of myosin. The active contractile properties of muscle are derived from the active sliding of one type of filament over the other.

The signal transmission paths in the neuromuscular system are nerve fibres. A typical nerve fibre has a cell body, which contains the nucleus, and a number of so-called "processes" that may be more than one metre in length. Most nerve cells have several short processes, called dendrites, and a thinner and longer process called the axon. Electrical impulses are generated by a discharge at a receptor organ.

Activity at one point in a nerve fibre leads to activation of adjacent parts of the fibre and at points where a nerve divides the activity is carried into all branches. In electrical terms the activity at a given point takes the form of voltage impulses or "spikes", known as action potentials, separated by periods of inactivity. It is widely believed that neural information is transmitted using a form of pulse-frequency modulation.

Since the transmission of an action potential is an active process that depends upon ion transport across membranes there are usually significant time delays. The velocity of propagation of a neural impulse is directly related to the nerve fibre diameter. The point of connection between one nerve cell and another is known as a synapse. A single impulse in one pre-synaptic fibre may not have sufficient effect to activate the post-synaptic cell and there is a resulting change in the pattern of activity. In many ways a synaptic junction may be regarded as analogous to an ideal summing junction with additional attenuating properties. The pre-synaptic connections are classified as either facilitatory or inhibitory and are equivalent to inputs with positive or negative signs, respectively, at the summing element.

The sensory elements of the stretch reflex feedback system are of two main types. These are the muscle spindle receptors and the tendon organ receptors and they are known to be very different, both anatomically and functionally. It is believed that the muscle spindle receptors have a particularly important role in the stretch reflex system and in the overall control of the complete neuromuscular system. It has for long been accepted that a more complete understanding of the muscle spindle receptor would throw light on many unsolved problems concerning neuromuscular control mechanisms.

The muscle spindle receptors are sense organs that lie in parallel with the main load-bearing (extrafusal) muscle fibres. Each spindle consists of a group of special-purpose muscle fibres, known as intrafusal fibres, with a set of receptors located in the central region of the spindle. Although the complexity of the muscle spindle is considerable and some of



Fig. 1: Diagram of pathways connecting the muscle spindle and its parent muscle to the spinal cord (adapted from diagram in [2])

its features are still not understood, there are some points upon which there is general agreement. It is recognised that intrafusal fibres vary in size and fall into two distinct categories in terms of visible microscopic features, these being known as nuclear bag fibres and chain fibres. Nuclear bag fibres consist of a central region which has few myofibrils and is thought to have properties that are essentially elastic. This central region is linked in series on either side to contractile regions that have visco-elastic characteristics. Nuclear chain fibres are more uniform in structure and therefore are not expected to show such markedly viscous responses. In most mammals the muscle spindle consists of two or three bag fibres and several chain fibres.

A typical anatomical muscle may have a large number of muscle spindles, in some cases as many as eighty. The properties of the intrafusal fibres are similar to the properties of ordinary extrafusal muscle, except that the intrafusal fibres have their own motor pathways through which the central nervous system can influence the spindle quite independently of the main extrafusal fibres. In the context of neuromuscular control, one of the most important points that has been established about the muscle spindle receptor is that there are several different types of efferent nerve fibre leading to intrafusal fibres. In addition to the well-known gamma efferents, of which there are two types, muscle spindles may also be innervated by collaterals of the alpha motor nerve fibres which supply the extrafusal fibres of the main muscle. These collaterals are known as beta fibres.

The sensory regions of the muscle spindle can therefore be affected both by mechanical changes in the main loadbearing muscle and by activity in the gamma and beta pathways. Functionally, the muscle spindle receptors are thought to be somewhat analogous to strain gauges and provide neural output signals in response to an applied extension of the main muscle or in response to neural activation of the intrafusal fibres of the spindle itself.

There are two types of sensory receptor in the mammalian muscle spindle and these are known as primary and secondary endings. The significant difference between primary and secondary endings is that while the primary endings are found to display a high sensitivity to the velocity of applied stretch, the secondary endings are not so velocity sensitive. The primary output of the muscle spindle is usually transmitted to the spinal cord by a single large axon that winds around the nuclear bag. Mechanical deformation of the sensory endings of this axon leads to the generation of neural impulses at a frequency that is related to the velocity of applied stretch as well as to the static deformation. This rate sensitivity is thought to be of considerable significance in the study of the neuromuscular control system. Secondary endings do not display any marked rate sensitivity.

Tendon organs lie in series with the main load-bearing muscle and the neural ouput from this sensory receptor is directly related to the overall muscle tension. The tendon organ is known to have an inhibitory effect on the the motoneurones of its own load-bearing muscle, whereas the spindle output has a facilitatory effect.

In the 1950's a servomechamism hypothesis of neuromuscular control was formulated by Hammond, Merton et al [3] in which it was suggested that some movements are produced not as a direct result of neural signals from higher centres acting on the alpha motoneurones, but indirectly through the gamma pathway to cause contraction of the intrafusal fibres of the muscle spindle. The spindle output signal is transmitted to the motoneurone pool and thus stimulates the motoneurone which causes contraction of the load-bearing muscle. Although Hammond et al. [3] were the first to employ the servomechanism analogy and carried out experimental investigations that supported this theory, the existence of closed-loop neuromuscular systems had been recognised for many years before that.

The supposed advantage of the indirect route through the gamma efferent pathways is that feedback through the muscle spindle receptor facilitates the maintenance of the conditions appropriate to the fusimotor input in the presence of external disturbances. The system becomes largely independent of changes in load and will be relatively insensitive to muscle fatigue. Later work suggested that the action of the stretch reflex system must be more complicated than the description given by Hammond et al.. It is known, for example, that beta motoneurones innervate both extrafusal muscle and the intrafusal fibres of the muscle spindle. A number of researchers pointed out that the controlled variable cannot be defined in a simple way because of the complex dynamic interactions in the muscle spindle that acts as the comparator within the closed-loop system. Examination of the possible consequences of either position or tension control suggests that neither of these would be suited to the tasks performed by skeletal muscle. More recently, the ideas of pre-calculated feed-forward control have contributed to an understanding of some aspects of the action of the cerebellum in controlling rapid movements in the presence of delays in the feedback pathways.

4 Computational models of the neuromuscular system

A number of mathematical and computer-based models have been proposed for describing control mechanisms in the neuromuscular system and these models have been reviewed recently [4]. A modular approach is generally favoured and has been highly successful in studies of the dynamic properties of individual neural elements of this complex nonlinear dynamic system and of interactions between them [2]. What remains uncertain in this closed-loop system model is the nature of the controlled variable and the origin of properties that produce insensitivity to changes of external load and smooth movements under a very wide range of experimental conditions. These apparently adaptive features are of enormous interest for robotics research. If these and other features of the biological system could be more fully understood and reproduced in the design of control systems for manipulator robots there could be significant benefits.

It appears likely that the key element in providing insight about the complex properties of this biological system is the muscle spindle receptor. This appears to function as a comparator and also as a combined sensor that provides a signal that depends primarily on muscle length and velocity but has dynamic properties that are dependent on the gamma efferent signal from the central nervous system. There is substantial evidence [2, 5] that the fusimotor inputs to the muscle spindle are responsible for changes in the dynamic response of this sensory element in the system. This gives a strong clue about the possible source of some of the complex self-adaptive properties observed in the intact closed-loop system but it is still not known how the overall behaviour of the control system is influenced by these changes at the spindle. The difficulty in approaching this problem lies in the inherent nonlinearities of the system, the complex nature of the interactions at the spinal cord of the various sensory feedback pathways together with spinal pattern generators and local neural feedback pathways. There is also evidence from experiments on intact human subjects that the neuromuscular system changes the loop gain with the load on the system – a feature that is not replicated in many present-day man-made control systems. This has been the subject of a further servomechanism hypothesis by Marsden, Merton et al [6], one of whom (Merton) was closely associated with the earlier hypothesis of Hammond et al [3].

Within the control paths for a single muscle, the alpha motoneurone provides the actuating input that causes the extrafusal muscle fibres to contract. The alpha motoneurone output is determined by a number of feedback loops as well as the inputs from higher levels in the central nervous system. These loops include the feedback from muscle spindle receptors and Golgi tendon organs, feedback from joint receptors and skin receptors and local neural feedback pathways at the spinal cord, involving what are known as Renshaw cells which provide recurrent inhibition. Also, while the activities of the alpha and gamma motoneurones are essentially independent, the activity of the beta motoneurones causes contraction of the intrafusal fibres as well as being related to the activity of the alpha motoneurones, which leads to contraction of the extrafusal fibres. The situation becomes even more complex when the model attempts to represent the coordinated action of a pair of muscles working together on a joint.

It is far from clear how the sensory feedback signals and descending control signals interact. The effect of a single feedback pathway on the overall system behaviour may appear obvious but the complex nonlinear properties of muscle and the threshold phenomena that are know to be present in some of the feedback pathways make any meaningful quantitative analysis or simulation very difficult. Detailed models that attempt to incorporate the anatomical features of the real biological system require knowledge of a large number of parameter values, many of which are difficult, or impossible at present, to estimate from experimental data. Access to intermediate variables for model validation is also extremely difficult in models of this type.

5 An alternative nonlinear model of the neuromuscular system

One common assumption in models of the neuromuscular system is that the efferent signal that actuates the load-bearing muscle is a pulse-frequency modulated signal of strength proportional to the weighted sum of the relevant afferent signals. There is, however, considerable well-established experimental evidence that suggests that the force developed by a muscle is controlled primarily by regulation of the number of motor units that are active. As the required force in the muscle increases, more and more units become active. It is known that in a single unit the pulse frequency increases with the overall load force until a saturation level is reached beyond which no further increase occurs. If recruitment of motor units is the most significant factor in the development of tension in intact muscle this feature must be incorporated into neuromuscular control system models. For a muscle subjected to an imposed constant tension p(t), it is reasonable to assume initially that the applied tension is shared equally by the active units. If the number of active units is n(t), the tension in each single unit of the load-bearing musle must be p(t)/n(t). Traditional models of muscle, make the implicit assumption that all the motor units are stimulated in synchronous fashion, with the complete muscle acting as a single unit. This is clearly incorrect, except in some rather artificial experimental conditions, but it is reasonable to take such a description as a model of a single active motor unit [6]. This provides a basis for quantitative assessment of the ideas of a variable gain system that are implied by the hypothesis of Marsden et al [7].

It is first of all assumed that, as in most previous models, the outputs of the muscle spindle are proportional to the overall length of the main muscle. It is also assumed that some form of composite afferent signal from the muscle spindles and other sensory receptors controls the number of active motor units in the extrafusal load-bearing muscle. Possible models representing the stretch reflex under conditions of controlled tension or controlled length would have the forms shown in Figs. 2 and 3 respectively. If, in the model for isotonic conditions, the fusimotor input is such that at a particular level of applied tension p_1 , the number of active units is n_1 and the muscle length is l_1 , then an increase of applied tension will cause the length to increase and this will give rise to an increase in the number of active units. Any increase in the number of active units will tend to reduce the tension within each active unit because of the action of the divider in the forward path of the nonlinear feedback system. This



Fig. 2: Block diagram showing stretch reflex model for conditions of controlled muscle tension



Fig. 3: Block diagram showing stretch reflex model for conditions of controlled muscle length

shows clearly that a form of negative feedback exists. In Fig. 3 no feedback path exists involving the active skeletal muscle.

It has already been noted that there is evidence that suggests that the stretch reflex system is not a simple position servomechanism. It is possible therefore that muscle spindle receptors have a component of their output that is proportional to muscle tension as well as the component proportional to muscle length. An increase in either length or tension would then tend to increase the number of active units. The overall characteristics of the stretch reflex models for isotonic conditions is altered very little by the introduction of the tension path, but for isometric conditions the presence of such a pathway completely changes the properties of the model and introduces the possibility of an unstable condition that would not otherwise exist. Such instabilities have been observed experimentally under conditions of controlled muscle length.

Tendon organ receptors have been included in this recruitment model. The inclusion of tendon organs provides an additional feedback pathway under isometric conditions. A tendon organ discharge is known to have an inhibitory effect on the motoneurones of its own muscle and must therefore act in opposition to the signals in the muscle spindle pathway. The fusimotor input influences the spindle primary discharge and thus may be expected to influence the number of motor units that are active. For conditions of controlled tension any increase of fusimotor activity will cause the muscle to contract until a new equilibrium condition has been reached. Conversely, any reduction of fusimotor activity will cause the muscle to lengthen.

Although fusimotor inputs are known to play an important part in the control of the neuromuscular system, it is also known that active movements can occur in preparations in which the feedback path from the muscle spindle receptors has been opened. It follows that the system must incorporate a second input that acts directly at the motoneurone pool. This is the feed-forward input, referred to previously, that appears to be used for the initiation of urgent movements for which the time delays inherent in the gamma pathway would introduce an unacceptable lag. Although the two inputs are normally considered as being functionally distinct it is known that they often work together and it has been suggested that the relative amounts of activity in the two pathways may be controlled by the central nervous system to suit the task being performed. It has been suggested that the feedback properties of the neuromuscular system only apply for fusimotor inputs but this is not necessarily true since in many forms of feedback control input signals applied at different points in the loop may have similar effects at the output. If, however, the contraction initiated by the direct pathway is so rapid that the spindle discharge ceases, the system does become transiently open-loop. This is one of the interesting effects of neural signal encoding through pulse frequency modulation. Large negative effects at a sensory receptor may cause the neural discharge to cease complete and the feedback signal to be interrupted. This gives rise to a subtle form of nonlinearity in a closed-loop system that may produce open-loop operation on a transient basis.

The models of the muscle spindle currently available suggest that there is a significant rate sensitivity component on the spindle primary output but no similar rate sensitivity in the secondary output. This could be regarded as a form of rate feedback that would be beneficial in compensating for the effects of pure time delays and other lags. There is a further interesting observation that can be made here about the pulse frequency modulated form of signal in neural pathways. Since the signal strength is proportional to the inverse of the time between adjacent spikes observed in the afferent nerve, the primary discharge from the spindle could be said to show a form of adaptive sampling. When the mechanical input is changing rapidly the instantaneous frequency in the afferent nerve will be high due to the rate sensitivity and the sampling rate will be correspondingly high. This is clearly a desirable situation in a control system.

The changes in the dynamic response of muscle spindles observed when fusimotor inputs are applied suggests that, in addition to producing changes of the equilibrium point, these inputs may provide some form of adaptive compensation of the stretch reflex through parametric changes at the muscle spindles. It is possible, therefore that the static and dynamic fusimotor pathways to the intrafusal fibres provide a means of relatively independent control of the operating point and the dynamic characteristics of the complete closed-loop system.

Discussion

One technological development that makes the study of recruitment phenomena in muscle particularly interesting is the recent development at a number of research centres, including SRI International at Menlo Park, California [8], of actuators that expand and contract in response to electrical stimuli. The development of a robotic actuator made up of many strands of plastic artificial muscle fibre provides interesting possibilities for a wide range of applications. The inclusion of large numbers of sensors that can be embedded in these artificial muscles, like muscle spindles and tendon organs in real muscle, opens up the further possibility of highly redundant control systems that resemble the structure of the neuromuscular system and could begin to offer the high levels of redundancy in terms of sensors, actuation units and signal transmission pathways that seem to offer advantages in the neuromuscular control system. Adaptive systems that vary the gain and other system parameters with load would become relatively easy to implement.

The pulse frequency modulation that is inherent in nervous signal transmission has interesting consequences in terms of control. As already pointed out, extreme conditions can effectively cause feedback pathways to open transiently and switch the control system from closed-loop to an openloop mode. This form of nonlinearity could well be beneficial in some circumstances and has already been introduced in some forms of bang-bang control system, which effectively operate in open-loop mode with maximum control effort applied except when the error becomes small and they change to linear closed-loop operation. As has been noted above, the pulse frequency modulation inherent in nervous signal transmission can also be regarded as a form of adaptive sampling and this has interesting implications in terms of information transmission in man-made systems.

Conclusions

Any engineer must inevitably have respect for the excellence of the design that can be seen in biological systems. Unfortunately much of the fine detail is well beyond our present understanding and cases where the results of these natural design processes may appear to us to be imperfect or puzzling may simply represent our failure to comprehend fully the complexities of the situation.

One of the key elements of engineering design is that, generally, we aim for simplicity and accept added complexity only when we have convincing evidence that the gains exceed the losses. Such losses may be associated with the inevitable extra development or materials cost and the possible added system life-time costs caused by a reduction in reliability and additional maintenance demands. In contrast, complexity does not appear to be expensive in nature, where prototype testing is carried out on such a very large scale that an enormous number of refinements can be tried [9]. While biological systems appear to have a major advantage in terms of the scope for innovative design by evolutionary optimisation, there are important limitations set by the kinds of materials that can be used and the mechanisms that can be employed.

Biological systems are also limited, if one accepts evolutionary principles, in the sense that every new feature must develop from an existing feature and there is no chance of making the sudden giant developmental leaps that are so common in the history of technology. Nevertheless it is clear that the processes of evolution can tell us much that is of value for specific engineering control systems that have functions equivalent to easily identifiable biological systems. The neuromuscular system is an excellent example because of the obvious link to manipulator robots and other types of robotic system. The subtle nonlinear and self-adaptive features of the neuromuscular control system are beginning to be understood and better knowledge of the biological control system could well provide further ideas to explore in the development of improved forms of man-made system. Improved understanding of the normal operation of the neuromuscular control system could also assist rehabilitation engineers in developing ways of assisting, through controlled functional electrical stimulation techniques, those who have lost the functions of normal limb control.

References

- von Gierke H. E., Keidel W. D., Oestreicher H. L. (Editors): *Principles and Practice of Bionics*. AGARD Conference Proceedings No. 54. Slough (England): Technivision Services, 1970.
- [2] Rosenberg J. R., Murray-Smith D. J., Rigas A.: "An introduction to the application of system identification techniques to elements of the neuromuscular system." *Trans. Institute Measurement and Control*, (U.K.), Vol. 4 (1982), p. 187–201.
- [3] Hammond P. H., Merton P. A., Sutton G. G.: "Nervous gradation of muscular contraction." *Brit. Med. Bulletin*, Vol. **12** (1956), p. 214–218.
- [4] He J., Maltenfort M. G., Wang Q., Hamm T. M.: "Modelling neural control." *IEEE Control Systems Magazine*, Vol. 21 (2001), p 55-69.
- [5] Halliday D., Murray-Smith D. J., Rosenberg J. R., Rigas A.: "A frequency-domain identification approach to the study of neuromuscular systems-a combined experimental and modelling study." *Trans. Institute Measurement and Control*, U.K., Vol. 14 (1992), p. 79–90.
- [6] Murray-Smith D. J.: An Application of Modelling Techniques to the Neuromuscular Control System. PhD Thesis, University of Glasgow, Chapter 6, 1970.
- [7] Marsden C. D., Merton P. A., Morton H. B.: "Servo action in human voluntary movement." *Nature*, Vol. 238 (1972), p. 140–143.
- [8] Anon: "Artificial muscles: Expansive thinking." *The Economist*, February 5th 2000.
- [9] French M.: *Invention and Evolution*. 2th Edition. Cambridge, Cambridge University Press, 1994.

David J. Murray-Smith

Centre for Systems and Control

Department of Electronics and Electrical Engineering Rankine Building The University of Glasgow Glasgow G12 8LT, Scotland, U.K.