UNIVERSITY OF COLORADO, BOULDER

Department of Computer Science

Gerhard Fischer

Human-Computer Communication and Knowledge-based Systems

Appeared in:
Otway, H.J. and Peltu, M. (eds.): "The Managerial Challenge of New Office Technology". A publication from the INSIS Programme of the Commission of the European Communities. Butterworths, London.

1984, pp 54-79

ECOT 7-7 Engineering Center • Campus Box 430 • Boulder, Colorado 80309 • (303) 492-7514

3

Human–Computer Communication and Knowledge-based Systems

Gerhard Fischer

Introduction

The microelectronics revolution of the 1970s made computer systems cheaper and more compact, with a greatly increased range of capabilities. Computing moved directly into the workplace to the fingertips of everyone doing office work. Chapter 2 described the rich variety of office tasks that can be assisted by computer-based computing power. Much of this power is wasted, however, if users have difficulty in understanding and using the full potential of their new systems. Too much attention in the past has been given to technical aspects which have provided inadequate solutions to real world problems, imposed unnecessary constraints on users and been too rigid to respond to changing needs.

This chapter examines how to improve the interaction between the user and the system in order to increase effectiveness. It explains why human–computer communication is a crucial determinant of the usefulness of systems and offers guidelines on the ergonomic criteria that should be used when developing and evaluating systems. It emphasizes the importance of software in managing the dialogue between users and computers, particularly in relation to workstations used by managers and professionals. The chapter also explains why, in order to develop systems which fit naturally into office work environments, more 'intelligent' software is needed which has knowledge about the user, the tasks being carried out and the nature of the communications process.

The importance of human–computer communication

The management challenge of introducing computer-based systems is not primarily a technical one. It involves handling a variety of complex human and organizational changes, as well as technical innovations. This means that the design, development and evaluation of new information systems should start with an understanding of the overall social and technical *(socio-technical)* environment in which any particular new technology is embedded. *Ergonomics* and *human factors engineering* are disciplines which pursue this aim of having people and technology working in harmony to meet the desired performance. This *user-centred* approach starts on the 'outside', examining human psychological and behavioural needs, then moves, inwards, through the work tasks carried out to the specific technical details.

A crucial determinant of the effectiveness with which computers are applied to assist people at work is the nature of the *human–computer communication*; that is, the interaction between the user and the system, also known as *Man Machine Interaction (MMI)* or the *user interface*. Managers should understand the key issues in human–computer communication. This will help them to steer the design and implementation of new systems in order to create the appropriate socio-technical environment for the groups under their responsibility. It will also enable them to differentiate between spurious advertising slogans and really important features when evaluating systems. In the past, methodologies for creating and assessing computers were *computer centred*, which is why so many failed to match their actual operating environments (*see* Chapter 6). They started by considering what the hardware and software could do, then built the final system around these computing capabilities. Managers have a responsibility to avoid this attitude in the future.

A systematic approach to human–computer communication should be an important mechanism in implementing strategic plans in order to match successfully the capabilities of new office technology with organizational and personal goals. Detailed prescriptions or check-lists cannot be provided to cover all aspects of human–computer communication because so much is dependent on human *cognitive* abilities – how people behave, think and perceive the world. Such subjective factors are not amenable to be being measured and predicted with the same precision that is possible with elements in the physical environment. This chapter does provide, however, principles of good human–computer communication which can be used as the basis for judging the inevitable trade offs that have to be made when weighing up the advantages and disadvantages of different systems.

Costs of ignoring ergonomics

Concern about the ergonomics of computer-based systems first came to prominence in relation to the possible health hazards for operators using VDTs for clerical and typing tasks. Managers became aware that failure to examine ergonomics could lead to staff anxieties and resistance to technology. Research into hardware ergonomics and the physical work environments (chairs, desks, lighting, ventilation, etc.) clearly showed that poor ergonomics could lead to inefficiencies at work as well as to some physical discomfort and mental stress for users.

If the human–computer communication is too complex, the user will be unable to understand and exploit many of the facilities available. It has been found that *less than 40 per cent* of the potential range of functions is ever used on many systems. If the user interface is difficult to grasp and remember, unforgiving to even the smallest error and generally unfriendly, users will require a great deal of training and the system will be error prone and inefficient in action.

Ergonomics views all elements of a system as a whole. It requires that hardware, software, an individual's psychological needs, group behaviour and dynamic social interactions are considered in a systematic and integrated fashion. By putting the user at the centre of the design, ergonomic and socio-technical methods identify what is needed before looking at how it is done. This leads to important strategic guidelines, independent of any particular technology. For example, it recognizes that individuals and organizations evolve over a period of time. The system should therefore be capable of adapting to meet different requirements, such as allowing a person to move from being a novice to an experienced user within a smooth, consistent framework of human–computer communication.

An ergonomically designed system should enrich jobs and reduce stress (*see* Chapter 5). If the system is forgiving towards user mistakes, say allowing erroneous actions to be corrected through an 'undo' command, users will feel more relaxed and willing to investigate a wider range of applications of the technology. On the other hand, a technology centred design can lead to computers becoming a straight jacket, determining what can and cannot be done. For example, if a workstation handles only text, there is a tendency for the user to ignore other forms of information presentation. This has led to a diminishing use of graphics in papers and reports prepared on word processors. The word processors may have been more productive in terms of the time taken to produce a document. The end-result, however, may be less useful because of the omission of graphics, which can be such an effective means of presenting information.

In order to avoid the costs and problems that can occur when socio-technical and ergonomic needs are ignored, it is important to incorporate the necessary evaluation criteria from the start of a project. Ergonomics research should indicate wrong developments at an early stage. After installation, continuous evaluations should assist the system to evolve in tune with social, human and organizational needs.

The nature of human–computer communication

A user-centred approach to computer systems requires an understanding of:

(1) The skills and knowledge of different types of user.
(2) The structure of tasks to be performed, for example, whether a task can be defined by a clear, predictable specification or is ill structured, with many ambiguities and unexpected occurrences.
(3) The technology involved, say whether the workstation has a basic keyboard and single frame screen or a mouse and keyboard with a multiwindow screen and icons (*see* Chapter 2).

Until the 1970s, the relationship between the user and the computer was so remote that it could be compared more to correspondence by letter than to a conversation. Today, users and computers usually interact directly, in a similar fashion to a conversation. The styles of some interactions are restricted and allow tasks to be accomplished using only a narrow range of techniques. The users in such systems are regarded essentially as operators, be they typists operating a word processor or children manipulating the control stick of a video game. A new era of human–computer communication began when microelectronics decreased the cost and increased the availability and capabilities of hardware. This made it feasible to use computational resources not only to provide particular functions, but also to assist in making those functions usable.

There is a growing understanding that the cognitive limitations of the user are as important to communication with machines as the technology of the machine itself. The increasing richness and complexities of possible communication means that systems designers are often faced with having to resolve conflicting requirements, such as:

(1) Balancing what is best in terms of a person's cognitive thought against what is most efficient for the computer.
(2) Providing systems that must be easy for most people to use, but also must have sufficient power to allow the skilled user to exploit the system for a variety of different and more complex purposes.
(3) The necessity to remain compatible with existing systems while also exploiting the power of new systems and techniques.
(4) Being easy for beginners to use as opposed to the needs of experienced users who require less hand-holding.
(5) Having tight integration between different subsystems but still allowing systems to be composed of independent modules that can be flexibly interlinked and rearranged.

Human–computer communication can become a bottleneck that restricts the growth of successful uses of new office technology and limits the extent to which new information processing and communications technologies can be integrated into our working and living environments. Many techniques are being developed to assist communication.

Advantages of knowledge-based systems

Computer techniques have traditionally been constructed from the logical information handling capabilities of hardware, which are most suited to dealing with factual data and other information and calculations amenable to digital encoding. *Artificial intelligence* (AI) methods, on the other hand, start by using human behaviour as the model of how computers should act. People acquire knowledge through experience and learning and then apply that knowledge to solving problems, communicating, making decisions and acting. *Knowledge-based systems*, also known as *intelligent knowledge-based systems* (IKBS), aim to emulate these human characteristics. Knowledge-based systems have two main ingredients: the store of knowledge and a means of processing that knowledge using programs and rules based on how people reason, deduce and infer. This requires techniques for:

(1) *Knowledge acquisition*. How knowledge can be acquired

most effectively from human experts and data gathered by instruments. This may involve a *knowledge engineer* to interpret the expert's knowledge to the computer. The expert could, however, create and manipulate the knowledge base directly.

(2) *Knowledge representation*. How to represent knowledge in a form that can drive the computer but which is still understandable by users. Traditional software code is generally comprehensible only to specialists who understand programming rather than those who know about the subject (*domain*) to which the programs are being applied.

(3) *Knowledge utilization*. How the knowledge base can be 'browsed through' and relevant knowledge found. The extent to which the system uses its knowledge to assist the user in finding what is needed must also be determined.

One particular IKBS development is the expert system which attempts to match the performance of human experts in a specialized domain. Like a human specialist, it should also be able to communicate, to explain and to give assistance, and so on. A knowledge-based system without good human–computer communication is like a human who knows everything but cannot talk about it. Good human–computer communication without knowledge behind it, however, is like a person who talks all the time but does not know anything.

Knowledge-based systems offer the most promising approach to improving human–computer communication to the level of effectiveness expected when people communicate with each other. They recognize that people and computers have different attributes and so form a cooperative partnership.

Towards better human–computer communication

The following are the key areas where human–computer communication has failed to match the effectiveness of human interaction:

(1) People are able to understand each other and make reasoned judgements although all the elements in the communication have not been made explicit. In all but the simplest exchange, it is likely that a substantial portion of the communicated message is not made explicit and that the information given is incomplete. Nevertheless, people are able to deduce or supply additional information and correct mistakes through their knowledge of the context of the

communication, past experience and so on.

(2) When faced with a problem presented in broad general terms with many details missing, people are still able to apply their reasoning and problem-solving power. People can make sense out of unexpected situations whereas computers have to be preprogrammed to anticipate all eventualities.

(3) If there is a misunderstanding, people are able to articulate the reasons for it and realize the limitations of their own and their partner's knowledge.

(4) People can provide explanations to others of how they reached a conclusion or why they behaved in a particular way.

(5) People can solve problems by taking imaginative leaps, for example by conceiving of an analogous situation of similar characteristics with which they are more fmiliar.

What makes human communication so successful is that it takes place between individuals who have their own knowledge bases to draw on. A manager expects that the people around him or her know what they are doing in their jobs. A secretary, for example, should know whether or not to put a call through to the manager depending on what the manager is doing, who is calling and the importance of the topic the caller wants to discuss. An office worker who has no understanding of the business being carried out is of little support to a manager because of the effort required by

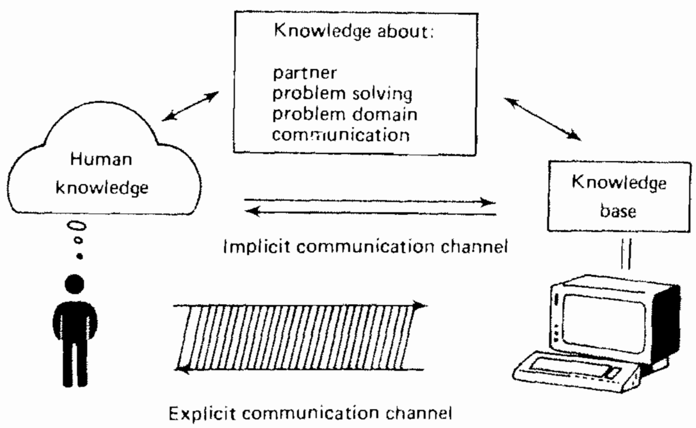

Figure 3.1 Use of a knowledge-based system to assist human–computer communication

the manager when delegating a task. In a similar way, computer aids for managers should also have some knowledge and understanding of what they are doing.

Figure 3.1 summarizes the most promising approach to meet fully the design criteria provided in the remainder of this chapter, although some goals can be met with other techniques. Compared to previous computing methods, the advantages illustrated in this diagram are:

(1) *The explicit communication channel is widened.* This covers the direct interactions with a workstation. Through the use of colour, graphics, icons, multiwindow displays and so on, the workstation has a much wider range of options which can be used to manipulate, explore and analyse information and knowledge held in the computer.

(2) *An implicit communication channel* is opened. This allows for the crucial ability of partners in a communication to understand each other without having to have every last detail spelt out explicitly.

Human–computer communication can therefore become an exchange between two knowledgeable entities, one human and one computerized, rather than between an intelligent person and a dumb machine.

Guidelines on user requirements

A computer-based system cannot be evaluated in isolation from the person using it or from the tasks that it is expected to support. A system that performs well when being used to carry out a narrow range of relatively well-structured tasks is unlikely to be suitable for managers whose workload is varied, unpredictable and unstructured.

Ergonomics research therefore examines the following, in addition to particular technologies.

(1) *The behaviour and perceptions of the user.* In comparison to computer information processing systems, people have both strengths and weaknesses when they deal with information. For example, people have limited short-term memories and a tendency to make errors for a variety of reasons. On the other hand, people have powerful visual systems which give great scope for improving the way information is displayed on workstation screens and for human–computer com-

munication. Some people suffer from fears and anxieties about the technology, which must also be taken into account.

(2) *The nature of the tasks being performed.* The job roles carried out by people in offices consist of a number of tasks. Clerical and secretarial work tend to have a narrow range of tasks which occur in a more predictable and structured way than those of managerial and supervisory work. Systems for office work must be flexible enough to be tailored to meet unexpected situations. They must also be responsive to changing patterns of work and the evolving needs of each user.

Managers are sometimes referred to as *casual* or *discretionary* users because they do not have to operate the system continuously and they have a degree of freedom in choosing when to use computing aids. The work of managers and professionals, which form the main focus of this book, can be differentiated from other office roles by the following characteristics:

(1) *Problems dealt with are often 'fuzzy'.* The precise specification of the problems and challenges to be resolved can be difficult to define in advance. Solutions may be found by moving from a partial solution via a learning from experience to an evolving understanding of the problem.

(2) *Formal analyses are insufficient.* Problems to be faced usually cannot be adequately understood in advance. Techniques of formal analyses of work tasks, such as those favoured in operations research, are inadequate to define job routines and communications channels in sufficient detail to be translated into software.

(3) *Importance of innovation.* The unpredictable and varied nature of the work means that innovative new solutions often have to be generated which depart from systems based on analyses of past behaviour.

(4) *Complex decision making.* Responsibilities for making decision are often difficult to delegate because they cannot be described well enough for an assistant to do them.

How to evaluate the usability of office technology

The human–computer communication ability of a system cannot be defined by a simple measure of 'goodness' or 'badness'. The aspects discussed above indicate the many factors and interactions involved. The following dimensions should be considered when assessing systems.

(1) *Usefulness*. The system should be as supportive as possible in aiding the user to meet her or his goals; knowledge is therefore needed about these goals.

(2) *Functionality*. The range of tasks or functions that a system can perform is known as its functionality. All required functions should be smoothly integrated within a consistent framework.

(3) *Uniformity of interaction*. The style and format of human–computer communication should be consistent and uniform between different tasks and subsystems. This is particularly important for a casual user who does not have the time to learn and understand new techniques.

(4) *Flexibility and adaptability*. The system must not only be able to meet immediate short-term objectives but should be capable of being enhanced, adapted and extended in the future in an evolutionary way, building on skills learnt and systems used in the shorter term.

(5) *Learning and training*. The time needed to learn how to use functions should be short, particularly for busy users who have little time to spare. The aim should be 'no threshold and no ceiling': there should not be a threshold step too high and too complex for a novice but there should also be no ceiling on the degree of complexity to which the system rises. This can be achieved by constructing systems that grow with the experience of the user.

(6) *Error handling*. When an error occurs, the system should generate a meaningful response relating to the task, not to the internal state of the machine. Advice should be provided, clearly and unambiguously, so that the user is able to take corrective action to overcome the problem.

(7) *Robustness*. Errors should not lead to a total breakdown of the system. Users should also be offered the chance to retract (*undo*) previous actions without losing all the relevant information or causing a major disruption in processing. This reduces stress on the user and enhances her or his willingness to innovate and explore new aspects of the system.

(8) *Speed of response*. The computer should respond to the user with sufficient speed so that the user does not get anxious wondering what has happened, but not so quickly that the user is put under unnecessary pressure. If there is likely to be a delay in response because the computer needs time to perform a complex calculation or search through a large

database, the user should be given an indication that the system is still active and, if possible, an idea of when the response will be provided.

(9) *Helpfulness*. The system should provide guidance and advice if the user is unsure what to do next. Such help should be provided in a language that is understandable by the user, and it should be unambiguous and relevant to the situation the user is in at the time.

(10) *Time to implement*. A system should not only provide the appropriate functions and quality but should minimize the amount of time and personnel resources needed to apply the technology efficiently and effectively.

(11) *Quality of service*. The quality of the final service or product, such as a customer enquiry service or printed document, should be at least as good as anything that could be provided by non-computer means.

(12) *Acceptability*. All users should feel the system genuinely enhances their jobs and work environments.

(13) *Group support*. Systems should provide adequate assistance to the sharing of information amongst members of a group, in addition to helping individuals.

(14) *Self-explanatory power*. A knowledge-based system should have the ability to understand what it is doing and to explain to the user how it made inferences which lead it to recommend a particular conclusion.

(15) *Conviviality*. The user must be in control of the system and be able to modify it as required to meet activities and situations not anticipated in the programs.

(16) *Symbiosis*. People and computers should be able to unite in a harmonious, symbiotic relationship where the computer augments human activities rather than replaces people.

Eight pillars of ergonomic wisdom

The following summarize the major principles that have emerged from ergonomic research into the behaviour of users and the tasks they perform. They provide a basis for developing and assessing adequate human–computer communication.

(1) *The limiting resource in human processing of information is human attention and comprehension, not the quantity of information available*. Modern information and communication technologies have dramatically increased the amount of information available to individuals. An important function

in human–computer communication is to allow for the selection of the information you actually want and presentation of it in the most appropriate way.

(2) *In complex situations, the search for an optimal 'rational' solution is a waste of time.* There are limits to the extent to which people can apply rational analyses and judgements to solving complex, unpredictable problems. It is insufficient to ask people to 'Think more clearly' without providing new tools, such as knowledge-based systems, which help extend the boundaries of human rationality. The aim is to achieve the most satisfactory solutions given current knowledge, accepting that 'better' solutions will emerge as the result of experience and enhanced knowledge and understanding.

(3) *The nature of human memory mechanisms are important design considerations.* The limitations and structure of human memory must be taken into account in designing human–computer communication. People have relatively limited short-term memories. Dialogues should, therefore, be constructed which do not expect the user to remember everything and which reinforce, prompt and remind the user of necessary information in a supportive but unobtrusive manner. The way people *recognize* information visually is different to how they *recall* other information. The different recognition and recall memory structures are relevant to judging the advantages and limitations of different user interactions, such as comparing the use of a function key to initiate an operation compared to a menu-based interface (menus are examined later in this chapter).

(4) *The efficient visual processing capabilities of people must be utilized fully.* Traditional displays used with screen-based workstations have been one dimensional, with a single frame on the screen usually filled with lines of text. New technologies have opened ways to exploiting human visual perception more fully, say through the use of multiwindow displays, colour, graphics, icons and mice (*see* Chapter 2).

(5) *The structure of the computer system must be understandable by people using it rather than requiring the user to learn by rote the functions that can be performed.* An adequate understanding of how a system works gives users the knowledge and confidence to explore the full potential of a system, which can have a vast range of possible options. Learning by rote may train the user to operate a limited number of functions but makes it difficult for the user to cope

with unexpected occurrences and inhibits their exploitation of the full potential of the system.

(6) *There is no such thing as 'the' user of a system: there are many different kinds of user and the requirements of an individual user grows with experience.* Computer systems built to a static model of the nature of *the* user of the system are too rigid and limited to meet the demands of a rapidly growing and diverse user community or the evolving needs of each user.

(7) *The 'intelligence' of a complex computer aid must contribute to its ease of use.* Truly intelligent and knowledgeable human communicators, such as good teachers, use a substantial part of their knowledge to explain their expertise to others. In the same way, the 'intelligence' of a computer should be applied to provide effective communication.

(8) *The user interface in a computer system is more than just an additional component: it is an integral and important part of the whole system.* Human–computer communication must be considered at the very earliest stage of the design process so it can be adequately integrated with all other elements.

Ergonomic criteria for human–computer communication

The crucial unresolved issue for human–computer communication in the office of the future is software ergonomics. Hardware and software are closely related and there are important questions in hardware ergonomics still to be investigated. Hardware, however, is no longer the main limiting factor in computing developments, as it once was. Progress in microelectronics and other 'hard' technologies have opened huge spaces within which software can manouevre and which are still under-exploited.

Rapid hardware developments have frequently overtaken ergonomic research. The technology being examined may become obsolescent before the research reaches a conclusion or capabilities previously available only on a few costly systems suddenly become common at a price within the reach of all users. For example, detailed research into the optimal design for keypunches used with punched card equipment is of little value today because keypunches are dying out. At one time, much effort was spent investigating the importance of having terminals and printers which provide both upper- and lower-case characters. Now, most devices offer small and capital letters at a reasonable cost.

Computers now have enough computational power in a compact and low cost form to overcome previous major hardware restrictions such as the lack of graphics, colour and multiwindow displays. The prime question is how to develop software which makes use of this extra power to meet actual user requirements.

Workstation hardware requirements

Research and development in hardware ergonomics and the physical work environment have made substantial progress and is backed by checklists which can be used with a reasonable degree of confidence[1]. This is possible because the human physiological and perceptual systems involved in operating equipment are fairly well understood and more readily amenable to measurement than the subjective factors involved in software ergonomics. These guidelines generally focus on VDT designs for secretarial and clerical tasks and emphasize characteristics such as:

(1) Keyboard detached from screen for optimal positioning, with the display capable of being rotated and tilted.

(2) Avoidance of glare and reflection from the screen.

(3) A flat keyboard (no more than a 15 degree slope) to avoid excessive loads on the hand and arm, with properly shaped keytops labelled with legends that are readable and understandable.

(4) Presentation of stable, legible characters on the screen. The most common (and preferable) method of creating images on a screen, also used for domestic televisions, is *raster scan* generation. The cathode ray tube sweeps an electron beam over the phosphor coated inner surface of the screen, illuminating tiny dots on the screen to form the required characters or other shapes. The image, however, quickly fades and has to be continually refreshed. If the *refresh rate is too low, characters are likely to flicker; a minimum of 50 Hertz (preferably 60 Hertz) is needed for negative presentation* (light characters on a dark background) or 80 Hertz for *positive presentation* (black on white) of flicker-free character displays.

These basic requirements also apply to the personal workstations for use by managers and professionals which will be the dominant feature of office systems in the future, providing considerable local computing power as well as being linked to local area networks and external communication links. It is therefore of great importance that workstations adequately integrate the broad range of tasks required. The following characteristics for workstation ergonomics should therefore be considered, in addition to the general VDT guidelines already mentioned.

(1) *Screens and output*
  (a) The display should be large enough to show a full page of A4 text and make effective use of a multiwindow capability (*see* Chapter 2).
  (b) A graphical capability is vital for many applications and a desirable option for most tasks. Raster display techniques are preferable to *character* or *vector generation*, which produces images as sets of lines rather than groups of dots.
  (c) Colour is also a desirable option for many tasks and a key requirement for specialized activities where it is necessary to differentiate between many objects or types of information.
  (d) It should be possible to identify and select objects on the screen through a pointing device, such as a mouse, a light pen or touch-sensitive screen (*see* Chapter 2).
  (e) Screens and keyboards should have similar functional capabilities so that What you See Is What You Get (WYSIWYG, pronounced 'whizzy-wig'). For example, the screen should have *proportional spacing* where each character has a different width (such as 'm' being wider than an 'i') as is common on many printers. If the screen has a graphics capability, then the printer should also be able to produce similar images; laser printers, for instance, can satisfy this need.

(2) *Input devices*
  (a) Facilities should be provided to handle text (most commonly a keyboard); pointing devices; and choice devices (for example, special buttons or keys which initiate complete functions). Voice and handwriting input will enter commercial applications when they achieve appropriate performance levels; initially they are being used in specialized tasks, for example where a limited speech vocabulary is sufficient or where hand-printed, block characters are written on special forms.
  (b) The way a variety of input methods are used together must be carefully planned and integrated; for example,

there should be a smooth transition when switching between a mouse and keyboard.

(3) *Growth potential*. New technological developments are continuously offering new functions and capabilities. The workstation design must therefore be sufficiently flexible to incorporate enhancements as they become available and are required by the user. The design strategy must also provide a sufficiently coherent and consistent framework so that new facilities can be integrated smoothly with existing capabilities. This can be achieved by following the design criteria for human–computer communication recommended in this chapter.

Software: the key to effective computing

Software mediates between the sophistication of computer-assisted office systems and the human ability to interact with them in a natural, productive manner. It is the bridge between what the technology can do and how the user expects the system to behave. It is, therefore, partly concerned with human behaviour and partly with the technicalities of the system. The sturdiness and usefulness of this bridge is the concern of *software ergonomics*, although this also encompasses psychological and other factors[2]. The subject became recognized as a crucial aspect of computing when word and text processing systems began to be used widely in offices by people who had no previous computing experience. Before that, computer users generally had specialist knowledge and an interest in computers which made them willing to overcome awkward and inefficient interfaces.

Software can be regarded as the implementation of a model of how a system is expected to behave (*see* Chapter 6). In terms of software ergonomics, the challenge is to bring together three models which play a major role in human–computer communication each providing a different perspective on the interface.

(1) *The model of the systems designer*. The person or group designing a system has a concept of the purpose of the system, the kinds of users who will work with it, the tasks and performance to be achieved, and the most appropriate forms of interaction. This, perhaps imperfect, model has a crucial influence on the nature of the systems and in the past has been dominated by the views of technical specialists who often failed to consider the users' needs in full. There is still, however, much that is unknown or unpredictable in relation

to human behaviour, of how an office operates in information processing terms, and of the underlying structures of tasks being carried out. Steps towards better understanding are emerging and it is important that designers are kept on the right track through the involvement of users in the design and assessment of systems, even if most users still have difficulty in articulating the details of their own working methods, let along specifying the future system they would like to have.

(2) *The user's view of the system.* Users often find systems designed by the technically oriented specialist impenetrable. They may learn how to use some of the facilities by memorizing the operating instructions but they have little appreciation of how information is organized within the system, the processing mechanisms, and so on. In some systems this can lead to over 60 percent of the computing power being wasted. This waste may grow as systems become more complex unless a model of the system is presented which clearly explains to the user the limitations, as well as the scope of the technology. For example, if the human–computer dialogue can take place using a limited form of natural language vocabulary and grammar, it should be made clear to users that they still cannot communicate with the computer with the same freedom as they can chat to colleagues.

(3) *The system's expectation of the user.* As has already been asserted, there is no such thing as *the* user of a system. There are many types of user and individual users change over a period of time. Systems should therefore have the ability to be tailored to particular *profiles* that define the tasks to be carried out by a particular user. The more knowledge the system has about the user and his or her tasks, goals and understanding, the better should be its ability to adapt its behaviour to match the varied requirements of different user needs.

Convivial human–computer partnerships

A key aim of software ergonomics is to develop systems that are symbiotic and convivial. Symbiotic systems combine human skills and computing power to carry out a task more effectively than it could be done by the human or computer alone. Convivial systems give users the power to adapt the system. Knowledge-based systems will help to profile systems in a deeper way than by simply specifying and adjusting the values of particular *parameters* (variable factors, such as the size of a file). With knowledge-based convivial systems, users will be able to create new programs and adapt designs without needing specialist training.

Symbiotic systems exist in many different areas. For example, computerized axial tomography is a computer-controlled scan-ningtechnique which can present three-dimensional X-ray images of the body. It operates as a partnership between the computer, which performs an immense amount of mathematical calculations, and the doctor, whose experience and visual perception enable the information to be interpreted in ways which discriminate between subtle differences in aspects of the image. In financial forecasting, a knowledge-based system can perform calculations, advise the user of errors, explain the origins of particular data and generate proposals according to constraints formulated by the user in an 'intelligent' dialogue, as would take place in discussions with knowledgeable colleagues.

Balancing on the trade-off tightrope

While there are a number of clear principles and trends for guiding software ergonomic design, choosing the best solution in a particular case is often like walking a tightrope. As the system progresses, the balance needs to be restored as different forces come into play.

Even when there seems to be a clear-cut advantage in a particular technique, the details of the implementation must be carefully examined. For example, in early text processing systems the following instruction may have been needed to replace a word four lines down and six characters to the left of the current position of the text pointer with the five-letter word 'green':

(4n6fsi"green")

Such a command has all the characteristics of poor ergonomics: it is complex, hard to remember, liable to cause errors and is difficult to relate to what is actually happening.

A more direct and preferable form of performing this type of editing is to use human visual abilities to move the cursor directly to the point where text is to be deleted, changed or inserted. There are, however, many ways of controlling the cursor. On some personal computers, two keys need to be pressed for each movement, such as the control key and the I, J, M or K key, which bear no obvious relation to particular movements. Generally, however, there are special control keys with arrows to indicate the direction moved by the cursor when that key is pressed on its own. From an ergonomic point of view, the cluster of keys should be arranged so that they are easy to access and their positioning should indicate their function (top key for upward movements, left-most key for left movement, etc.)

In other areas, the pros and cons can be more finely balanced. For instance, with the principle of WYSIWYG (what you see is what you get), the user does not have to keep transforming what is on the screen to the form it will take in printed hard copy which is obviously a benefit. On the other hand, the user may have to do more work when directly manipulating a format than if separate formatting software was being used to translate the screen information to printed form. It may also be too costly or impractical to have a printer that gives the same functions as the screen, and *vice versa*, although this drawback is likely to diminish in time.

Another example of a design conflict arises from the need to have systems that allow the manipulation of various information objects, such as text, graphics, programs, mail messages, and so on. There should be a uniformity and consistency in interactions handlingdifferent functions but there should also be provision to manipulate specific objects in particular ways, which are not needed for other types.

Help systems

The extent to which a system helps the user is an indication of the degree to which the computer is being used to filter, summarize and diffuse information selectively rather than merely adding more data to a world already overloaded with information.

When a person looks for assistance from a colleague or from a computer, it is to answer questions like: 'How can I do X?', 'What happens if . . . .?', 'Why did Y occur?', or 'Can I undo the effects of Z?'. On many systems, the help facility is initiated by a special HELP key or typing in the command HELP. Unfortunately, the computer's response is often to present a lot of information, much of which is irrelevant to the question in the user's mind. Finding the answer may be time consuming, if it is there at all. It is as if a colleague responded to a question such as 'Who is responsible for signing this requisition form?' by presenting you with a manual containing the full organization chart.

To be able to answer requests for help in an appropriate way, the system should have some knowledge about the dynamic context in which help is requested. The user should be able to describe to the computer what is wanted and the system should also offer an explanation based on an informed model of the user. The help facility should not be intrusive, giving advice when the user does not want it and getting in the way of other work being carried out. Given that users can be unaware of about 60 percent of the functions of some systems, the help service may need to be *active*, volunteering advice that is relevant and unintrusive. Most help systems, however, have been *passive*, activated only on the user's initiative. The development of suitable active aids is a complex task which needs to be done well or not at all.

Menus: making the right decisions

One of the major lessons from human–computer communication research is that there are no optimal solutions, only trade offs. The variety of interrelated issues that affect the user and the task being formed means that solutions need to be approached systematically but with sensitivity to the particular circumstances in question. Prescriptive formulae applicable to all systems are not possible or desirable. The nature of the trade offs that need to be made can be illustrated by examining the pros and cons of menu-based interactions.

A menu is a list of alternatives which appears on the screen, similar to the items on a restaurant menu. This reminds the user of the options available at a particular stage, from which one or more can be selected. Selection may be done by keying in the number associated with the desired item(s) or pointing to it with a mouse or using a touch-sensitive screen. Ideally, the menu system, like a help service, should be aware of the dynamically changing context in which it is being used. This means that it can intelligently select information valid in that situation to limit the possibility of the user making an error by selecting an inappropriate option. The menu can also be employed as a means of reminding the user of options that can assist exploring unknown parts of the system.

Alternatives to menus include command and natural languages. Command languages make available a number of commands and instructions, usually within a very restricted format. Natural languages allow theuser to interact in a way which is similar to human languages, although usually with a limited vocabulary applicable only to particular tasks. Where a small set of alternatives exist, menus can be an effective alternative or supplement to command languages. Where it is necessary to construct a complex phrase to make the user's needs clear, say, when making an enquiry of a database, command and natural languages are generally better.

The selection of a suitable system is often a matter of 'horses for courses' – different users have different requirements. Provided the words used in the display are meaningful to the user, menus provide particularly valuable assistance to new and casual users because they do not rely on memorizing commands, limit the amount of typing (and so avoid and minimize the likelihood of errors), and constantly remind the user of the options available. On the other hand, more experienced users often complain about menus although they can be an effective means of reducing the complexity of sophisticated systems. If the user knows what to do, it is a waste of time to request the menu, wait for it to appear, and then to look through it for the appropriate option. In these cases, the user would prefer a command or natural language to express his or her requirement directly.

Once it is decided that a menu approach is suitable, there are still many open questions in relation to the technology with which it is implemented, as well as the groups of users and the sets of tasks for which it is applied. Managers must avoid regarding this openness as a reason for dodging the issues and attempting to find simplistic answers. Careful thought should be given to weighing up factors relevant to particular circumstances, including the following aspects:

(1) *Technology used*. Managers and users often have to make the most of systems already available rather than obtaining a more suitable technology. To be really effective and efficient, with the minimum chances of error, menus should operate with multiwindow systems and pointing devices. A traditional single-window screen has all its other information wiped out when the menu is presented and relies totally on keyed responses. Nevertheless, if the user has no other feasible option, the content and structure of menus should be carefully considered to make the best use of a traditional keyboard and screen VDT.

(2) *Number and order of menu items*. Too many items are confusing, too few can limit the range of choice unnecessarily. Elements in the menu can be ordered alphabetically, by functions or randomly. The optimum number and order depend on the tasks involved and the experience of the user;

these may need to be adapted for different users and applications.

(3) *Hierarchical menus*. If it is necessary to trace a path through a number of menus to get to the desired point, the user should be given the opportunity to go directly to the place needed, otherwise he or she has to re-tread a tedious path through the networks of menus. Provisions should also be made to limit the chance of the user getting lost in the hierarchy and providing help if the user is in doubt as to what to do next.

(4) *Presentation and positioning*. In multiwindow systems, the menu may be ever-present or only pop-up when required. Decisions also need to be taken on how close the menu should be to the window(s) of information to which it relates, and whether it should always be in the same place.

(5) *Icons*. When it is appropriate to use icons in menus rather than verbal descriptions.

(6) *Default values*. A decision is required as to what action the system takes if the user fails to specify one or more of the number of items expected to be selected on each occasion the menu is activated.

(7) *Conviviality*. The extent to which users are allowed to modify the system, for example, by extending the items in a given menu, choosing default values or deciding where to position the menu. Too much choice can be confusing, too little can mean the system is too rigid for necessary local adaptations.

Future improvements to user performance

This chapter has shown that user performance can be improved not just by providing new interfaces to the system but also through a combination of increasing the understanding of the user and the knowledge built into the system. There are, however, important issues which need to be investigated in future research into human–computer communication:

(1) *User training*. We have started to live in a world where people are likely to learn new skills several times in their working lives. One of the aims of software ergonomics is to cut the training load by creating systems which are easy to use, helpful, and forgiving of mistakes. Short periods of hands-on training with well designed systems should be a

cost-effective alternative to investments in formal training courses or trying to invent mythical 'ultimate' interfaces which require the user to act totally naturally (*see* Chapter 9).

(2) *Genuine productivity improvements*. There is a limit to the gains in productivity that can be made by simply carrying out existing operations at a faster pace. A broader analysis must also be made of work and organizational procedures and structures (*see* Chapters 5 to 10).

(3) *A firmer theoretical foundation*. The application of computer-based technology on a wide scale amongst non-technical users is still in its infancy. It is, therefore, to be expected that research into human–computer communication, which must draw on real experience as well as experiment, is in an embryonic stage. Technological developments have also moved much faster than present psychological and behavioural research, and have been given much more support. In the future, an ergonomic methodology should emerge which is more systematic, more detailed and more reliable than is possible at present.

(4) *Managing technological innovation*. New technologies, such as graphics and multiwindow screens with a pointer device, give the designer increased scope for creating enhanced interfaces before clear guidelines have been established even for the much more limited older technologies. Increased technological power and sophistication can make interfaces far better than before, or make things much worse because designers and users are unable to master the new-found possibilities.

(5) *Identifying real limiting factors*. At present there is a general awareness that there are some limits to human and computing abilities. This needs to be explored further.

The first phases of computerization, from the 1950s to the 1970s, were dominated by the view that people, work procedures and organizations had to re-orientate themselves to behave in ways most acceptable and appropriate to the computer's mode of behaviour. The advent of low cost office systems, personal workstations and home computers has changed this. With good ergonomics, computers will cease being logical but rigid and insensitive dictators. Instead, they will become supportive, knowledgeable, helpful and adaptive partners and aids, as indicated in *Figure 3.2*.

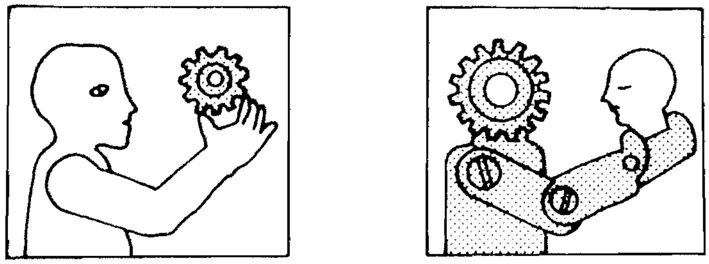

Figure 3.2 User-centred versus computer-centred systems (based on a design by Peter Hajnozcky in Zürich)

Recommendations

(1) Managers must play an active role in the design and implementation of systems to ensure they achieve the results desired by the organization as a whole and by each group and individual. This means having a good understanding of the issues involved in human–computer communication in order to act effectively, as follows:
   (a) Provide guidelines and requirement specifications to systems designers.
   (b) Create a working environment which is more supportive, less stressful and more productive than existing systems.
   (c) Select the right computer systems for themselves and other users of the technology.

(2) In order to gain acceptance of the technology amongst all managers, professionals and staff, the system and its application should be designed to be genuinely usable and useful in practical work situations.

(3) Managers should not expect to obtain foolproof ergonomic guidelines which will inevitably produce an optimal solution: there is no such thing. The process of developing and assessing good human–computer communication should be viewed as an attempt to balance trade offs between what is known of different users' needs and of the technologies available. The following are desirable characteristics for any computer-based system:
   (a) *Adaptability*. Computer systems should be able to evolve, grow and change to meet different and changing needs, preferably under the direct control and navigation of the user(s) most involved.

(b) *Knowledgeability*. The more a system knows about the behaviour of users, the tasks to be performed and the limits of its own capabilities, the more effective it is in interactions with users.
(c) *Helpfulness*. The system should assist the user to sift out relevant and useful information, provide prompts and advice on how to exploit computing capabilities to the full, and should limit the possibilities of making errors.
(d) *Forgivingness*. The user should be allowed to make mistakes without causing the whole system to collapse. The system should be friendly and helpful in informing the user of the causes of errors or breakdowns and the user should have the ability to retrace steps to undo some errors.

(4) System design should be user-centred, starting from the user on the outside and moving inside to the more technical requirements.
(5) More and more managerial and professional work will be carried out through personal workstations. It is therefore important that workstations not only have the ability to carry out the required functions (now and in the future) but also that these functions are integrated in a consistent framework.
(6) Knowledge-based and expert systems should be understood and carefully examined because they provide the most promising route for resolving many of the problems that have previously existed in human–computer communication.

References

1. Armbruster, A., Ergonomic requirements, in Otway, H.J. and Peltu, M. (eds), *New Office Technology: Human and Organizational Aspects*, Frances Pinter, London; Ablex, Norwood, N.J. (1983)
2. Jensen, S., Software and user satisfaction, ibid

Bibliography

The following publications are also relevant to topics discussed in this chapter.

ACM Computing Surveys, complete issue on interactive editing systems and document formatting systems, **14**(3) (1982)
Balzert, H. (ed), *Software Ergonomie*, Teubner, Stuttgart (1983)

Card, S.K., Moran, T.P. and Newell, A., *The Psychology of Human–computer Interaction*, Erlbaum, Hillsdale, N.J. (1983)
Feigenbaum, E.A. and McCorduck, P., *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*, Addison-Wesley, Reading, Mass. (1983)
Infotech State of the Art Report, *Man/Computer Communication*, Vol 1 and Vol 2, Infotech International, Maidenhead (1979)
Moran, T. (ed), The psychology of human–computer interaction, *ACM Computing Surveys*, **13**(1) (1981)
Simon, H.A., *The Sciences of The Artificial*, MIT Press, Cambridge, Mass. (2nd edition, 1981)