VOLUME 13 PART 2 JULY 1994

BRITISH TELECOMMUNICATIONS ENGINEERING

Included in this Issue

Human Factors in Telecommunications Engineering

New Network Infrastructures for the 21st Century

Speech Recognition for Speech Services

The Journal of The Institution of British Telecommunications Engineers

BRITISH TELECOMMUNICATIONS ENGINEERING

Contents

VOL 13 ■ PART 2 ■ JULY 1994

Published by: The Institution of British Telecommunications Engineers, 2–12 Gresham Street, London EC2V 7AG **ISSN 0262-401X**

GUEST EDITORIAL

Martin Cooper

Putting the User First

The customer sees, hears, feels, and ultimately judges, a system by its user interface

Readers of *British Telecommunications Engineering* will be familiar with developments in semiconductors, optics, radio, satellites and networks that are sweeping away the limitations of restricted bandwidth, distance, processing power and memory. Telecommunications and IT are advancing at an exponential rate and we might expect our customers to welcome the benefits of advanced services—better quality of life, reduced environmental impact, better international cooperation and lower costs. However, a barrier stands between these customers and the benefits of technology—the user interface.

Almost everyone has met an unfriendly system interface at some time: the video cassette recorder with obscure operating procedures, a computer with unfathomable command sequences to log the user off—even a simple device like an overhead projector can prove to be a challenge for a presenter under stress! The user interface of many services *is* the service from the customers' point of view. No matter how powerful or elegant the underlying technology, it remains just that—underlying—for all but the technophile. The customer sees, hears, feels, and ultimately judges, a system by its user interface. Ease of access and operation encourages a high level of usage. In a competitive situation, user friendliness can be the critical factor judged by the user in the selection of a supplier.

In round terms, information processing technology doubles in power each year. The abilities of people who use systems are not increasing anywhere near as fast—it takes millions of years for a similar doubling of our biological abilities. We are effectively rooted in the biological past, while acquired skills are stimulated and developed by new experiences, new tools. Technology can support learning and change the way we think, but only if we can understand and use it. If we do not harness the power of technology to make it more accessible to people, we risk leaving many behind—as non-participants in the future world of telecommunications—or as customers of competitors who *do* match the technology to users' needs and abilities.

It is misleading to think of the user interface as a thin layer, spread on top of a system. We cannot paste on user friendliness at the last minute to mask shortcomings in system design any more than a coat of paint can overcome the difficulties caused by a badly designed building. Only by involving users throughout the development of a new system can we be sure of identifying and satisfying their needs. To succeed, we must master the science and art of providing systems that people can use with the minimum of training, that are easy to use, that help users achieve their goals without undue effort. Many of the people responsible for commissioning, developing, delivering and selling BT systems have recognised the importance of these issues. The series of articles beginning in this issue of the *Journal* on the Human Factors theme will survey the field, describe techniques for identifying user needs, designing and prototyping user interfaces, evaluating usability and the application of these methods to telecommunications engineering. I hope that they raise your awareness of human factors, demystfying the theory and practice of matching systems to people and convince you of the vital importance of putting the user first.

Martin Cooper
Head of Human Factors,
BT Development and Procurement

Martin Cooper

Human Factors in Telecommunications Engineering

Telecommunications systems only succeed if people can use them easily to achieve their objectives. Theoretical findings have limited value as guidelines for usable system design and so a design process that involves users is required to fill the gaps. This article reviews the contribution that human factors knowledge and processes can make to the development of successful telecommunications systems.

Introduction

Telecommunications services generate sustained streams of revenue if, and only if, they enable their end users to achieve their goals. Increases in the raw power of technology increase the scope not only for delivering services that amplify human capabilities, but also for confusing users with obscure behaviour or complex user interfaces. Centuries of study have provided information on human capabilities and limitations. However, the complexities of people make it impossible to predict anything other than their basic performance as users of technology. In consequence, we cannot yet guarantee a usable system—first time, every time—without applying empirical methods that involve the user. Keeping the user firmly in mind throughout the development life cycle eliminates the causes of user difficulty and allows us to provide systems and services that people can, and will, use—repeatedly.

People do not want telecommunications services! They want to maintain relationships with distant friends and relations, to locate and understand information, order goods, strike deals, and much, much more. Telecommunications helps them achieve their goals, but very few people use telecommunications services as an end in itself. The aim must be to provide services that support human goals without getting in the way, to transparently amplify and extend human capabilities across the planet. People must be able to get the best from telecommunications services without years of technical study or weeks of training. Some customers will not even look at the instruction book before trying to operate the products and services and will reject them if this is not possible—especially if a competing product *is* easy to use.

Understanding People

We enter the world with certain basic abilities that unfold with time. These innate abilities to act, see, hear, think, and so on are modified by experience. We learn to interpret sights and sounds, to plan and act to achieve our goals. System designers can build on these skills to help users navigate around and interact with the worlds of information and communication services. We learn the arts of conversation, asking and answering questions, explaining what we mean and what we want through years of experience, observation, guidance and practice. The conventions of conversation, negotiation and delegation can also inspire telecommunication systems design.

Of course, the world has been shaped and changed by people and their tools over millions of years, and the pace of change is accelerating. Consider the portable telephone, the pocket calculator, the fax machine. Technology that once seemed strange and difficult is now more widespread and used easily by many people—especially the younger generation. Our skills are changing in response to these new tools, which in turn change

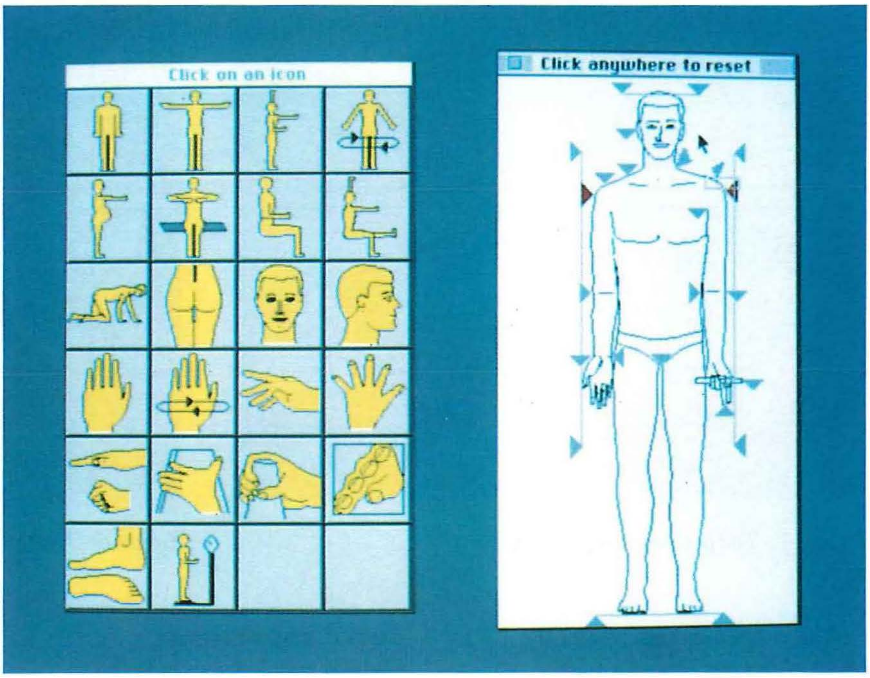


Figure 1—Anthropometric model

the tasks we attempt and the tools we need. This cycle continues as technology and its use evolve. There is never a single, all-time solution to a human need—only a solution that is right in a particular set of circumstances. The first step to finding the solution is an understanding of people—their physical, psychological and social characteristics.

Physical dimensions

Numerous studies have charted the size and shape of men, women and children from a wide range of ethnic backgrounds. For example, a recent anthropometric database[1] contains the distribution of over 200 dimensions of four major population groups and several age ranges (see Figure 1). This data is valuable for guiding the physical design of control panels, products, enclosures, furniture and so on.

Perception

All the human senses have been explored, mapped and documented. Studies on the basic sensitivity and discrimination of the senses provide baseline data to ensure that people will be able to detect the output of a system and distinguish its features. Models of the various perceptual systems can be used to predict human performance. For example, one study[2] has shown that people interpret colour codes for an ordered series of values with fewer errors if the colours consist of a series of mixtures of two colours. Codes that include a third hue are less easy to interpret.

Memory

Information is held in a limited-capacity working memory before being transferred into long-term storage. Many models of memory have been proposed that deal with the amount of information that can be stored, its decay with time and how this is affected by rehearsal and interference, coding schemes, association and retrieval mechanisms. Figure 2 shows the recall of lists 24 hours after they were memorised, as a function of the number of total lists learned. The previously learned material interferes with later learning—confirmation that it is difficult to teach an old dog new tricks.

Motor skills

Psychologists have extensively studied perceptual-motor skills. One series of experiments focused on repeatedly moving the hand from a starting position to a target. This culminated in the formulation of *Fitts's law*[3] that relates the time per movement, T, to a relationship between the distance, D, to the target and its size, S. Welford[4] developed a more accurate variant of the law (see Figure 3):

$$T = I_M \log_2 (D/S + 0{\cdot}5),$$

where I_M is a constant.

Figure 2—Interference of previously learned material with later learning

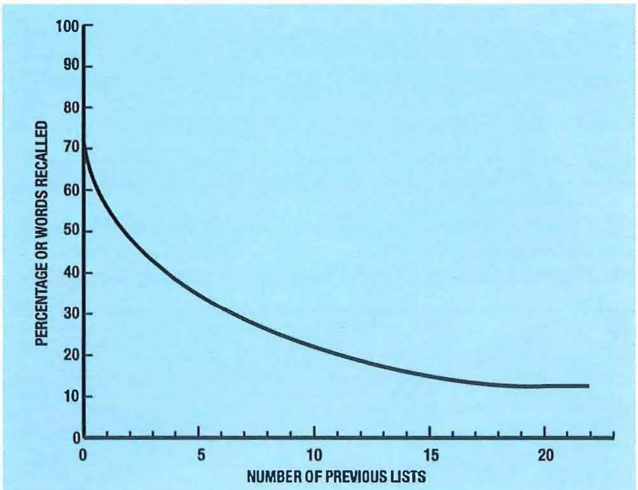
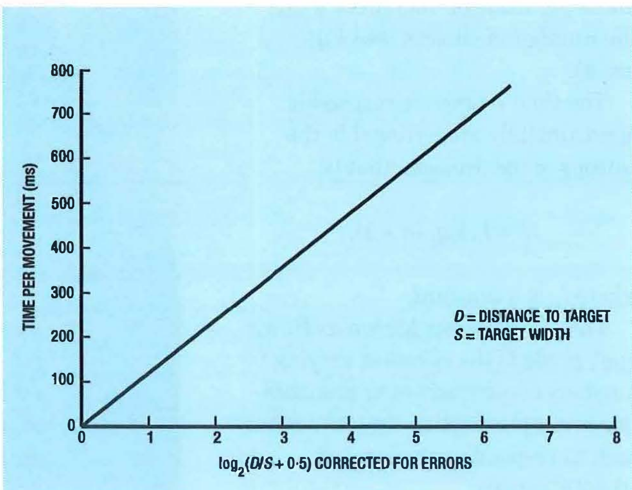


Figure 3—Welford's variant of Fitts's law



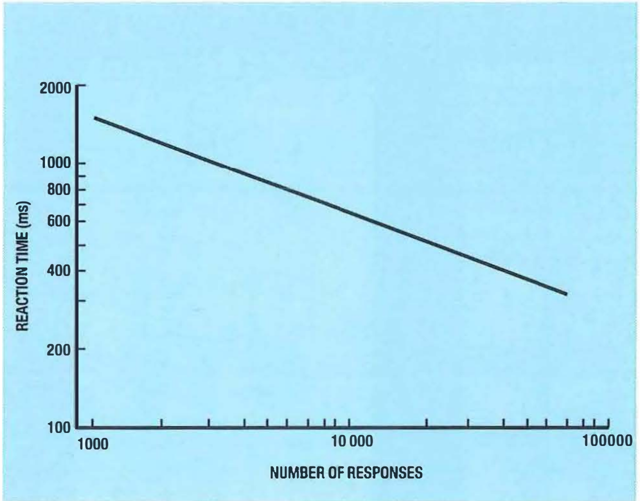


Figure 4—The power law of practice

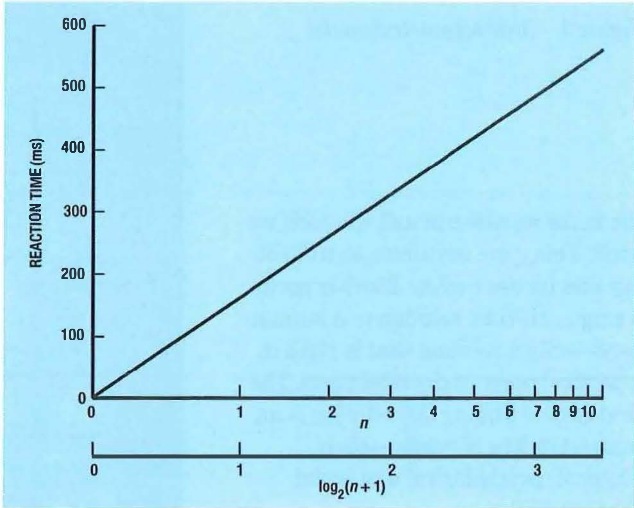


Figure 5—Hick's law of choice reaction time

These findings can predict the effect of the layout and size of controls on operating times.

The adage that practice makes perfect has been confirmed by studies of perceptual-motor learning. This led to the establishment of the *power law of practice*[5]—that the time T_n to perform a task on the nth occasion follows a power law:

$$T_n = T_1 n^{-a},$$

where T_1 is the time to perform the task on the first trial and a is a constant.

Figure 4 shows the result of applying this law to the improvement of reaction time with practice on a task in which subjects pressed combinations of ten keys in response to patterns of lights directly above the keys.

Simple decision making

If a user has to choose between several responses, the time taken to make the decision increases with the number of choices (see Figure 5).

The time T taken to respond is approximately proportional to the entropy of the decision, that is:

$$T = I_C \log_2 (n + 1),$$

where I_C is a constant.

This relationship, known as *Hick's law*[6], predicts the effects of varying numbers of alternatives or probabilities in simple reaction-time situations such as responding to visual or acoustic signals.

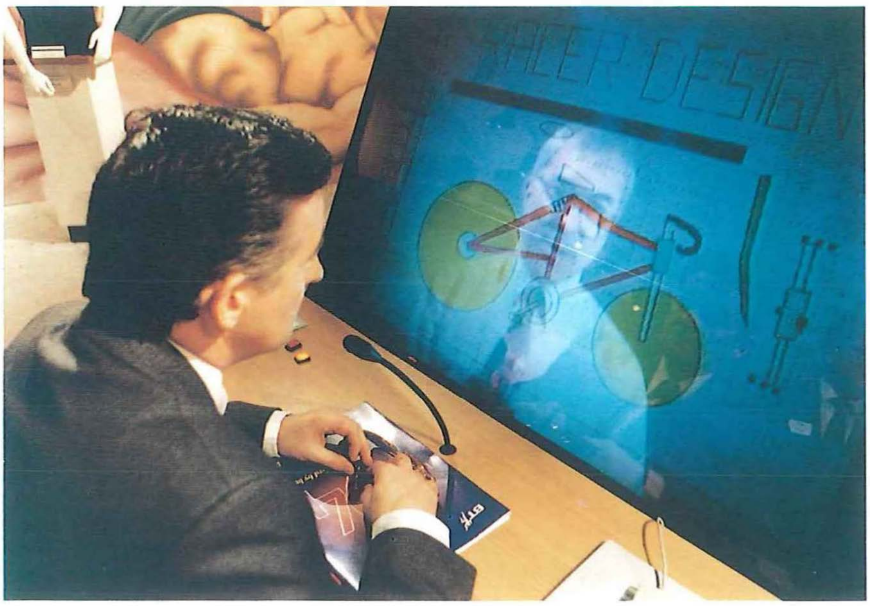
Social psychology

Social psychologists study the behaviour of people interacting with other people. This is considerably more complicated than interacting with the physical world owing to the rich complexity of human relationships. Topics include how people perceive the causes of the behaviour of others and of themselves, of social behaviour patterns and stereotypes, the processes involved in forming an impression of another person, of influencing others and of group decision-making. In view of the role of telecommunications as a vehicle for interaction, it is not surprising that social psychology has much to say of interest to the telecommunications engineer. For example, the importance of patterns of gaze to communications, persuasion, deception and perceived character have implications for videophone design. Figure 6 shows a prototype videophone in which a camera effectively looks directly at the caller and eliminates the 'averted eyes' effect caused by a displaced camera angle.

People are difficult to understand

People are very difficult to model. Tens of billions of heterogeneous, parallel, statistical processing elements in the human brain are programmed by gigabits of stored information interacting with a similarly huge amount of input data coming from the senses every second. Both remembered and incoming data are different in some ways for each individual. Human interaction by means of telecommunications is

Figure 6—Through-the-screen videophone



Understanding the situation in which they will be used forms the basis for designing usable systems

further complicated by social factors and the effects of the telecommunications technology itself. The result is a chaotic, complex, unpredictable, non-linear system. An integrated model of this system that could guide the design of a telecommunications service, system or product would be invaluable, but is not yet available. In the interim, we have to rely on the qualitative results of social psychology and the low-level findings of the sort outlined above, refined by means of empirical usability engineering—which is described below.

Techniques to Develop Usable Systems

Analysing the content of use

Understanding the situation in which they will be used forms the basis for designing usable systems (see Figure 7). The three key aspects are the users, their tasks and their environment. User characteristics help define the requirements of the system. The physical, psychological and social environment imply physical, information content, presentation and communications requirements for the system. Establishing a rich picture of who will use the system, what their objectives will be when they use it and the environment in which it will be used pays off many times during the later stages of development. A set of usability requirements, in terms of the target performance and the opinion of particular types of people completing specified tasks in typical environments, can be derived from the analysis for use later in validation and testing.

User interface design

Designing usable systems is a craft, not a science. The starting point is the environment of use—a noisy environment may rule out a speech interface, a mobile user needs a portable system. The next step is to determine the flows of information across the interface. The users' view of the world is translated into the objects, behaviours, interaction and dialogues. Storyboards sketch out the steps in each transaction and communicate the evolving design to team members and representatives of the user population. Consistency is an important consideration. When users encounter systems that do similar tasks or are used in proximity, their experience with one system either helps or hinders their use of the others. Systems are coupled by their common users and this must be considered during their design. The application of common design principles and agreed common elements (such as welcome screens) minimises inconsistencies between systems.

Usability evaluation

Each step of the process, from the establishment of the context of use, to the delivery of the final system, benefits from usability evaluation. Available techniques include cognitive walk-throughs, heuristic evaluation, laboratory tests with users (see Figure 8) and field testing. All evaluations need to be carefully designed to extract useful information with the least cost. Tests can be conducted on paper specifications,

Figure 7—Human factors in the development process

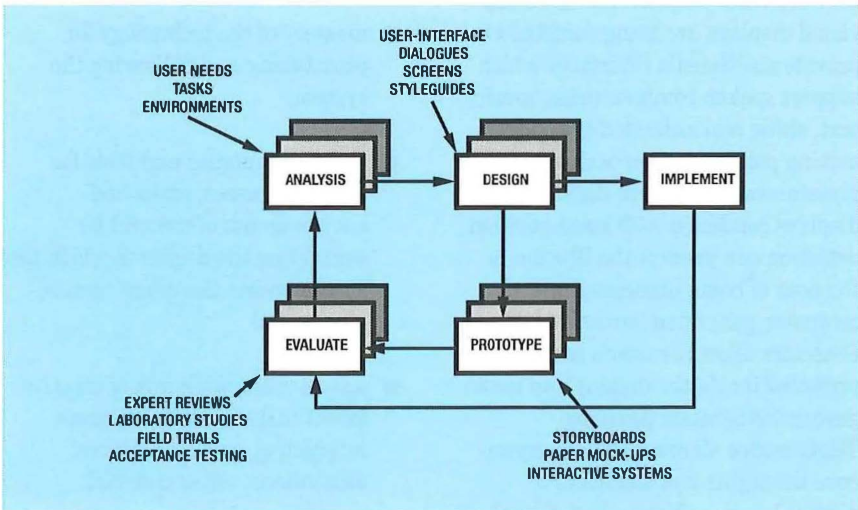
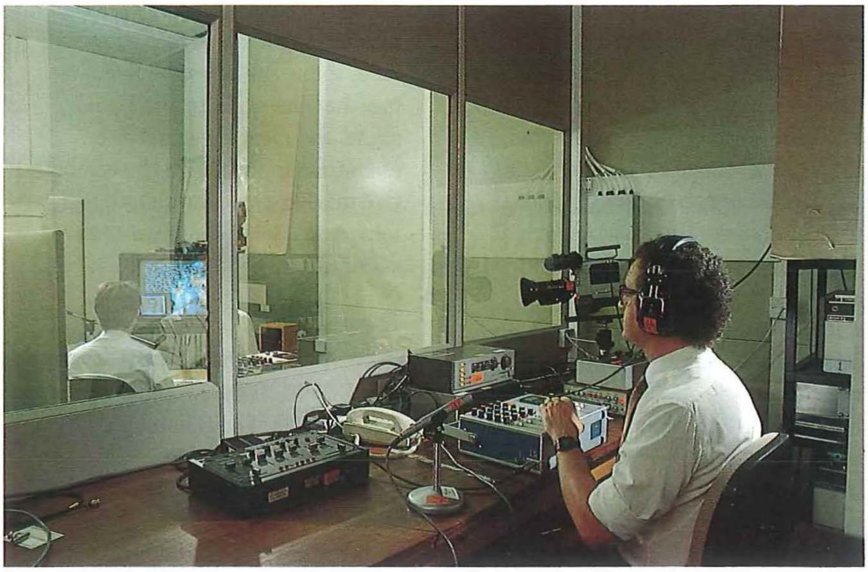


Figure 8—Usability test laboratory



Increases in the power and reductions in the costs of computing, display and control technology have made new types of user interface possible.

storyboards, mock-ups and more sophisticated prototypes (including the Wizard of Oz technique where a person imitates the behaviour of the system). Finally, the implemented system itself can be tested to confirm that it meets the usability requirements.

User Interfaces for the Future

Human interface technology

Human interface technology is continually expanding. Increases in the power and reductions in the costs of computing, display and control technology have made new types of user interface possible. Acoustic and visual displays are being combined to provide multimedia interfaces which support spoken commentaries, music, text, static and animated graphics, moving pictures and speech and video communications. Three-dimensional displays combined with head-position detection can present the illusion to the user of being immersed in a computer-generated 'virtual world'. Detectors allow the user's hand to be projected inside the display and make gesture recognition possible. Telepresence allows a user to experience the sights and sounds of a distant location. Networked virtual-reality systems can project an image of a user into a virtual world where meetings take place between people from anywhere on the planet.

Alongside the development of display and control technology, advances in computing power and artificial intelligence are increasing the sophistication of conversational user interfaces. Text, particularly if the topic of discussion is constrained, can be understood quite well by natural language processing systems. Recognition of limited spoken vocabularies is already in use, though unconstrained spoken input is still some way off. Intelligent agents are being developed that can carry out tasks on behalf of a user in much the same way as a personal assistant carries out delegated tasks. Applying the user-centred methods outlined in this article reduces the chances that this new technology will fail due to the human factor.

Design skills for usable new technology

The convergence of computing, communications, television and publishing has major implications for the design of user interfaces. These developments blur the distinction between the control of an information system and the content of the information it provides. Factors determining the acceptability of these systems to their users include:

- mastery of the technology for prototyping and delivering the system;
- skills, techniques and tools for analysing users, tasks and environments of use and for evaluating the degree to which the system meets the users' requirements; and
- access to the wide mix of creative talent to design and integrate interactive sounds, pictures, animations, video material.

The talents of engineers, human factors people and designers are needed to ensure that the telecommunications systems of tomorrow will meet the needs of users. Bringing together the skills to elicit, understand and meet these needs is essential for success in the competitive market-place for today's and tomorrow's telecommunications services.

References

1 PEOPLESIZE. Friendly Systems Ltd. 1994, Loughborough, UK.

2 May, K. and Travis, D. A Closer Look at Intuitive Colour Scales. BT internal report, 1992.

3 Fitts, P. M. The Information Capacity of the Human Motor System in Controlling Amplitude of Movement. *Journal of Experimental Psychology*, 1954, **47**, pp. 381–391.

4 Welford, A. T. Fundamentals of Skill. 1968. London: Methuen.

5 Snoddy, G. S. Learning and Stability. *Journal of Applied Psychology*, 1926, **10**, pp. 1–36.

6 Hick, W. E. On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology*, 1952, **4**, pp. 11–26.

Biography

Martin Cooper
BT Development and Procurement

Martin Cooper joined the British Post Office in 1967 as a Student Apprentice. In 1971, he graduated from Southampton University with a B.Sc. Honours degree in Electronic Engineering and gained an M.Sc. in Work Design and Ergonomics in 1975. He is a Fellow of the IEE and the Ergonomics Society. Since graduation, he has worked on a variety of projects in the human factors field. In 1988, he was awarded the Sir Frederic Bartlett medal of the Ergonomics Society for his contribution to the application of ergonomics in industry. His current role is Manager of BT Development and Procurement's Human Factors Unit, a team of 60 psychologists, ergonomists, designers and engineers, where he continues to study ways of matching future telecommunications systems to the needs of their users.

Ian Hawker, Vivek Tandon, David Cotter and Alan Hill

New Network Infrastructures for the 21st Century

Part 1: Ultra-high-capacity wavelength-routed networks

In the last 20 years, advances in telecommunications technology have been staggering. However, increasingly the traditional approach of operating networks is becoming unsuited to the needs of customers and network operators. Radically-new network solutions are envisaged for the next century which will realise greater simplicity and higher reliability.

Introduction

Despite the revolutionary advances in transmission and switching technologies witnessed in the last 20 years, the way in which telecommunications networks are operated remains fixed in the age-old traditions of narrowband telephony. Increasingly this established approach will be found to be fundamentally unsuited to the needs of customers and network operators. This will be due to:

- greatly increased requirements for network capacity through intensive use of broadband interactive services, and
- the need to lower costs by drastically reducing network complexity, software burden, and management.

At a fundamental level, optical-fibre technologies have already overcome the problem of insufficient network capacity. However, in practice their capacity is limited by concatenated electronic switches and regenerators in the fibre path. Future photonic networks will offer greatly increased capacity as they will allow optically transparent, end-to-end connections between customers, giving direct access to a larger proportion of the immense fibre bandwidth[1]. Hence bandwidth is expected to become a cheap and plentiful commodity. In fact, bandwidth availability is already on the brink of changing the network and service design paradigm. In the future, network designers will break with tradition and take a radically new approach—searching for options realising greater simplicity and higher reliability to reduce control and management costs—at the expense of bandwidth efficiency.

This article describes radically new switching solutions for broadband networks based on optical Lambda-PONs (multi-wavelength interconnected passive optical networks)[2], and how an almost 'switchless' national network can be achieved using either passive or active optical routing nodes. Part 2, to be published in the October 1994 edition of the *Journal*, will describe ultra-fast optical packet routing, based on primitive autonomous network nodes. These provide a degree of self-organisation that realises a dramatically simplified network management regime.

Optical Lambda-PONs within Metropolitan Areas

The future expansion of broadband services implies that both the business and residential user will generate increasing amounts of data traffic and require access to higher-capacity data-channels. It is widely accepted that the introduction of these new services will require transport and processing capacities far larger than those available using today's electronic-based technologies[3].

Providing point-to-point optical links to the residential user is expensive owing to the opto-electronic infrastructure required and the cost

The routing of information between customers is performed entirely in the optical domain with no electronic switching or buffering of the signals required.

of laying and providing the optical fibre. Passive optical networks (PONs) offer a method of sharing the transmission costs of the access network among a large number of customers by allowing fibre and bandwidth sharing[4]. Implementations of PONs to date are primarily based on time-division multiplexing (TDM), most notably in the form of TPON[5]. An alternative approach is to use wavelength-division multiplexing (WDM). This exploits the vast bandwidth of fibre by allowing traffic routing functions to be performed in the optical domain. The problems of non-linearities and crosstalk[6] at very high bit rates, which ultimately limit the number of wavelengths available, can be overcome by wavelength reuse.

Consider the interconnection of PONs using two unidirectional fibre rings; one for clockwise and one for anticlockwise traffic (Figure 1). Such a configuration could be implemented within a metropolitan region linking clusters of users. Users have access to the fibre rings via a *passive routing unit* (described later) and wavelength reuse is achieved by tunable wavelength selective filters which can be turned on and off within milliseconds to define the wavelengths used and the areas within which they are confined. Wavelengths allocated to calls originating and terminating on the same PON or Lambda-PON can then be used elsewhere in the Lambda-PON. The network shown offers the following benefits:

- It allows customers to construct an optical connection between any two (or more) terminals. The routing of information between customers is performed entirely in the optical domain with no electronic switching or buffering of the signals required.
- Wavelengths are not pre-assigned to the customers[7] but allocated to customers dynamically as required.
- Customers can be allocated bandwidth on demand since all the devices used to route the information are bit-rate independent.
- Once the wavelengths have been allocated, the delay is negligible since the scheme does not use time-division multiplexing or packet switching. Call set-up time is limited only by the network controller's processing speed, the tuning speed of the transceivers and the propagation delay (typically a few milliseconds).
- Routing within the network is achieved by the use of controllable tunable wavelength selectors which are required to operate only at the speed of the call set-up process.
- Storage of network control information is achieved by distributed network control units. This provides network resilience and reduces the propagation delay during the call set-up phase.
- The network does not require synchronisation or wavelength conversion.
- The customer experiences low blocking, low delay, has true real-time bandwidth on demand and high reliability.

Figure 1—A future Lambda-PON showing eight interconnected PONs

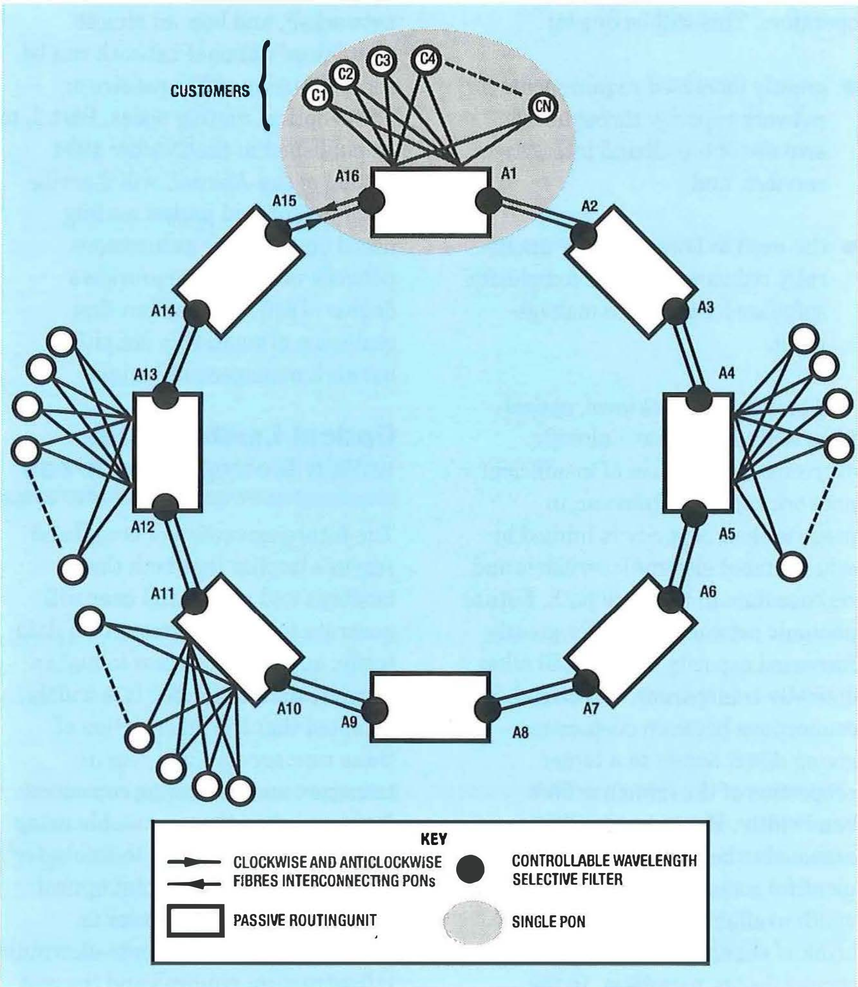


Figure 2 shows the occurrence of five unrelated circuit-switched calls† between 10 users (A to A', B to B', α to α' etc) using only four wavelengths. By turning on the wavelength-selective filters A1 and A16 to absorb

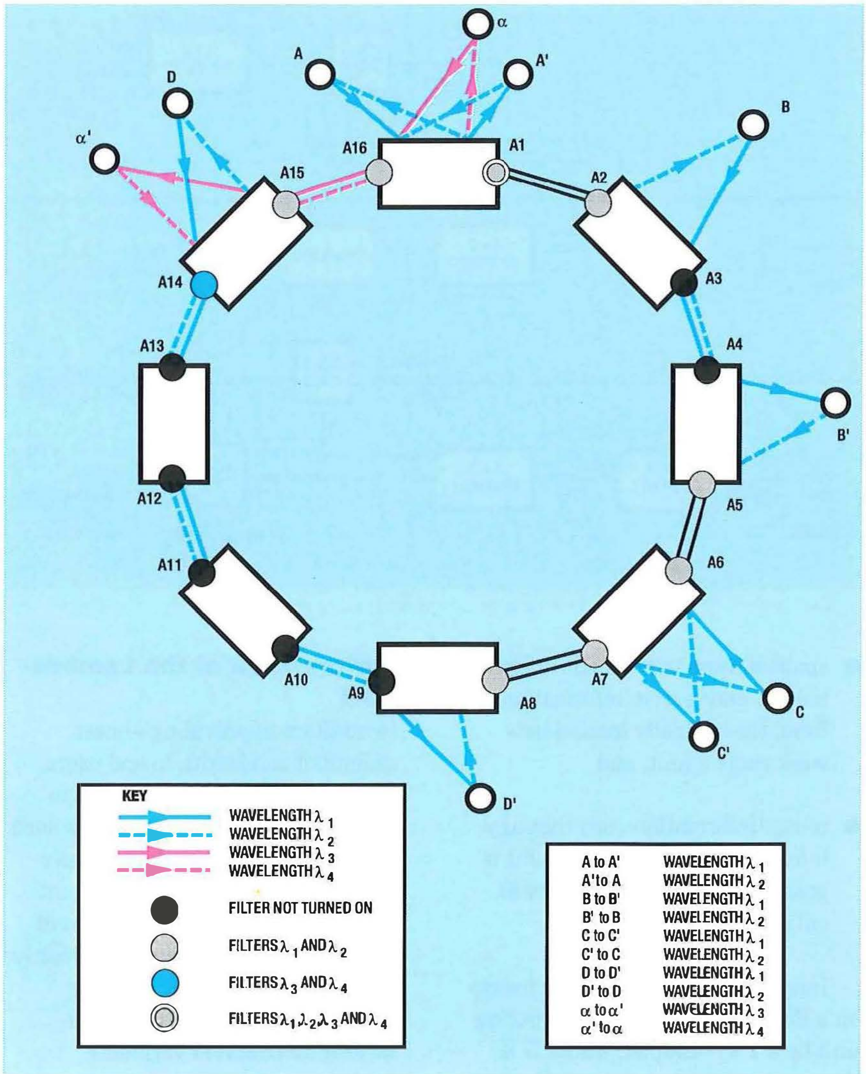


Figure 2—Wavelength reuse within a Lambda-PON

out wavelengths λ_1 and λ_2, the two wavelengths are localised to that PON and hence can be reused elsewhere. In the event that two paths overlap, a new pair of wavelengths to interconnect the users is required; as shown by the connection between α and α'. In practice, several hundred users are connected to each PON and all the information entering or leaving a PON is broadcast to every user on that PON due to a $1 \times N$ splitter (where N is the number of users on a PON) located in the passive routing unit (described later).

Controlling the network

Controlling the tunable wavelength selective filters and the dynamic allocation of wavelengths on a per-call basis can be achieved by the use of a network controller located at each Lambda-PON (Figure 3). The network controller contains a routing table to keep track of the wavelengths in use on the various routes within the Lambda-PON, the state of the users (engaged/available) and the current state of the absorbers (on/off). In addition, the network control unit has a direct link to each of the controllable absorbers enabling it to turn them on and off to absorb the desired wavelengths. Accessing the network control unit is accomplished by transmitting a control channel packet on a common control wavelength. Contention for the network control unit is resolved by using a suitable protocol such as ALOHA.

The passive routing unit (PRU)

The passive routing unit (Figure 4) performs three main functions:

- enables users to transmit and receive information from other users on the network,

---

† The term *call* refers to any traffic generated which is transmitted into the network whether it be straightforward voice-telephony type traffic or some future multimedia service

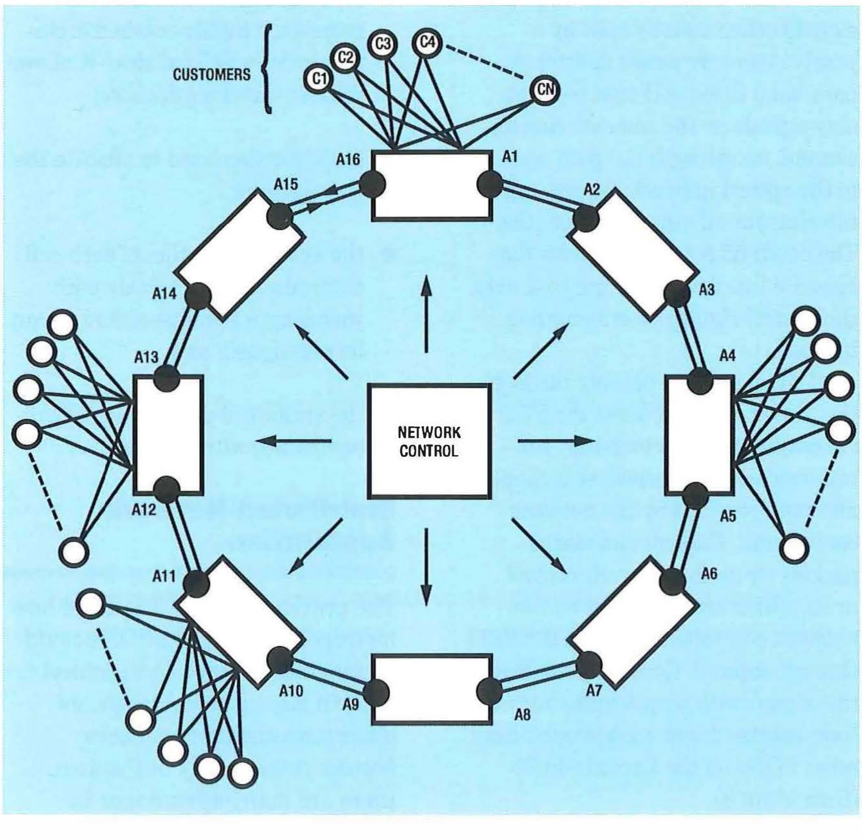


Figure 3—Connection of PONs to the network control unit

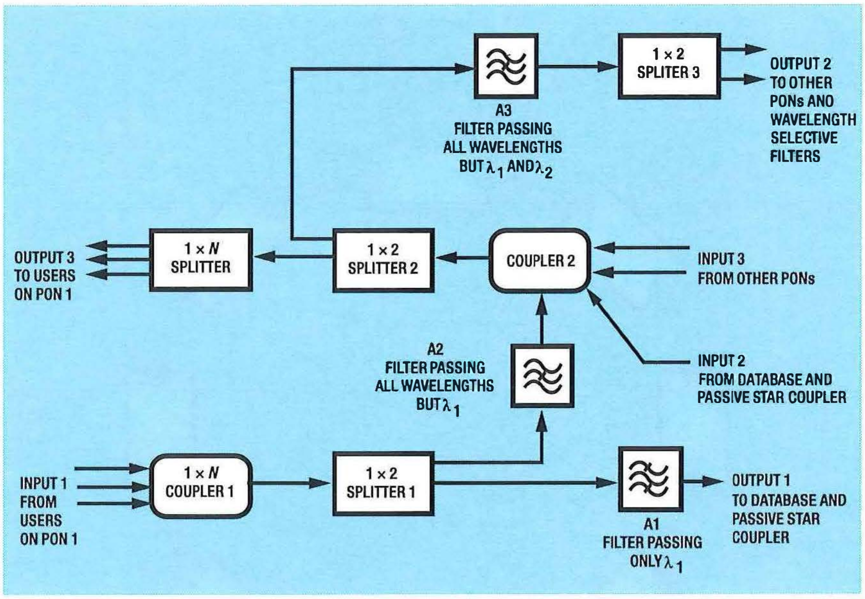


Figure 4—The passive routing unit

- enables users to transmit information to, and receive information from, the centrally located network control unit, and
- routes information such that any information input into the unit is passively routed to the relevant output ports.

Information transmitted by users, on a PON, enters the passive routing unit by a 1 × N coupler, where N is the number of users connected to each PON (labelled as 'input 1' in Figure 4), which combines all the signals onto one fibre. This combined signal is then equally split by a passive two-way power splitter. A pass-band filter (A1) ensures that only signals on the common control channel wavelength (λ_1) gain access to the optical network control unit by filtering out all other wavelengths. The notch filter (A2), performs the opposite function, ensuring that only the control channel wavelength is blocked.

Control packets passing through the pass-band filter leave the PRU via output 1 after which they are rebroadcast by a passive star coupler and then received by the network control unit. Control and status packets (from the network control unit), which are broadcast by the network controller, arrive at the PRU through input 2. Coupler 2 combines this signal with wavelengths received from splitter 1 and wavelengths from other PONs on the Lambda-PON (from input 3).

Performance of the Lambda-PON

In addition to providing almost unlimited bandwidth to end users, the Lambda-PON must be able to meet other performance metrics such as blocking probabilities (typically about 2% for speech in the present network) and delay. The amount of wavelength reuse achievable directly affects the wavelength blocking probability and is dependent on several interrelated variables:

- the topology of the network—at present a simple bidirectional ring topology has been discussed; however, a highly-meshed architecture is beneficial since it allows greater wavelength reuse;
- the algorithm used to allocate the wavelengths;
- the average duration of each call determining how quickly each wavelength is released and so can be reassigned; and
- the statistical distribution of call occurrence with distance.

Distributed Network Restoration

The previous section discussed how metropolitan Lambda-PONs could be controlled from a local control unit. In more general terms, as telecommunication networks become increasingly automated, there are many advantages in distributing the higher-level management/control processes such as network restoration. These include speed of response, robustness and reduced software complexity.

A radical approach of achieving these higher network functionalities is to embed the rules/algorithms within the individual network elements and switches enabling restoration to be achieved automatically in response to a network event/request. This method eliminates the involvement of the network management centre for restoration. Since basic intelligence is built into the network elements, network restoration can be performed very quickly and efficiently.

One method studied involves distributed link/node restoration which uses a simple but efficient flood search to identify alternative routes quickly and automatically[8]. Any new line system or node added to the network is automatically protected, subject to spare capacity availability, since there are no protection plans to modify. Each node knows only its unique identity and contains a simple set of rules that tell it how to react when it sees an alarm or message from a neighbouring node.

Simulation results (Figure 5) indicate that distributed span restoration in a network is feasible within 1 second, offering restoration within the call drop-out threshold of System X switches to provide customers with an uninterrupted service. A processing time of 5 ms and a cross-connect time of 20 ms are typical, representing modest modifications to current cross-connect specifications. In addition, a combination of pre-planned (that is, distributed restoration routes created continually in background mode) and real-time flood searching reduces overall restoration time by at least 50%. The restoration mechanisms described can handle node failures based on concurrent implementation of end-to-end path restoration. Whereas node loss is expected to be relatively rare, it is potentially catastrophic!

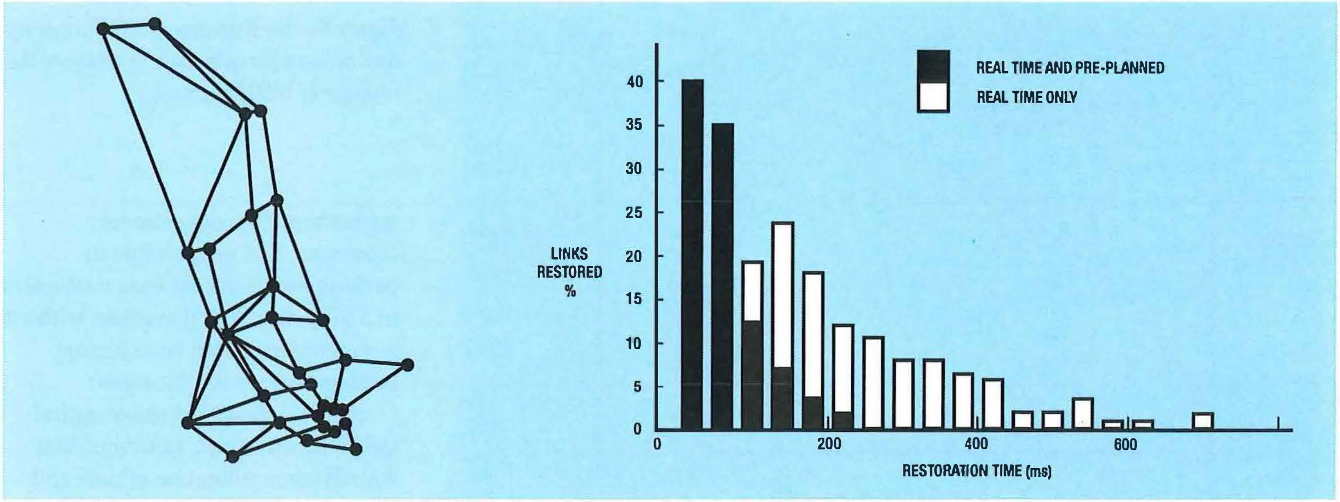


Figure 5—Simulated restoration times on a UK network

Towards the Switchless Network

Optical transparency eliminates the need for unnecessary opto-electronic conversion hardware and high-order electronic multiplexers, and could substantially reduce the number of network active switching nodes through the much increased reach of PONs compared with copper cables in the access network. Indeed, it may eliminate switching nodes completely! Such a minimal approach would take the traditional trade-off between transmission and switching to its logical extreme, where switching is performed in a fully distributed manner, such that switching and transmission become indivisibly one task. With few or no switches to control, the software burden of operating such a network will be greatly reduced, particularly in relation to network management.

The initial use of amplified PONs (henceforward termed *transparent optical networks* (TONs)) will provide multiplexing and consolidation of access traffic to reduce the number of switch nodes. But, with optical amplifiers, TON capacities will be sufficiently large to make distributed switching an attractive possibility by allocating time and wavelength channels dynamically to individual customer–customer connections, without the use of electronic switches. There is, however, a difference between the capability of upstream and downstream TONs which should be considered.

Downstream TONs

A very large customer base can be served in the downstream direction. The possibility of broadcasting downstream from a single head-end, to tens of millions of customers has been demonstrated, over 527 km[9]. The experiment demonstrated the ability to deliver 40 Gbit/s capacity, shared between 16 wavelength channels with 1 nm channel spacing, sufficient for 256 uncompressed video channels (or perhaps as many as 8000 compressed to about 5 Mbit/s). In such a network, channel selection at customer terminals would be a combination of tunable optical filter and electrical time-slot selection. With such capacities, a single CATV head-end could serve an entire nation. Indeed, computer models suggest that, in principle, optical amplifiers with flattened wavelength responses could support 400 or more wavelength channels across the erbium-doped fibre amplifier window, over 500 km, with operation at 2·5 Gbit/s. This would give 1 Tbit/s downstream capacity, sufficient for at least 200 000 compressed video channels. Nonlinearities in the fibre restrict this to about 0·5 Tbit/s[10].

Such capacity within a single TON could be used in various ways. For example, it would be sufficient to provide a near-demand video library service, by staggered, yet simultaneous, broadcasting of multiple copies of thousands of film titles. All this could be from a single server, giving national coverage—with no switching requirements. Another possibility with so many channels, assuming sufficiently low penetration and high concentration, would be for the introduction of a nationwide video-on-demand service from a single national server. An initial service could begin with a single server, and additional servers would be installed as required, as the customer base expanded.

Alternatively, the same capacity corresponds to about 13 million telephony channels, which would be sufficient to support the entire telephony traffic of the UK, with relatively low levels of concentration. If the same number of channels could be supported in the upstream direction, the entire telephony service could be provided over a single looped-back TON, without any central switch node.

With few or no switches to control, the software burden of operating a network will be greatly reduced

Upstream TONs

For conventional topologies the achievable split and capacity of upstream TONs is the limiting factor in moving towards a switchless network. Optical amplifiers impose major differences between upstream and downstream TONs, because of quantum processes within the amplifiers, that generate amplified

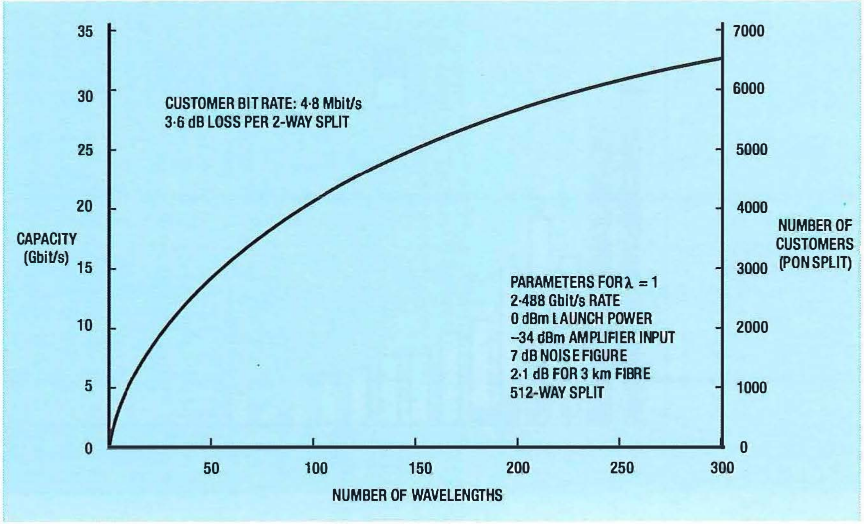


Figure 6—Total upstream PON capacity and achievable split, as a function of the number of WDM channels

spontaneous emission. Figure 6 shows the achievable split and capacity for customers provided with a simultaneous average bit rate of about 5 Mbit/s, which is likely to be sufficient to support high-quality compressed video. With just one operating wavelength channel, the power budget should be sufficient to combine 512 customers, at an aggregate rate of 2·5 Gbit/s. But to support more customers the aggregate bit rate must be reduced, because a larger split will reduce the power budget. For example, at 155 Mbit/s, about 5200 customer fibres can be combined. However, to maintain the same calling rates and degree of concentration, 160 wavelengths are needed to carry the additional capacity, and to compensate for the reduction in bit rate. The total TON capacity is about 25 Gbit/s at a bit rate of 155 Mbit/s on each of the 160 wavelengths.

Taking the extreme, where the aggregate bit rate exactly equals the customer bit rate, and every customer connection is allocated its own wavelength (that is, no electrical multiplexing employed), the TON split could support some 94 000 customers at about 5 Mbit/s each. The total upstream TON capacity would reach a maximum of 0·5 Tbit/s.

Minimum switch requirements for a broadband national network

The fundamental asymmetry between upstream and downstream TONs means that, for conventional topologies, it is not possible to perform switching between customers in a fully distributed manner, without a centralised switch node (except perhaps at very low bit-rates).

Figure 7 shows a possible optical switch design which, in conjunction with dynamic allocation of time and wavelength channels at customer terminals, is capable of interconnecting the entire broadband traffic for 26 000 000 customers within a single switch node. The design is based upon the MONET back-plane principle[11], using on-off spatial light modulators (SLMs) as the switching devices. The switch performs as a time- and wavelength-shared, multi-stage rearrangeable space switch, in which each SLM array represents a stage of 2 × 2 optical change-over switches. The relevant parameters are given in Table 1. The switch requires just 15 SLM arrays to provide a network capacity of 13 Tbit/s, equivalent to

Figure 7—Time- and wavelength-shared multi-stage rearrangeable space switch, using three-dimensional spatial light modulators

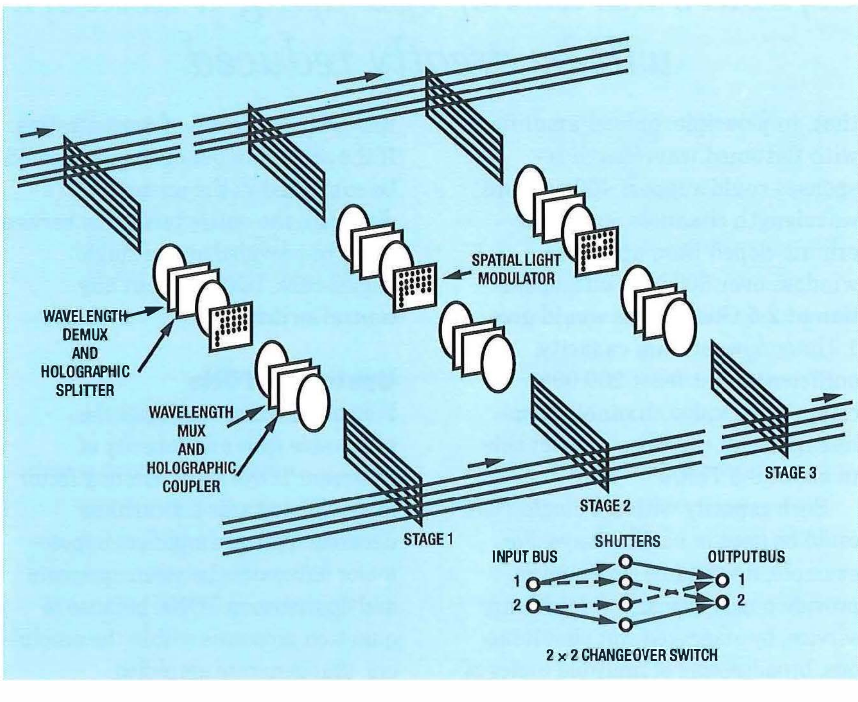


Table 1 Design Parameters for a Single 13 Tbit/s Broadband Switch for the UK. (The capacity is approximately equivalent to 120 160 Gbit/s throughput ATM switches.)

Number of customers	26 000 000
Number of simultaneous 5 Mbit/s broadband connections (10:1 concentration)	2 600 000
Customer bit rate	300 Mbit/s
Time-sharing factor (for example, number of time-slots)	60
Number of wavelengths	208
Number of buses	208
Number of space/wavelength channels	43 264
Shutter array size	208 × 208
Number of 3D SLM optical switches	15

The ability to eliminate all active switch nodes from the network is achieved by applying the wavelength-routing technique between every TON in the network

about one hundred and twenty 160 Gbit/s throughput asynchronous transfer mode (ATM) switches.

An almost switchless broadband national network

The ability to eliminate all active switch nodes from the network is achieved by applying the wavelength-routing technique between every TON in the network (Figure 8).

Individual TONs are all taken to a single, central node (in practice a number of stand-by nodes would also be required) consisting of one wavelength-division-multiplexing device—a purely passive optical component. No active switching is required. Traffic is routed from an upstream to a downstream TON by tuning the customer transmitter to the appropriate wavelength to couple between the correct pair of fibres on the multiplexer. The number of TON ports on the multiplexer, n, depends on the total network customer base, N, and the upstream split m of each TON. An additional set of dummy ports is required to allow more than one wavelength channel to couple simultaneously, to satisfy the traffic requirements.

Figure 9 shows the total number of ports required on the multiplexer for N=20 000 000 customers, for simultaneous average customer bit rates of 5 Mbit/s. There is clearly an optimum bit rate of about 50 Mbit/s which minimises the total number of ports, and hence the number of wavelength channels required to about 2000. With this number, the maximum network capacity is about 160 Tbit/s. If non-linearities and amplifier noise accumulation were to allow end-to-end operation at 155 Mbit/s, a formidable 2350 Tbit/s would be available for long-term capacity growth (that is, 20 000 000 customers each with 118 Mbit/s simultaneous capacity). But further studies are needed to determine the maximum operational bit rate that could be supported transparently across the network with so many wavelength channels. Also, the system described is fibre lean! Introducing a far higher degree of space-division multiplexing—fibre rich—would increase the above capacities beyond any effective limit. However, the fundamental elements to realise such an option are still under investigation. Ultimately, there is the potential for unlimited bandwidth and connectivity!

Figure 8—A broadband national network, employing wavelength routing between customers through a single passive wavelength routing node

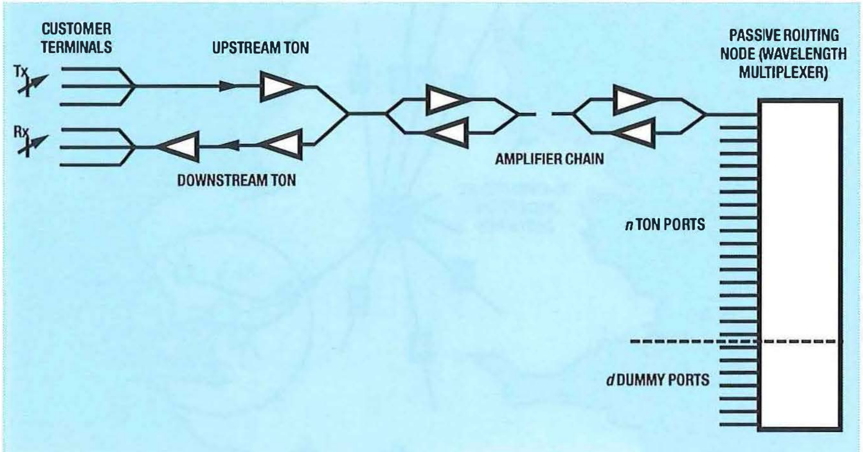


Figure 9—Number of multiplexer ports

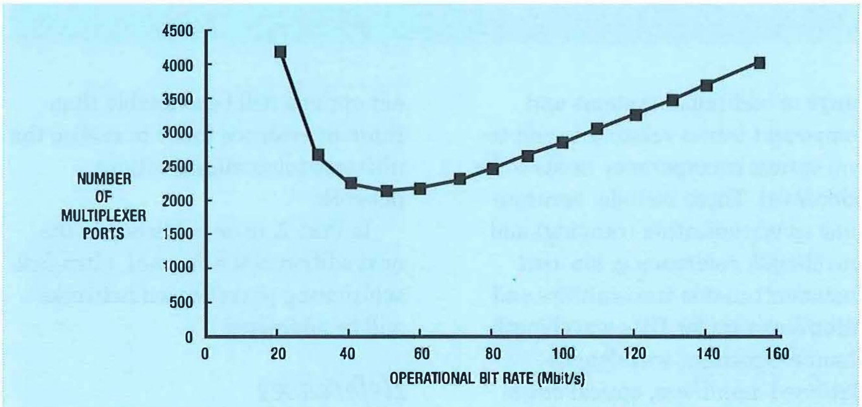
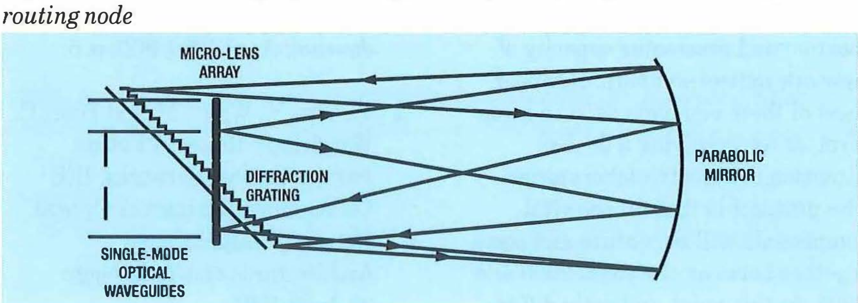


Figure 10—Possible design of wavelength multiplexer for a passive wavelength routing node



The passive wavelength-routing node

The passive wavelength multiplexer performing the wavelength routing function between transparent optical networks is in principle a very simple device (Figure 10), similar to commercial optical spectrographs and monochromators.

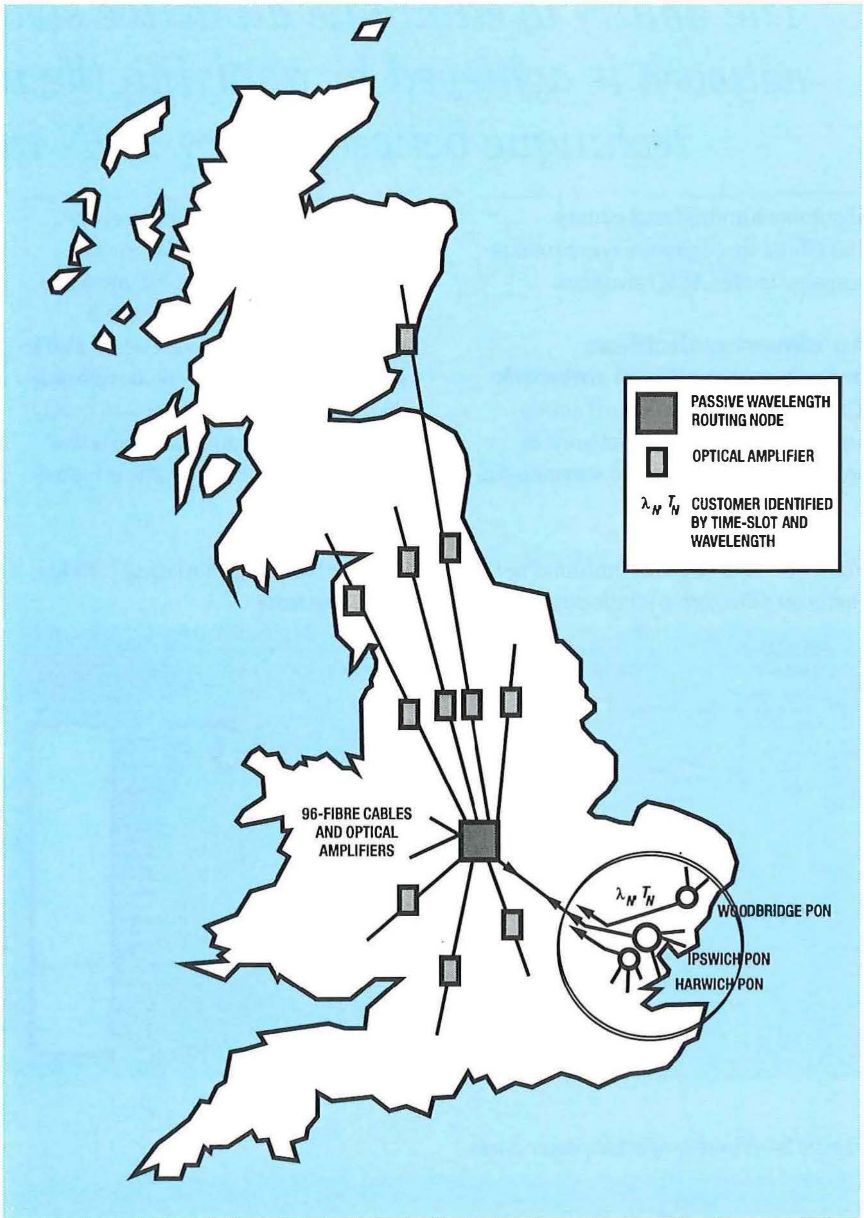


Figure 11—Schematic illustrating the concept of zero switching within the national network

The device would be used in a far more powerful way than is conventionally employed for wavelength multiplexing. Instead of just combining light from a number of input fibres into a single output fibre ($1 \times n$ connectivity), the device couples light from all input fibres to all other fibres ($n \times n$ connectivity). Its capacity as a simple multiplexer is multiplied by the number of fibre ports. As already discussed this can lead to an enormous information throughput in a simple device. However, the detailed optical design of the multiplexer, with such high wavelength resolution requirements, is a challenging task. Fortunately, the coupling efficiency between wave guides need not be high since optical amplifiers can compensate for the losses.

Figure 11 illustrates the concept of zero switching within the UK. Customers are allocated a spare wavelength and time-slot from a local control unit, and thus gain access to long-haul fibres via a local PON splitter network and optical amplifier. Traffic is passively routed by the passive wavelength routing node (PWRN) to the required fibre to reach its destination town or city. Here it is broadcast over a downstream splitter network. Only the intended customer terminal is allowed to select and receive that wavelength and time-slot. In practice, several geographically separated PWRNs would be meshed to provide the necessary redundancy and reliability.

Final Remarks

Two views on the development of future networks have been presented in this article, both depend strongly on the number of wavelengths available. For a small number of wavelengths, the optical Lambda-PON offers a method of achieving a high degree of wavelength reuse.

For a large wavelength availability, the 'switchless' network should be achievable. To make the switchless network a reality, a whole range of additional systems and component issues relating to end-to-end optical transparency needs to be addressed. These include: accurate time synchronisation (ranging) and wavelength referencing, low-cost customer tunable transmitters and filter/receivers for GHz wavelength channel spacings, wavelength-flattened amplifiers, optical cross-talk, power level control (similar to today's PON systems), signalling requirements, path-searching and channel allocation algorithms, location and processing capacity of network control software. However, most of these elements exist in some form, or are receiving a deal of attention in research laboratories. The prospect is that all the vital components will be mature and come together between the years 2000 and 2005. At this point, radically different options will be available than those in evidence today to realise the ultimate telecommunications network.

In Part 2, to be published in the next edition of the *Journal*, ultra-fast self-routing packet-based networks will be addressed.

References

1 Cochrane, P. Future Directions In Long Haul Fibre Optic Systems. *British Telecom Technology Journal,* April 1990, **8**(2), p. 5.

2 Tandon, V., Wilby, M. and Todd, C. Wavelength Reuse In Future Passive Optical Networks. IEE Colloquium Transparent Optical Networks: Applications, Architectures and Technology. 22 April 1994.

3 VALIANI, M. The Fading Distinction Between LANS and WANS. *Telecommunications,* April 1993.

4 BEAUMONT, A. R., CADE, I. R., FFITCH, P. K., JENKINS, P. and PAYNE D. Passive Optical Network For Telecommunications. EFOC/LAN 1991, London, 19–21 June 1991, pp. 481–423.

5 ALBITT, D. J. and SENIOR, J. M. Performance Evaluation of a TPON Network. Fifth Bangor Symposium On Communications 2–3 June 1993, pp. 268–271.

6 CHRAPLYVY, A. R. Limitations on Lightwave Communications Imposed By Optical Fibre Non-linearities. *Journal of Lightwave Technology,* Oct. 1990, **8**(10).

7 AURA Ganz and ZAHAVA KOREN, WDM Passive Star—Protocols and Performance Analysis, IEEE Infocom 1991. The conference on Computer Communications, Bal Harbour, Florida, 9–11 April 1991.

8 JOHNSON, D., BROWN, G. N., BEGGS, S. L., BOTHAM, C. P., HAWKER, I., CHNG, R. S. K., SINCLAIR, M. C., and O'MAHONY, M. J. Distributed Restoration Strategies in Telecommunications Networks. ICC 1994.

9 FORRESTER, D. S. *et al,* 39·81 Gbit/s, 43·8 million-way WDM broadcast network with 527 km range. *Electron. Letts.,* 1991, **27**(22), pp. 2051–2052.

10 CHRAPLYVY, A. R. and TKACH, R. W. What is the actual capacity of single-mode fibers in amplified lightwave systems? *IEEE Photonics Technology Letters,* 1993, **5**(6), pp. 666–668.

11 HEALEY, P. *et al.* Photonic switching and interconnect for future network nodes. *BT Technology Journal,* Oct. 1991, **9**(4), pp. 19–29.

Biographies

Ian Hawker
BT Development and Procurement

Ian Hawker received B.Sc. and Ph.D. degrees from Leicester University in 1973 and 1976. He joined BT Laboratories in 1976 where he worked on the design of high-speed integrated circuits for TAT-8 submarine systems and for inland 565 Mbit/s transmission systems. He then undertook studies in design, control and quality-of-service issues for synchronous digital hierarchy (SDH) networks. He currently leads a team researching into future telecommunications networks including advanced network management issues and techno-economic modelling for SDH, ATM and optical networks.

David Cotter
BT Development and Procurement

David Cotter received B.Sc. and Ph.D. degrees from Southampton University in 1973 and 1976. He was a visiting researcher at the Max-Planck Institute for Biophysical Chemistry in Göttingen, Germany, during 1977–1980. In 1980, he joined BT Laboratories to research into the influence of optical nonlinearity in telecommunications systems and ultrafast optical nonlinearity in semiconductors, organic materials and glasses. His current interests are the development of radically new network solutions for the future using ultrafast photonic technologies. In 1993, he was elected a Fellow of the Institute of Physics.

Vivek Tandon
BT Development and Procurement

Vivek Tandon received his B.Sc. degree in Physics at Imperial College (University of London). He joined BT in June 1993 and is researching the development of future revolutionary packet and non-packet based optical networks. He is also pursuing a part-time Ph.D. at the University College of London developing a broadband wavelength-division-multiplexed passive optical network (Lambda-PON).

Alan Hill
BT Development and Procurement

Alan Hill received B.Sc. and Ph.D. degrees in Electronics from London University in 1973 and 1976 respectively. In 1973, he joined the Post Office Research Laboratories (now BT Laboratories). His career has covered a wide range of optical systems activities, including receiver design, laboratory and field testing, development of micro-optic components such as connectors, switches, tunable transmitters/receivers and wavelength-division multiplexers, studies of linear and non-linear optical crosstalk and the development of advanced optical switching and distribution networks. He has more recently been involved in the broadband upgrade of telephony passive optical networks (TPON/BPON) by WDM, including responsibility for BT Laboratories' activities in the RACE II MUNDI project for fibre-to-the home and building.

Michael Lyons and Michael Gell

Companies and Communication in the Next Century

The accelerating development of technology is placing increasing strains on traditionally structured organisations. The old vertically integrated industrial giants are experiencing great difficulty in operating, competing and surviving. To be successful in the next century, companies will have to alter drastically their structure and mode of operation. The speed with which this will occur will depend critically on the availability and cost of telecommunications services.

Introduction

The world is undergoing a major globalisation with a restructuring of the Fordist production process and a fracturing and dispersion of all the sub-processes across national boundaries. Realisation of an integrated structure will require communications systems to provide the 'glue' essential for global coordination. As a result, companies are under increasing pressure to rethink their operations to position themselves well in the new competitive era. This means a faster delivery of better products, of increasing sophistication, at lower cost, on demand. Common features of the evolving industries are:

- shorter product cycles—to provide a faster response to market and technological change; and
- increasing development costs for new products—carrying greater risks, but a potential for greater rewards.

As the competitive environment reduces the time to take action, faster decision making will be essential, requiring the intelligent processing of more data, under conditions of growing information pollution. All of this is underpinned by advances in communication and computer technology within legislative and policy frameworks that may allow or restrict such technology advantages.

Market Dynamics

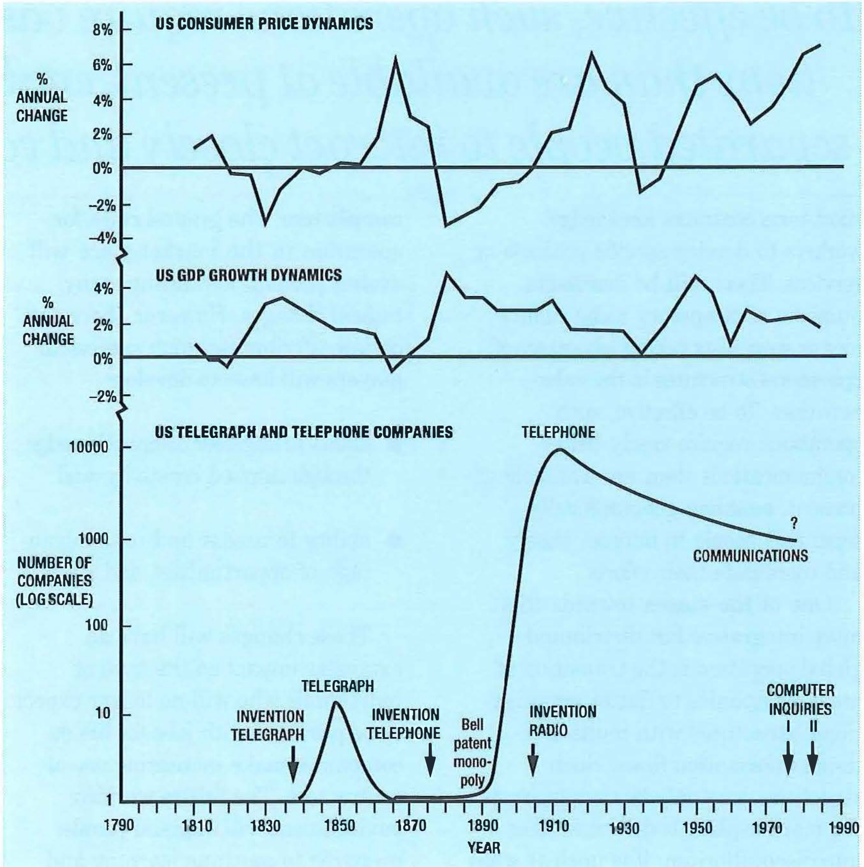
The US provides an interesting perspective on the interrelations between the evolution of communications and the economy. In Figure 1, consumer price, growth in gross domestic product (GDP) and communication company dynamics are shown for the past two centuries. Both the GDP and consumer price have exhibited broad 'cycles' of about 50 year duration, with a successive ratcheting between accelerating and decelerating growth/prices. During this evolution, the telegraph, telephone and radio have come into existence, leading to the generation of numerous companies exhibiting spectacular growth during periods of competitive evolution. The US is now entering a third communications age dominated, not by the telephone, but by a diversity of competing media. This new age of communications will be challenging not only from the point of view of technology but also from that of the US economy.

At the start of the century, the massive restructuring of the telephony sector was not able to arrest the decline in growth of the US economy (Figure 1). With the imminent burgeoning of the number of communications companies under a regime of intense competition and globalisation, the ways in which communications are harnessed to restructure economies into a mode of increased growth will prove challenging. Communications services will be critical for all sectors of the global economy.

Technology

With the imminent convergence of information technology (IT), entertain-

Figure 1—Time variations of US consumer price, GDP and communication dynamics



Adapted from: Berry, B. J. L. Long-Wave Rythms in Economic Development and Political Behaviour. John Hopkins University Press, 1991.

ment and telecommunications, many new opportunities are emerging:

- telepresence displacing travel;
- increasing telecommunications mobility;
- natural language recognition and machine input/output and conversation;
- automatic language translation;
- humanised interfaces—everyone will become computer literate;
- integrated terminals for work and pleasure—hifi, TV, virtual reality (VR), personal computers (PC), mobile IT;
- the supercomputer on your desk (wrist)—by 2010 the PC will be 1 000 000 times more powerful than that of today;
- computers you wear—communication and computing on the move;
- intelligent office with data sorting, visualisation, filtering, decision support, search—find—classify—anticipate—war games support; and
- mobile medicare—real-time monitoring of your health metrics.

Such advances will require the rethinking of organisational structure and processes.

Value Networks

The emergence of an integrated global economic structure will play havoc with established markets with global developments both dominating and driving national economies. Existing value chains will become less static and disordered value networks will take their place. Such networks will be dynamic, having many complex features, such as cross or floating links, reflecting the evolving market structure. These will be made even more complex through greater cultural mixing of business and management styles and the need for new forms of taxation for global companies. Company structure and operations will have to mirror this dynamic environment.

Communications will play an essential role in enabling companies to respond effectively in a dynamic, real-time economy in which the unifying managerial resource is information. The vertically integrated company, characteristic of much of current industry, will become less viable. One process leading to the demise of older structures is the inability of companies to retain strategic information as the convergence of industries necessitates increasing collaboration; already we see companies collaborating in one sector while competing (with different partners) in another.

Virtual Organisations

The new companies will be product- and service-based with different organisations contributing complementary skills. Unified companies could disappear and become primarily contracting organisations with manufacturing, research and development, marketing, sales and support run as separate profit centres, offering services competitively to parent or rival companies. Departments brought together to produce a specific product would form a 'virtual company' linked, not by geographic proximity, but by a global communication network that permits the low-cost transfer of ideas and knowledge. This mode of working is already evident in the software industry and echoes earlier developments in electronics manufacture.

As the virtual company becomes more common, we could see specialised companies dominating particular activities such as design, planning, marketing, or situations where companies, employing a small core of workers and wielding tremendous financial power, hire design, planning and manufacturing capacity, on a world basis, to create their own products. This could be particularly important as the 'means of production' shifts from the owner of production tools to the owner of the knowledge to control those tools—the information worker. In an even more extreme form, the 'virtual company' could employ, on

To be effective, such operations require vastly better communications than are available at present, enabling geographically separated people to interact closely and coordinate their efforts.

short-term contracts, knowledge workers to develop specific products or services. There will be increasing numbers of temporary niche companies or specialists taking advantage of ephemeral structures in the value networks. To be effective, such operations require vastly better communications than are available at present, enabling geographically separated people to interact closely and coordinate their efforts.

One of the stages towards this more integrative but distributed global operation is the transition of major companies to flatter organisational structures with multidirectional information flows. Such structures may only be temporary as the market-place is driven further into disequilibrium; it is unclear what types of organisational structure will be viable under such extreme conditions. It is also unclear what regulatory and legislative frameworks will best suit the requirements of this global economy, characterised by intense competition and widespread corporate restructuring.

Impact

As the global competitive process stimulates the atomisation of companies and institutions, counteracting cohesive processes such as cooperation between companies and nations, and increasing emphasis on business ethics, will be essential for maintaining global stability. Where businesses fail to meet these needs, they will be faced with increased environmental regulation and governmental control. Such cohesion can be enhanced through:

- more rational/democratic self-regulation of companies, and
- communication referenda and televoting.

This globally competitive environment, reliant on interworking communication networks, will depend critically on cooperation between competitors. The ground rules for operation in the market-place will evolve, perhaps exhibiting many radical changes. However, there are certain attributes which successful players will have to develop:

- ability to engineer competitive edge through applied creativity, and
- ability to assess and take advantage of opportunities and risks.

These changes will have an extensive impact on the lives of individuals who will no longer expect or be provided with jobs for life as companies make increasing use of contractors. The future working environment will demand people prepared to continue learning and developing throughout their lives. Individuals will have to develop their own strategies for coping with this demanding environment in which multi-tasking will become commonplace. Many people will become both employers and employees.

Communications Technology

The imminent deployment of novel optical technologies such as fibre amplifiers and soliton systems will change the course of telecommunications as we approach the next century, by stimulating radical and fundamental changes in systems and networks. The release of fibre bandwidth will bring about the realisation of the 'transmission engineer's dream': near infinite bandwidth; near zero physical space; near zero material requirement; near zero power consumption; all at near zero cost!

The key element to this dream is optical path transparency through the use of photonic amplifying devices (semiconductor and fibre) and photonic switching which overcomes the electronic traffic bottlenecks that would otherwise result as the enormous capacity of fibre releases the latent demand for communication. Radically new forms of telecommunication networks—well beyond the simple point-to-point systems that exist today—will be possible, with a progressive migration of intelligence, control and service creation towards the customer and terminal equipment. Further migration into the realm of non-linear optics for devices, fibre, systems and networks opens a further level of exploitation that is likely to eclipse the progress of telecommunications technology and services experienced so far.

Probably the most radical commercial changes in the 21st century resulting from the further advance of optical technology will be associated with the introduction of new services, the realisation that distance and bandwidth will be irrelevant—service and time will become the measures for charging, services migrating to the periphery of networks, local calls extending across complete countries and perhaps spanning Europe or North America, and ultimately the globe. The emergence of low-cost, high bandwidth telecommunications coupled with the convergence of the information and communications sectors will provide an opportunity to synthesise disparate technologies to create new environments in which people can live and work.

Future office

The office of the future was the subject of an earlier paper in this *Journal* ('The Office You Wish You Had'—see bibliography). It can be realised with available technology integrated to satisfy currently known and well defined requirements for human orientated interfaces. Features include:

- *desktop videoconferencing* featuring high-definition TV (HDTV) screens to produce life-size images and an enhanced sense of presence;
- *hands-in screen* interface, to allow manipulation of three-dimensional images in a shared work-space; and

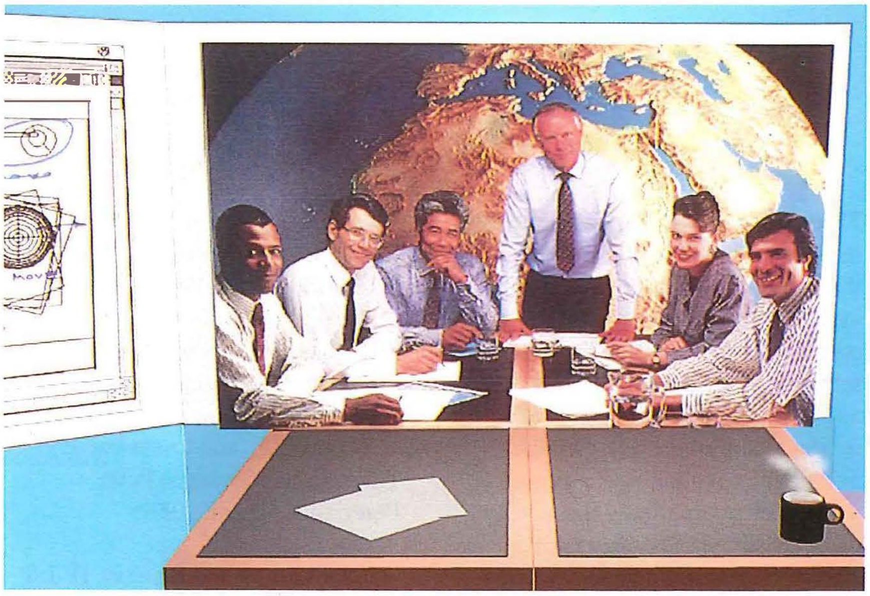


Figure 2—The virtual conference room—videowall and whiteboard

- *optical wireless* links to provide mobile broadband access within the office and eliminate complex and costly hard wiring.

Virtual conference rooms

Meetings involving several people will require a more formal environment than the desktop videoconferencing described above. Human behaviour in a real (single location) conference facility reveals a number of important requirements for maximising the usefulness of any meeting. To optimise the effectiveness of videoconferencing, this environment could be mimicked by providing a facsimile representation—a 'virtual' conference room.

Video windows

These are now commercially developed to a point where all the components are available. Ideally, human beings are presented in full size on a high-definition screen, as shown in Figure 2. By suitably arranging furniture and decor, the illusion of a continuous room or meeting place can be created. Moreover, by using electronic processing and steerable microphones, it is possible to focus the acoustics on any one speaker and arrange for the voice to emanate from the appropriate part of the image.

Electronic white board

This is currently lacking, or poorly realised, in teleconferencing. The solution is an electronic white board which allows people at remote sites to interact as if they were sharing the same board. Writing and drawings on one electronic white board appear on all other boards to which it is linked. Erasures by any of the users of linked electronic white boards are also possible. In one implementation, an image of the human being at the distant end is superimposed on the screen to increase the sense of personal interaction.

Instant fax

The ability to pass documents across the table during a meeting is essential for effective communication. In the tele-environment this may be realised in the manner shown in Figure 3 with a wideband (240 kbit/s) fax able to relay details effectively across the table instantaneously.

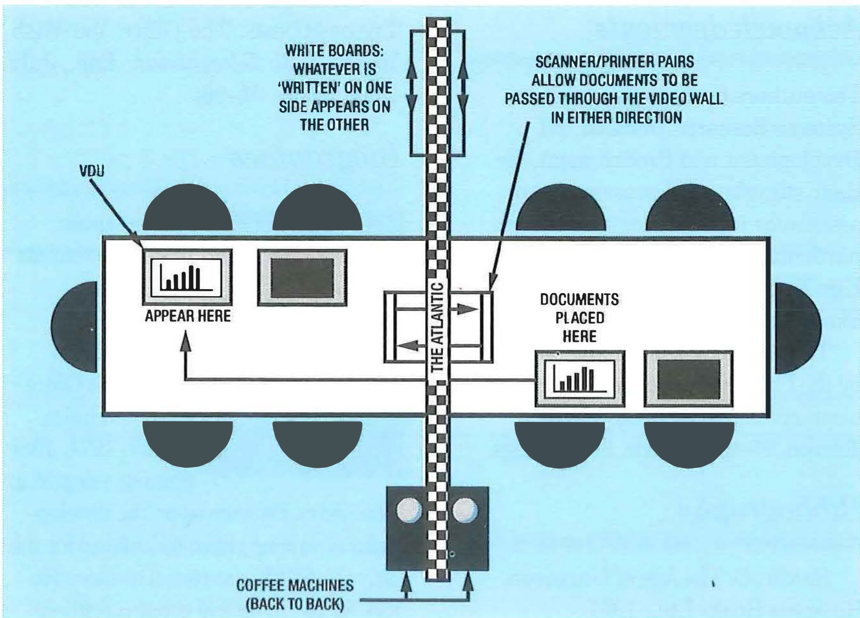
The telephone box of the future

Facilities of the kind described above will be costly at first and companies may be unwilling, or unable, to make the necessary investment. Consequently, there will be an opportunity for third parties to provide such facilities as a service. Obvious candidates would be hotels, who already provide seminar rooms and conference facilities for businesses. Just as in former times, when the majority of the population used a telephone box to make calls (because the cost of owning a telephone was too great), so local people and travellers will use the 'hotel-of-the-future' for telemeetings.

Virtual reality

The use of an extended version of immersive VR, with the mixing of real and computer generated images, may appear to offer the ultimate form of videoconferencing. However, the requirement to use headsets inhibits eye contact, thus creating unnatural interactions which may offset the advantages gained. Alternatively, lightweight and miniature elements such as active spectacles and contact lenses are already at the prototype stage.

Figure 3—Instant fax and video window schematic



Three-dimensional presentation
As we move into the 21st century, three-dimensional services can be expected to permeate into the computer-aided design and manufacture arena. Architects, car designers and aeronautical engineers have all indicated a need to 'walk through' new designs of buildings, cars and aeroplanes to experience the novelty and to examine the design merits. Three-dimensional imaging systems offer an attractive solution.

Conclusion

The world is going through a deep-seated transition which is impacting on all aspects of society. The major characteristics of the transition are a compression of time and distance coupled with the removal of delays through communications and computing technology. While the scale of the impact is principally evident on the open market process and on an individual basis, it is clear that all systems in society are having to adapt at great speed. However, the changes taking place can allow a new post-industrial society to emerge, but only if the opportunities offered by the widest spectrum of computing and communication technologies are allowed to flourish and are exploited effectively.

Acknowledgements

The authors thank colleagues in Systems Research Division, BT Development and Procurement, for their stimulating discussions and assistance in preparing this text; in particular, for the contributions of Kim Fisher, Keith Cameron, Sonia Skelton and Tom Lynch.

This article was originally published by the Canadian Business Telecommunications Alliance in the CBTA Telecon '93 Conference Proceedings.

Bibliography

HANDY, C. The Age of Unreason. Business Books Ltd., 1991.

DRUCKER, P. F. The New Society of Organizations. *Harvard Business Review.* Sept./Oct. 1992, pp. 95–104.

GILDER, G. Into the Telecosm. *ibid.* March/April 1991, pp. 150–161.

FLOOD, J. E. and COCHRANE, P. Transmission Systems. Peter Peregrinus for IEE, 1991.

COCHRANE, P. and HEATLEY, D. J. T. Optical Fibre Systems and Networks in the 21st Century. *INTERLINK 2000 Jnl.*, Feb. 1992, pp.150–154.

FISHER, K. *et al.* CAMNET—The First Telepresence System. *ibid.* Aug., 1992, p. 38.

COCHRANE, P. The Human-Machine Interface & Telepresence, *Euro Conferencing Jnl.*, Winter 1992, **1**(2), pp. 48–53.

CAMERON, K. H. *et al.* Telepresence—Visual Telecommunications Into The Next Century. 4th IEE Conference on Telecommunications, Manchester, April 1993, pp. 175–180.

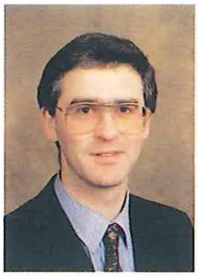
LYONS, M. H., Cochrane, P. and FISHER, K. Teleworking in the 21st Century. *Computing & Control Journal*, Aug. 1993, pp 170-180.

COCHRANE, P., FISHER, K. and TAYLOR-HENDRY. The Office You Wish You Had, *Br. Telecommun. Eng.*, July 1993, **12**, p. 91–96.

Biographies

Michael Lyons
BT Development and Procurement

Michael Lyons joined Post Office Telecommunications in 1974, after gaining a degree in Chemistry. He worked on the development of vapour phase techniques for the growth of III-V compound semiconductors for use in optical communications systems and was awarded a Ph.D. by London University for his work in this area. In 1982 he moved to the assessment section where he applied advanced X-ray diffraction techniques to the examination of semiconductor structures. In 1988, he started working on aspects of the future of telecommunications services including an assessment of the impact of new display technologies on visual telecommunications services and the use of telecommunications networks for environmental monitoring. He joined the Systems Research Division in BT Development and Procurement in 1991, where his work has included service and business modelling, and studying the impact on future society of advanced telecommunications services with a view to understanding possible usage patterns and traffic growth. He is a member of the Royal Society of Chemistry and is the author of over 30 scientific papers. He currently leads a group developing computer-based business modelling techniques.



Michael Gell
BT Development and Procurement

Michael Gell worked for seven years (1976-1983) with British Gas researching computer-controlled structures for real-time control systems. He then moved to Newcastle University where he carried out research in theoretical physics leading to the award of a Ph.D. in 1986. His research involved the development of theoretical and computational methods for quantum microstructures and was carried out in collaboration with teams at RSRE Malvern and IBM Research Centre, New York. Since moving to BT Laboratories in 1986, he has researched quantum silicon microstructures and, over the last four years, future communication systems. He is an editor of *European Transactions on Telecommunications* and holds a visiting Professorship at Staffordshire University.

Shara Amin, José-Luis Fernández-Villacañas and Peter Cochrane

A Natural Solution to the Travelling Salesman Problem

The principles of artificial life, self-organisation and the survival of the fittest can be used to solve a generic set of problems classified as 'The Travelling Salesman Problem'. Simple algorithms can search out and find the shortest tour length option between thousands of cities in a dramatically reduced computing time compared to previously published solutions. The technique can be extended into hyperspace to give real-time solutions for routing in mobile communication, network and circuit design or investment analysis.

Introduction

What is the most efficient route for engineers to repair reported faults or freight companies to deliver cargo to a fixed number of known locations? These are but two of many examples of the Travelling Salesman Problem (TSP), which, while apparently simple, is in fact deceptively complex and one of the hardest optimisation problems to solve[1]. During the past 50 years, many have tried to find efficient algorithms that provide good solutions[2]. Optimising the tour length when visiting a small number of locations is straightforward, and sometimes human capabilities can fulfil this task efficiently. However, the number of possible routes increases exponentially with the number of locations. For example; a 50 city-tour presents 10^{64} different possible routes. Even a supercomputer would take years to search every possible route and identify the solution that gives the least distance travelled.

The importance of the TSP problem stems from the wealth of applications areas where it can be applied. These range from pure mathematics to all aspects of commerce, industry, computing and telecommunications. In fact, it relates to any search and positioning system, from drilling a series of holes in a printed circuit board, to planning cable duct routes and searching electronic data libraries. The quest for an efficient solution has therefore been a target of considerable research effort worldwide for several decades. This article describes a new and novel approach.

Because it is generally thought that no exact solution can be found for very large TSP problems in a reasonable time, many have tried to construct algorithms that compute good approximations. However, for very large numbers of cities the problem is still difficult to solve and generally requires massive amounts of computing time. For example; the Lin and Kernighan[2,4] algorithm gives good results for a reasonable number of cities and recent attempts have used elastic nets[5], simulated annealing[6], neural and Kohonen-type networks[7,8,9]. Some of these approaches provide a good solution for a reasonable number of cities, but large problems remain intractable and particularly difficult to solve in real time.

Recently reported results have seen a new world record optimal solution for 3038 cities requiring about one and a half years of supercomputer time[3]. In contrast, the algorithm reported here converges to within 4% of this optimum solution within **25** minutes using a standard workstation.

A New Approach

The new approach combines a number of ideas from Newtonian mechanics and Kohonen-type self-organisation networks[10,5,8] with concepts from Darwinian evolution to create a new solution. To illustrate the power of this algorithm, a computer simulation has been developed for a random distribution of cities on a map of Europe (Figure 1). The aim is to establish the shortest path starting from one city, visiting each city once and returning to the departure point.

The assumed configuration is that of a two-dimensional distribution of x and y coordinates representing real

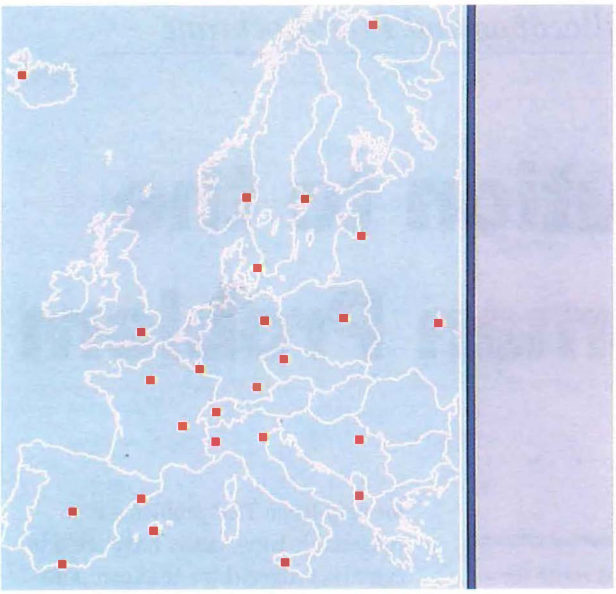


Figure 1—The Travelling Salesman Problem—how to visit 25 cities?

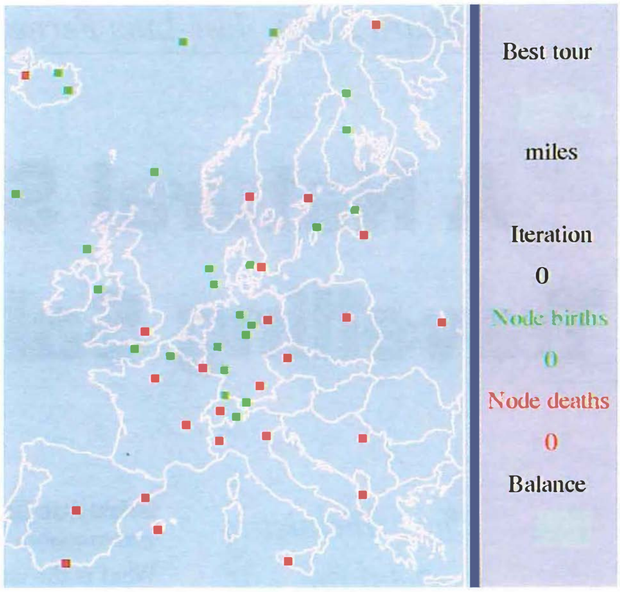
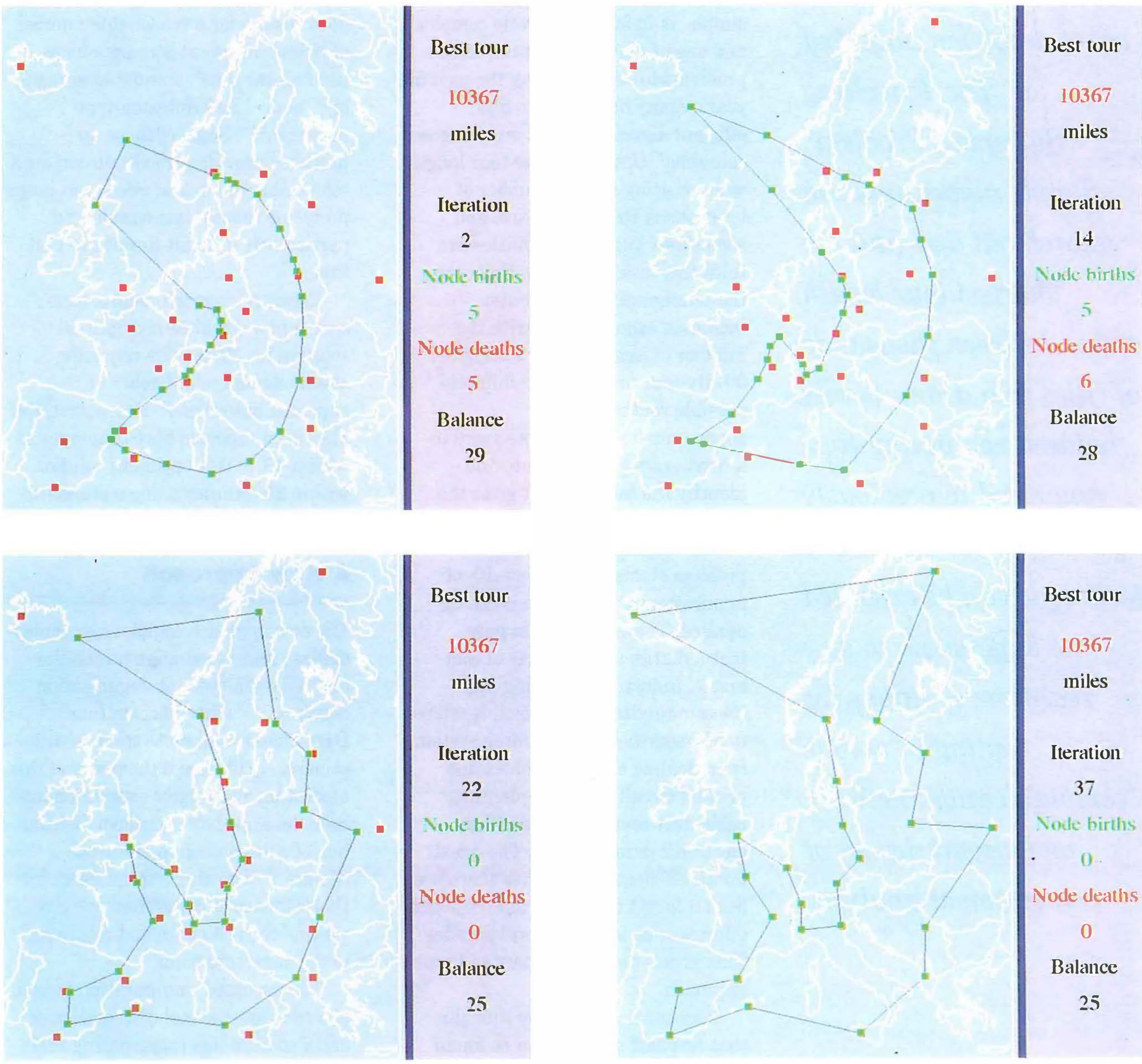


Figure 2—The ring of virtual nodes tries to match the 25 cities

Figure 3—The iteration solution to the 25-city problem. Top left—2 iterations; top right—14 iterations; bottom left—22 iterations; bottom right—37 iterations, final solution



cities. In general, it could be *n*-dimensional, depending on how many parameters have to be optimised. An equal number of virtual nodes (number of virtual nodes = number of cities) is randomly scattered in the same city coordinate space. This set of virtual nodes is ordered and constitutes the ring of green nodes (Figure 2) which is constantly being stretched. This movement is produced by Newtonian forces as each (red) city attracts the virtual (green) nodes. Every city competes for the virtual nodes as each is attracted to the strongest force. Each iteration is characterised by a temperature value, *T*, which, in thermodynamic terms, is a measure of agitation or random motion. At first the temperature is high, movement and the attraction is maximal. As the system converges the temperature cools and the attraction reduces. The system iterates until the ring of virtual nodes is stretched to coincide with the distribution of the cities[11]. This process is illustrated in the sample sequence of Figure 3.

Local Self-Organisation

The random movement of the virtual nodes, and their neighbours, preserves the neighbourhood structure and results in the system attempting shorter tours. This is caused by what is known as *local self-organisation*. For instance, as we drive a car we are only concerned with factors and events in our immediate vicinity—we do not worry about traffic in Manchester when driving in London! There is therefore a local set of conditions and rules considered by every driver. It is the combination of such single, small constituents that produces global traffic behaviour such as fluent traffic, jams, and waves. The TSP algorithm follows a similar pattern with localised actions attracting the closest virtual node to individual cities. The stretching is sensed by the nearby virtual nodes that are also being attracted by other cities. This results in a progressive and rapid optimisation as depicted in the sampled series shown in Figure 3.

The Survival of the Fittest

If, in any one cycle, the same virtual node is chosen as a winner by different cities then a similar virtual node with the same property is created (duplicated). Duplication occurs each time the same virtual node is chosen. The newly born virtual node is not allowed to give birth during the current cycle; however, it is allowed to move freely should any of its neighbours be chosen as a winner. In the next cycle, the new node is allowed to behave as an adult (that is, it can give birth and move freely if required). If newly born virtual nodes were allowed to replicate immediately after creation, the performance of the approach would deteriorate by up to 20% as they have to mature, like other life forms, to become effective!

By this process of creation, the number of virtual nodes is increased, but the final population of virtual nodes has to match the number of cities and there has to be redundancy or death. The 'reaper' mechanism is performed in the following manner: if a virtual node is not chosen as a winner in the current cycle, it is given a chance in the next cycle, and if not chosen, is terminated. The birth and death processes of the virtual nodes leads to the survival of the fittest.

This biological process ultimately drives the network towards a shorter tour and eventually to the final solution (Figure 3).

Sample Results

A large series of simulations have been completed on a variety of city configurations and configuration sizes. The cities were randomly distributed in a unit square box with a uniform distribution (geometric TSP). For such a case an estimate of the tour length is offered by Stein[12]:

$$L = 0{\cdot}765\sqrt{(NA)},$$

where N is the number of cities and A is the area of the square. This formula was used to obtain an estimate of the performance of the self-organising network. Tour lengths of random configurations of up to 30 000 nodes have been completed using the new algorithm. The deviation of results (Figure 4) from the estimated optimum derived using Stein's formula for the tour is uniformly less than 4%. Moreover, this deviation can be set to an arbitrary accuracy by parameter adjustment at the expense of computation time.

Another important consideration is that the approach is, in most cases,

Figure 4—Estimated performance compared to empirical results[12]

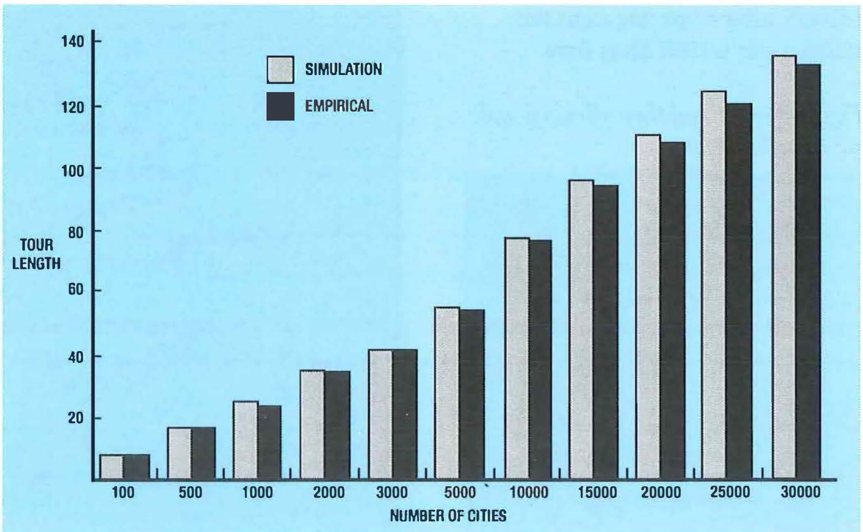
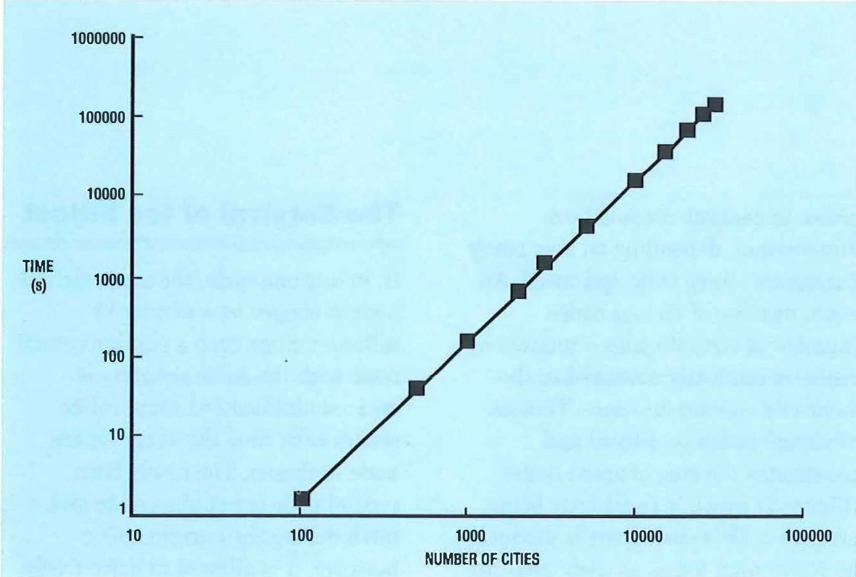


Figure 5—Size of the problem versus Sparc 2 computer run time

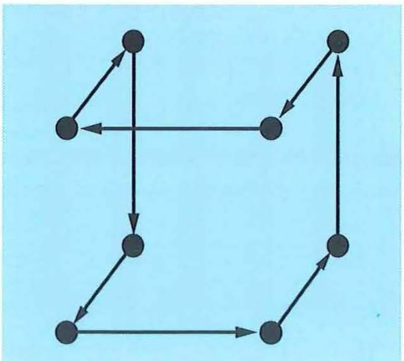


wholly insensitive to the choice of temperature and gradient. The same values apply when analysing a 10-city problem or a configuration with 30 000 (Reference 11). As shown in Figure 5, the rate of increase in computation time, with the number of cities, compares favourably ($t \propto N^2$) with Lin and Kernighan's[4] approach with $t \propto N^{2.2}$.

To obtain a comparable performance evaluation with alternative methods, sets of random cities were generated in the range 20–3000. In all cases, the method[11] outperformed all reported contenders including simulated annealing[6,13] which gets worse as the number of cities increases; in contrast, the new approach demonstrates a constant error.

It is also worth mentioning that the simulation time for the 3000-city set was 25 minutes by the new method and 3 hours and 10 minutes by a simulated algorithm on the same machine with the same compiler. Results were also tested against well-known, solved problems from the global TSP library (LIN105, Bier127, LIN318 and ATT532) for which optimal solutions are known[14]. Also, comparable results were achieved with fewer cooling steps. For example: optimal solutions for the Hopfield 10- and 30-cities problem[7] were obtained within 5 and 10 cooling steps respectively. Similar results were also obtained by using the simulated annealing technique. In the case of LIN105, after 47 cooling steps the solution was 0·188% away from optimal, while with simulated annealing it was 0·3481% away. The solution for ATT532 was obtained within 48 cooling steps and less than 1·678% from optimal, while the simulated annealing was 4·56% away. It is clear that the self-organisation approach gives better solutions in all the cases.

Figure 6—Optimal three-dimensional solution for a cube



Further Extensions

Studies of the basic algorithm have been expanded to include:

- *Multi-dimensional TSP* Where there is more than one parameter to optimise, for instance, cost, time, distance, etc. The optimal solution for a simple three-dimensional case found by the network is shown in Figure 6.
- *Multi travelling salesmen* The problem of a fixed number of engineers allocated to a large number of repair tasks has been solved for the non-constrained case (Figure 7).
- *Not-fully connected TSP* So far the approach has assumed full connectivity (airline-type network). In a road-like network where all routes are not possible, the best route can be found from isolated points A to B.

Hyperspace

A clear, and obvious, exposition of the potential problem complexity in this context is perhaps the threat scenario for a fighter pilot. It contains all the

Figure 7—The multiple travelling salesman problem. Optimised routes of four people from 'headquarters'

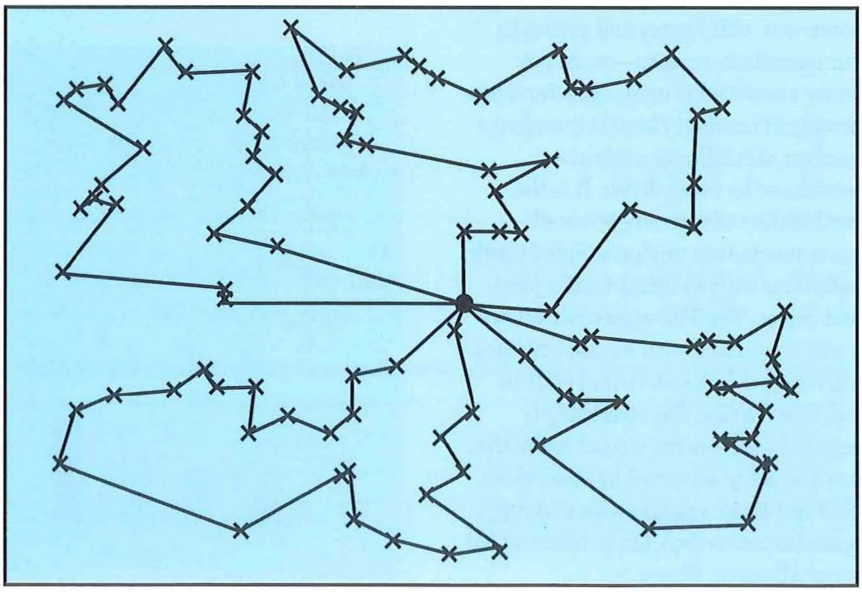


Figure 8—The fighter aircraft hyperspace problem

elements of spatial, temporal and stochasic space, plus a high degree of non-linearity and uncertainty well beyond any human ability to analyse in real time. Moreover, it also has an all-too-obvious death penalty!

Consider a squadron of fighters (Figure 8) at 30 000 feet under simultaneous attack from a ground-to-air missile battery, plus enemy interceptors approaching from two different directions, at different heights and carrying different, and perhaps unknown, weapon systems. As the squadron leader, what action do you take? Who would you assign to which threat, and in which order, to maximise the chance of survival? This is a non-linear hyperspace problem that can be tackled by an efficient TSP algorithm. It is also indicative of the scale of complexity we face in the future telecommunications environment.

The future of the industry will see more mobility, information, service diversity, competitors, technology change, and, perhaps most challenging, an increase in the number of concatenated carriers, service providers and terminal equipment types. On top of all this will be a changing regulatory framework, working practices and economic environment. It is therefore essential that solutions to the joint optimisation of such levels of complexity are found. The TSP in hyperspace looks like a viable contender in this context.

The Need and Future Applications

Exponential growth in every aspect of telecommunications networks, service and application provision makes it extremely unlikely that established engineering solutions will cope. As mobility is extended to tens of millions of users in the UK, and hundreds of millions worldwide, the search and locate, routing and management systems will require new and faster techniques. The same is true for billing systems and search algorithms for future electronic libraries. In each case, the customer tolerance to delay will predicate a change to new algorithms. The TSP algorithm described in this article is one contender able to provide a simple but very effective solution. It does this with a small sacrifice in accuracy that can be predefined. In many applications, such as mobile communications, the input data can be expected to be less than 50% accurate, and a small error (< 4%) from the optimum solution is thus generally acceptable in practice. This is certainly the case for mobile systems control and other stochastic databases where speed of response is paramount.

In the case of work allocation for installers and repair crews, or stand-by plant and spares holdings, the impact can also be expected to be extremely beneficial. But perhaps the most exciting application area is in the solution of non-linear hyperspace problems that defy human conceptualisation and understanding. This class would also include the optimisation of investment profiles, network and system design and the forecasting of future customer and traffic activity. To date, the algorithm outlined in this article represents an efficient technique available for such activities.

References

1 GAREY, M. R. and JOHNSON, D. S. *Computers and Intractability.* W. H. Freeman, New York, 1979.

2 LAWLER, E. L., LENSTRA, J. K, RINNOOY, A. H. G., and SHMOYS, D. B. *The Travelling Salesman Problem.* John Wiley & Sons Ltd, 1985.

3 Shortcircuiting the Travelling Salesman Problem. *New Scientist*, 27 June 1992.

4 LIN, S. and KERNIGHAN, B. W. An Effective Heuristic Algorithm for the Travelling Salesman Problem. *Operations Research*, 1971, **21**, p. 516.

5 DURBIN, R. and WILLSHAW. An Analogue Approach to the Travelling Salesman Problem Using an Elastic Net Method. *Nature*, 1987, **326**, pp. 689–691.

6 KIRKPATRICK, S., GELATT JR., C. D. and VECCHI, M. P. Optimisation by Simulated Annealing. *Science*, 1983, **220**, p. 671.

7 HOPFIELD, J. J. and TANK, D. W. Neural Computation of Decisions in Optimisation Problems. *Biological Cybernetics.*, 1985, **52**, p. 141.

8 ANGENIOL, B., LA CROIX VANBOIS, G. and LE TEXIER, JEAN-YVES. A Self-Organising Feature Maps and the Travelling Salesman Problem. *Neural Networks*, 1988, **1**, pp. 289–293.

9 FAVATA, F. and WALKER, R. A Study of the Applications of Kohonen-Type Neural Networks of the Travelling Salesman Problem. *Biological Cybernetics*, 1991, **64**, p. 463.

10 KOHONEN, T., Self-Organised Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 1982, **43**, p. 59.

11 AMIN, S., A Self-Organised Travelling Salesman. *Neural Computing and Applications,* 1994 (to appear).

12 STEIN, D. Ph.D. thesis, Harvard University, Cambridge, MA, 1977.

13 PRESS, W. H., FLANNERY, B.P., TEUKOLSKY, S. T. and VETTERLING, W. T., *Numerical Recipes in C.* Cambridge University Press, 1988.

14 TSPLIB is a library of travelling salesman and related problem instances. All data is publicly available from either Robert E. Bixby of Rice University (E-mail: softlib@rice.edu) or Gerhard Reinelt from Institut fuer Angewandte Mathematik (E-Mail: bixby@rice.edu).

Biographies

Shara Amin
BT Development and Procurement

Shara Amin graduated from Edinburgh University in 1988 with a Ph.D. degree in Condensed Matter Physics. He then joined the assessment section in BT where he developed advanced X-ray diffraction methods for studying semiconductor materials. He joined the Systems Research Division in 1991 and is currently working on optimisation problems. He has published extensively, among his awards being the first prize for a paper in a crystallographic conference (Oxford, 1990), and divisional and departmental prizes for his work on semiconductor materials.

José-Luis Fernández-Villacañas
BT Development and Procurement

José-Luis Fernández-Villacañas Martín graduated in Physics from Madrid University in 1984. He obtaining an M.Sc. and a Ph.D. in Astrophysics for his work on nucleosynthesis of heavy elements in the atmospheres of cool supergiant stars. In 1989, he became a member of staff in the Theoretical Physics Department, Oxford University, where he worked on Plasma Physics and radiation transfer in stars. He joined the Systems Research Division in January 1992 to research on artificial life systems. Currently, he is involved in optimisation problems and biological models for market and navigation systems in the network.

Peter Cochrane
BT Development and Procurement

Peter Cochrane graduated from Trent Polytechnic with a B.Sc. Honours degree in Electrical Engineering in 1973 and gained an M.Sc., Ph.D. and D.Sc. in telecommunications systems from the University of Essex in 1976, 1979 and 1993 respectively. He is a fellow of both the IEE and IEEE, a visiting professor to Essex and Southampton Universities and an honorary professor at Kent. He joined BT Laboratories in 1973 and has worked on both analogue and digital systems. He has been a consultant to numerous international companies on projects concerned with systems, networks and test equipment. In 1978 he became manager of the Long Lines Division and received the Queen's Award for Technology in 1990 for the production of optical receivers for TAT-8 and the PTAT-1 undersea cable systems. In 1991, he was appointed to head the Systems Research Division at BT Laboratories which is concerned with future advanced media, computing and communications developments. During 1993 he was promoted to head the Research Department at BT Laboratories with 620 staff dedicated to the study of future technologies, systems, networks and services. He is also the Development & Procurement board member for technology.

Tony Fowler and John Swingewood

BT's Service Management Centre for Broadcasters

The penalty for breaks in broadcast transmission can be severe in terms of lost revenues for the broadcaster and the network operator. A total of 30 seconds break in service per year might incur penalties of £2 million or more on the network operator. Building a system to manage a network to carry this high-value traffic requires not just technical but also organisational skills. BT has developed a management system for a nationwide network for carrying TV and radio programmes from the producers to their terrestrial or space-borne transmitters to improve customer service and revenue-earning potential of the network as competition intensifies.

Introduction

Distributing television and radio broadcast signals around the UK to transmitter sites has been the responsibility of BT and its predecessors for as long as there have been national broadcasting services. This has been an exacting task, since broadcasters and their viewers and listeners will not tolerate lengthy or frequent breaks in service. In recent years, both BT's business and that of the broadcasters have been marked by the rapid spread of competition in what were previously monopoly markets. Allied to this increasingly competitive marketplace has been the transition over the last decade to 24-hour broadcasting.

For the broadcasters, competitive pressure is for them to increase audience share and advertising revenues. In recent years, it has become fierce, with the arrival of satellite television direct to homes, and the building of cable television networks. At the same time, the licensing by the authorities of national commercial stations means that radio broadcasters are subject to similar pressures.

Alongside the rapid growth of entertainment broadcasting, the wider availability of the appropriate technology, together with liberalisation and deregulation, has generated a lucrative market for business broadcasting. Many large companies now want to operate their own services to keep their branches and dealers informed of business decisions and the latest market conditions. In many instances these organisations have also found that satellite communications and the rapidly spreading broadband infrastructure in the UK are also ideal for their data communications needs. Now they find that all along they have been travellers on telecommunications super highways that other nations are only contemplating constructing.

At the same time, deregulation and liberalisation have long since seen BT's monopoly whittled away, with rival telecommunications companies being licensed to provide the long-distance broadband services that customers now require. Even now, at least four high-capacity networks with broadband backbones are coming into service on a national scale, while a myriad of new operators are building local systems, and several more are building on a regional scale.

For a telecommunications company, broadcasting and data communications customers can mean extremely large revenues, with some used to spending tens of millions of pounds a year on their communications links. Simply providing cables and switches is not enough for the modern market. The networks must be both resilient and managed to provide 'no-break' services. The 24-hour service demands also mean that broadcast circuits and components cannot be withdrawn from service for routine maintenance.

In particular, the television companies demand practically 100% reliability. If there is a break in transmission longer than 10 seconds, then their viewers will use their remote controls to flip to another channel. So, for example, one of BT's prime customers, Channel 4, stipulates that no more than 30 seconds a year viewing time can be lost as a result of network failure. That is a level of service that

Figure 1 – SMC control room

BT Broadcast and Satellite Services (BSS) are happy to guarantee.

The Service Management Centre

Until the beginning of 1993, BT's broadcast customers did not have a dedicated service centre. Services were managed from a variety of fault reporting points all over the UK. This led to customers being given varying levels of services and responses dependent entirely on people at the various reporting points. In fact, BSS customers were serviced by some 60 different fault reporting points. It was difficult to deliver a consistent level of service to customers and virtually impossible for the management of BSS to have any idea of the level of service customers were being provided.

The technology employed to deliver broadcast services in the UK, while being highly labour intensive, consistently achieved a network availability of 99·998%. A clear target for BSS was therefore to replicate or better the traditional availability performance while reducing the cost base.

As a first step, it was decided that a service management centre (SMC) similar in concept to the service centres provided for telephony customers but dedicated to broadcast customers should be established with the following goals:

- to provide the customer interface for all broadcast customers;
- to provide management information systems enabling BSS management to be up to date and proactive when dealing with customers; and
- to house the new management systems to be deployed for imminent Channel 4 television distribution network and future broadcast networks.

A pilot 24-hour service point was established in April 1992. The first two customers were The Associated Press and Hutchison Paging, both of which have a satellite distribution service for their wire and paging services respectively. This pilot service also provided broadcast customers with an escalation point in the event of any unresolved problems on broadcast services.

When BSS moved to new accommodation in Vision House, at Euston in London, the SMC went into full operation in interim accommodation. In January 1993, the first managed network went live; this was the 34 Mbit/s digitally compressed national TV distribution network for Channel 4 Television. Systems deployed to support this network were based around the ServiceView network management platform from BT Development and Procurement.

While this operation continued, custom accommodation was being prepared in Colombo House near Waterloo, London. In February 1993, the SMC operation moved into Colombo House. Here the accommodation supports up to 20 network management personnel and their systems. A video wallboard displays the status of the managed networks that now include the Classic FM national radio service and will shortly include the Independent Television network contribution service, the Meridian TV and Anglia TV local TV networks. (See Figure 1.)

The main centre for broadcast TV and radio services in the UK continues to be the BT Tower, which is now undergoing a programme of modernisation and change. Major elements of the centre are being replaced, including a 26 year old Marconi mechanical switcher. The BT Tower is the largest switching centre of its kind in the world with three broadcast routers capable of switching more than 3000 TV and radio services.

In 1994, it is planned to integrate the SMC with the BT Tower operation. These locations will share common but resilient computer systems, communication systems and data networks. This will enable personnel from either location to access any of the information systems and to control all switching and network management systems recently deployed in the UK network.

While the SMC and BT Tower operate as the customer interface on broadcast networks, physical network maintenance and the repair of faulty circuit elements are devolved to BT Worldwide Networks.

To improve the reliability of the broadcast network, BSS has over the past 18 months replaced outdated mechanical switching equipment with state-of-the-art studio-quality routers. This has reduced the dependence on manual intervention for maintenance and switching activities. This national network of routers is controlled and monitored by technical staff at BT Tower. BSS believes that a major criterion for success is customer service. It is essential that BT is as proactive as possible in dealing with customers' services. The goal is to know of a problem on a managed network and resolve it before there is any disrup-

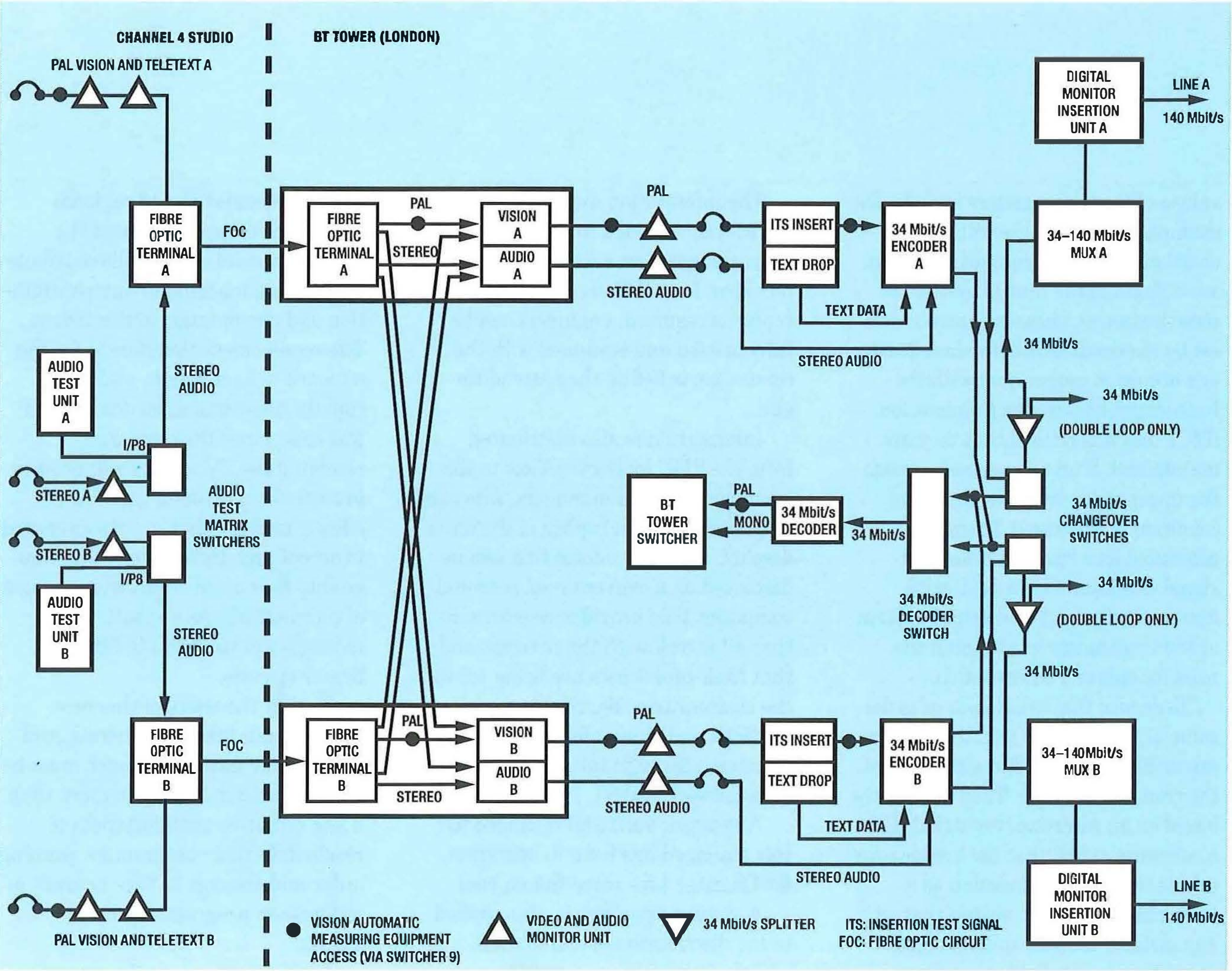


Figure 2—Channel 4 distribution arrangements (one region)

tion to service transparent to the user. On non-managed services, the target is to be aware of a problem as soon as a customer is and to mobilise resource to correct the fault whilst appraising the customer of fault progress. A design aim is to provide all future services with an element of network management.

There are a number of major elements to be addressed to achieve this goal:

- the configuration of the network to provide resilience;
- the use of specially developed sensors to monitor, measure and report on the performance of the network;
- the information technology to ensure that the information generated by the network about itself can be interpreted and applied; and
- an interface with customers that makes them aware of the quality of service they are receiving and reassures them that all is well at all times.

The Network

In practice, each customer requires a network to be constructed especially to meet their requirements, and that built for Channel 4 illustrates the approach taken by BSS.

The task is to feed the station's output to some 27 transmitter sites throughout the country. These are separated into six regional groups, and, apart from London, where transmissions are radiated from a single transmitter at the Crystal Palace site, each region includes several transmitters. Signals are carried to each region over a dedicated 34 Mbit/s bearer circuit, configured as continuous loops. This configuration provides two signal feeds to each transmitter site. (See Figure 2.)

The TV signal is transmitted around each loop in both directions. At the transmitter site, signals from each direction are continuously monitored and compared, with the highest-quality one being fed to the transmitter. In the event of one signal becoming degraded to the point that it is unsuitable for broadcasting, an intelligent control system, specially developed by BT Development and Procurement at Martlesham Heath Laboratories, can change to the alternate feed in less than 500 ms. This is the longest break in a broadcast that might be experienced should one of the circuits fail completely.

TV signal parameters are monitored by probes at various points on each circuit. These monitor up to 100 parameters that define the quality of the video, audio, NICAM and teletext components that make up the signal to be broadcast. Maximum and minimum values of these parameters include, for example, signal-to-noise ratio, differential gain, chroma content and delay are defined in the control system. In some instances, these parameters are set by the conditions of the broadcaster's licence in cooperation with the Independent Television Commission (ITC), and it is critical that they are maintained. If any parameter exceeds the upper or lower limits, an alarm condition is generated. Alarms generated as a result of broadcast signal degradation are dealt with automatically by the switching system at the transmitter site without the need for operator intervention.

To ensure that 'what goes in' is the same as 'what comes out', the signals are monitored by BSS at either end of the transmission line. They are mostly based on an insertion test signal (ITS), a reference signal that the broadcaster injects into the transmission as it leaves the studios. To ensure that, at any instant, the best-quality signal possible is forwarded for broadcast, each site is carefully programmed according to a decision table that takes into account permutations of the quality of the signal components from both feeds. For example, it may be that the audio on one feed is better than the other although the video is threatening to vary outside its limits.

The results of the measurements from both the TV studio and the remote transmitters are collected and returned to the SMC at Colombo House, via a GNS X.25 data network. The X.25 data communications link is provided independently of the main transmission line. At the SMC, returned data is collated for display in graphical form on ServiceView workstations. This allows alarms to be analysed rapidly in detail, and provides a very detailed historical report of the service levels provided. At the same time, the data link provides detailed telemetry information on the state of BSS equipment on the sites. In most instances, a potential fault is spotted and corrective action taken long before the customer is aware of it.

The information available is also sufficiently detailed to determine the exact nature of an actual or potential technical fault. Where a hardware repair is required, engineers can be fully briefed and equipped with the correct parts before they attend the site.

Information is also distributed from the SMC by ServiceView to the customers' station managers, who can be provided with a replica of the main display. This is in a form that can be displayed on a conventional personal computer. This provides reassurance that all is well with the network and that their broadcasts are being fed to the transmitters. ServiceView also provides real-time information to BSS managers through the in-house local area network (LAN).

A network built and managed to this standard has been in operation for Channel 4 for more than a year.

A similar approach is also applied to the distribution of stereo radio signals for another major BSS customer, Classic FM. This approach will also be adopted for other applications, including data services and both commercial and business broadcasting using satellite systems.

While the bandwidth requirements of the networks built and managed to date varies—Channel 4 carries compressed TV signals at 34 Mbit/s, Classic FM carries uncompressed audio at 2 Mbit/s, and private systems for companies such as Ford and TSB have their own special requirements—they have one thing in common, they are all distribution networks. This means that the signals they transport originate from a central point and are distributed to either single or multiple transmission sites.

A more complex problem to be addressed is a new network under construction for the Independent Television Association (ITV) which will provide a contribution service. (A contribution service is for example Coronation Street from Granada TV which is transmitted by all ITV companies.) In this case, the network will interconnect the 15 regional television companies around the country, each of which will contribute programme material to any permutation and combination of the others. The requirement therefore is for the network to be dynamic, and to be rapidly reconfigured on demand. In this case, since the companies receiving the TV signals will need to process them through their own editing studios, they are not prepared to accept any degradation of picture quality that could result from the use of compression. As a result, this network will use full 140 Mbit/s bearer circuits.

So that the users of this new network can take full advantage of its dynamic nature, changes must be effected with delays of no more than a few minutes, enabling them to reschedule their programme running order and timings to take account of unforeseen program and news events.

A prime example is the case of a sports programme that is cancelled at a few hours' notice because of weather or one that runs over time. In either case, scheduled programmes will need to be replaced with precisely timed substitutes sourced either from the network or from in-house resources. In the case of the latter, the company with a programme up its sleeve may wish to offer it to other broadcasting stations on the network, some of whom will want it, others will not. Another case, where split-second decisions and timings may be even more critical, is where a news story erupts that is important to another region and needs to be fed from one station to another. The distribution and contribution of regional or national advertising videos also adds to the need for flexibility.

To allow this degree of flexibility, BSS has adopted the novel solution of providing the companies with a measure of self-programming capability. Using a PC, a company will be able to revise schedules, and broadcast its changes to the network management centre and to the other

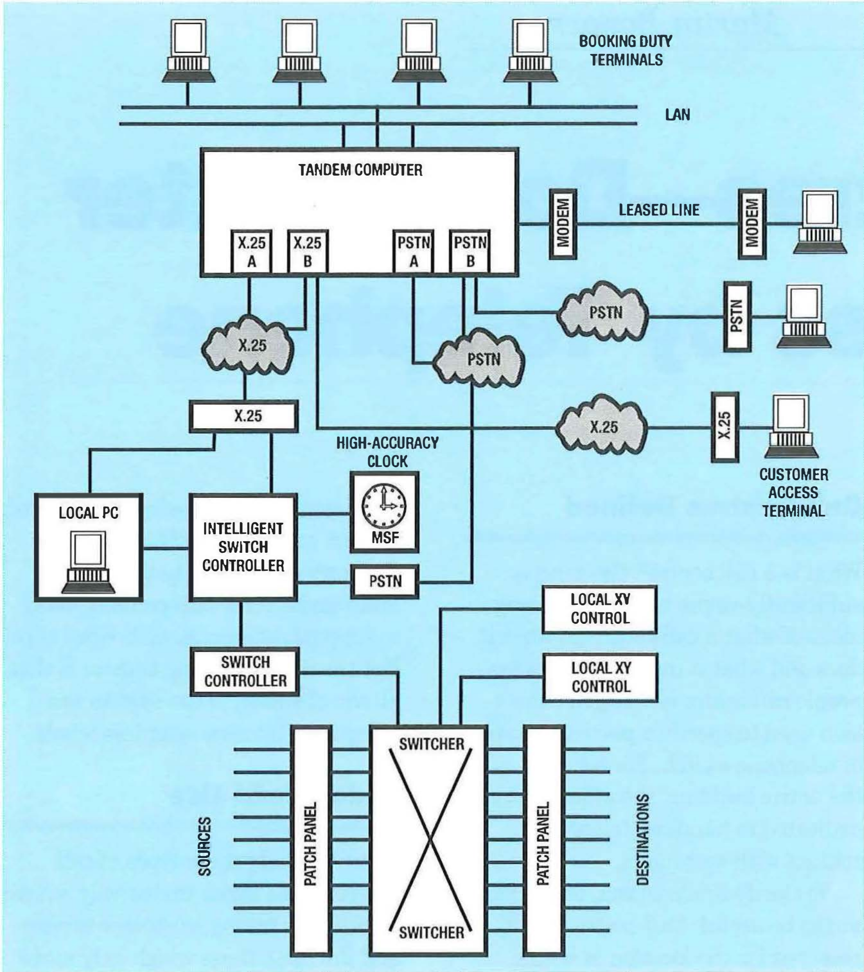


Figure 3 – National TV switching control system

companies. In general, the network changes requested will be actioned with no apparent delay. However, network configuration changes will not usually be made to the 'live' network, but will be set up in parallel, switching over at the precise time required. (See Figure 3.)

The 15 companies involved regard themselves as a 'federation' where no one is more significant than any of the others in its ability to impose its requirements. So overall control of the network may ultimately reside with BSS as a truly neutral network manager. This may well place a significant emphasis on recording all network events. In the past, the companies involved have been used to working on a network where changes were made by hand by using patch-panel techniques. The ability to make near instant changes may bring about major changes in their methods of working. One of the anticipated changes is likely to be increasing 'occasional' or unscheduled use of the network over and above the time or capacity each company has contracted for. Thus it will generate incremental revenues.

Conclusion

Whatever the complexity, BSS has, over an intensive two-year period, combined development with commercial success and gained total confidence in its technical and management systems. This confidence is such that the organisation can now stake more than reputation on its ability to provide its customers with precisely the service they need, offering compensation to customers if its quality of service falls down. In the case of the Channel 4 network, a total collapse of the network would cost BT some £2 million. At the same time, this level of service is not cheap, and customers will pay a premium for such a high grade of service.

However, the BSS managed broadband network service is not just for the giants. As the organisation and its reputation grows, its service will become increasingly attractive to smaller organisations at a time when the demands for multimedia and 'multi-stream' broadband communications services are poised to accelerate down the slip roads of the super data highways.

Biographies

Tony Fowler
BT Broadcast and Satellite Services

Tony Fowler joined BT as an apprentice in 1970. He spent eight years on private exchange technical support, five years with BT (CBP) Ltd. as international service manager and three years with BT's Customer Systems system integration business. He is currently responsible for the operation of BT's broadcast networks.

John Swingewood
BT Broadcast and Satellite Services

John Swingewood joined BT as an apprentice in 1972. He spent four years at university. He spent 10 years in Information Technology Systems before joining Visual and Broadcast Services in 1991. His current responsibilities include the operational and engineering aspects of all BT's broadcast and satellite products worth approximately £120M per annum. The services are increasingly being delivered into Europe.

Martin Bonner

Call Centres—Doing Better Business by Telephone

Uniting the power of the computer with modern telephone systems is one of the key ways of improving business efficiency, productivity and revenue-earning potential. Call centres, computer-telephony integration and a new appreciation of the power of the telephone are allowing organisations to improve their competitive position and provide better customer service.

Call Centres Defined

What is a call centre? The term is sufficiently vague to allow differing ideas of what a call centre is, what it does and what it includes. For some people *call centre* is a jargon expression used to specify a particular type of telephone switch. For others it is the entire building, the office that is dedicated to handling telephone contact with customers.

To clarify this position, definitions might be useful. *Call centre* is best reserved for the location at which telephone contact is managed. It includes the buildings, the staff, the desk and many associated components. The call centre is designed to manage large volumes of telephone calls. It is not an in-house telesales bureau or a jazzed-up name for an automatic call distribution (ACD) suite.

Customer contact system refers to the supporting infrastructure—the telephone lines, switch, computer systems—required to support the organisation in *doing business by telephone*.

Typically, even though there may be a prodigious number of telephone calls to answer, the purpose of the calls tends to be predictable and uniform. Staff (generally called *agents*), trained to handle only the types of call that the organisation experiences, rapidly become experts. Agents are equipped with telephone facilities and use the computer system either to access or input information.

The customer contact system, at its heart, integrates voice (the telephony) with data (the computers) but other technologies may be used to improve the agent's performance. For example, voice processing can be used to start and finish calls (as with directory enquiries). In the future, multimedia terminals could be used to support the agents with video clips. But the distinguishing feature is that all the elements of the system are integrated into one seamless whole.

Widespread Use

There are only three types of call centre: those which exclusively *receive* calls (order taking, customer service and the like), those which only *make* calls (telesales, market research) and those which handle both incoming and outgoing calls. A good example of this third category is the BT national telephone account management (NTAM) operation where the application is, as the name implies, account management.

Regardless of the type, all call centres have one thing in common: the organisation is being put into *direct contact with customers*. It is this feature that makes call centres the 'hot topic' with BT's business customers.

All commercial organisations, and an increasing number within the public sector, attempt to influence their customers. To do this they have to make contact and this is achieved through their chosen channels to market. Until now, organisations have had to choose between channels such as a sales force or a network of shops, television advertising or direct mail. Now there is a new channel to market: the telephone.

The power of *doing business by telephone* is that organisations are presented with a new channel, a new method of spreading their messages to their target customers.

the major cost benefit is provided by redesigning systems and processes to give a single point of contact for customers and right-first-time service

Real-World Users

Within the inbound, outbound and mixed categories of call centre there is a wide range of applications. They include:

- Customer service/enquiries
- Information service
- Sales-order processing
- Dealing/broking systems
- Account enquiries
- Handling of 0800 numbers
- Fulfilment centres for advertising response
- Insurance claims
- Appointment scheduling
- Hotel and ticket reservations
- Help-desk operations
- Telesales and telemarketing
- Debt collecting, debt counselling

The UK finance industry has been at the forefront of the move to telephone-based customer contact. Call centres are used widely by banks, building societies and insurance companies for handling customer service, providing information and improving client retention. Telephone banking is well established and buying insurance direct (meaning, by telephone) is the fastest growing sector of the financial-services market.

Typical call centre agent position

In other industries, call centres are already the norm; for example, the catalogue mail order companies, the emergency services provided by motoring organisations and advertising sales within the publishing industry. Similarly, hotel and airline reservations provide a clear illustration of how linking telephone agents to computer systems improves both customer service and the ability to sell effectively.

Benefits for Organisations

From an organisation's viewpoint the justification for introducing a call centre is, understandably, entirely commercial. Improving customer service is often the reason for embarking on telephone-based services, but just as frequently introducing a call centre enables a business to reduce costs and/or increase revenue.

The prime cost savings come more from the productive use by the telephone agents. Agents can handle calls more efficiently and their conversations, based on prompts presented on the computer screen, are more effective.

Further economies of scale arise from concentrating operations into a single coordinated location. But the major cost benefit is provided by redesigning systems and processes to give a single point of contact for customers and right-first-time service.

Cost savings can be dramatic. Productivity increases of up to 300 per cent have been claimed. One firm of factors (credit management) are now able to contact more than 200 defaulters per day compared with less than 60 before their system was installed. In the USA, one computer company cites the case of a system it installed for an electricity supply company; its client now claims savings of $3 million a year and managed a full payback on its investment within four months. A US bank, using an automated dialling system, increased the number of accounts it handled by 33 per cent during the first year of operations and it achieved the increase without adding any more staff. These achievements may sound extraordinary but they are not exceptional. They are certainly indicative of the technology's potential.

The scope for cost reduction is, however, dependent on the strategy underlying this move, which is why Syntegra, the systems integration business of BT, sets store by its consultancy service. Providing a single point of contact is only effective if the agent who answers the call is able to complete the transaction satisfactorily. Satisfying an enquiry may well require a number of separate actions to be taken by people who are dispersed across an organisation. The aim of a call centre operation is that a single telephone call should suffice to settle all these details. Empowering call centre agents—providing the tools to allow them to process enquiries fully—means stripping out layers of processing. The cost implications of this are significant.

Just as impressive is the impact on revenues. The call centre does not have to be dedicated to selling for the organisation to benefit. Even in the customer-service environment, customers may be persuaded to make a purchase. Agents can be prompted with on-screen information directed towards an additional sale. For example, in a computer supplies company, the agent may say: 'I remember, you're using Windows... well, we have a special that's not in this month's flyer and I thought you'd want to know we're offering PC Tools for Windows at a special price this week.'

This ability to 'cross-sell' is a significant advantage and allows the agent to introduce additional products and services into the dialogue and back them up with appropriate sales messages. Carefully structured scripts also ensure no points are missed and give greater control over what is said and when. This also avoids unstructured responses and promotes a businesslike impression in customers' minds.

Benefits for Customers

At the same time, the call centre provides a powerful means to improve customer service. This is not philanthropic: customer care is a live business issue and effective customer care can act as *the* differentiator between competing organisations. Traditionally, customer service has been seen as a cost to be minimised rather than the key element in customer retention. Efficient customer service, providing 'right-first-time' responses, is now recognised as an effective means of maintaining customer loyalty and it can be as much as five times more expensive to win a new customer than retain an existing one.

As a result, call centre operations, and the computer and telephony systems that support them, are gaining ground wherever organisations perceive customer service as a key objective. These organisations are not only retailers and financial companies: in the public sector the Citizen's Charter has made service a major goal for government departments and public utilities, and has set specific targets for speed of answering telephone calls.

But this trend to improved telephone service is still in its infancy. A recent market research programme sponsored by Syntegra reveals that 60 per cent of the major organisations contacted could not access their customer data either quickly or easily.

Growth Market

In a 1992 market survey, the research organisation Schema predicted that the European call centre market would grow from approximately $90 million in 1993 to almost $2 billion by 1999.

Not surprisingly, this kind of growth prediction has attracted many players to the call centre market. Network providers, computer vendors, software houses, consultancies, bureaux and systems integrators all compete for the same customers.

effective customer care can act as the differentiator between competing organisations

However, with the exception of the systems integrators, all the other suppliers are tied to their own products or services. And of the systems integrators, Syntegra is uniquely part of BT.

Meeting Expectations

The call centre concept is anything but new. BT has been handling customer service, by telephone, in call centres for the best part of a century. What is new is the convergence of telecommunications and information technology to provide new ways of handling large volumes of calls. This enables organisations to seize competitive advantage by installing modern customer contact systems.

Time is money as far as many customers are concerned. Customers' are becoming more sophisticated in their expectations of customer service. Once they decide to buy a product or service they expect rapid fulfilment of their needs. Advertising encourages this outlook, so companies interested in satisfying needs must be able to meet expectations—or competitors will.

An American consultancy, Oetting, concluded that up to 27 per cent of customers who cannot get through by telephone will either buy elsewhere or skip the transaction altogether. Call centres are playing their part in the crusade to make sure customers are handled attentively, without losing control of costs.

Meeting Customers' Needs: Call Sequencing and CTI

Callers ringing a busy number often assume that, if they are not answered rapidly, there is no one there to answer them, even though the real reason is that all staff are busy on other calls.

By installing call sequencing equipment, calls to an order office or help-desk service (to take typical examples) can be spread equally among all operators, and if calls cannot be answered within four rings, an explanatory recording cuts in offering an apology and possibly inviting them to leave their enquiry on a voice-mail system. As well as giving a more professional service, the equipment also traces the actual number of calls handled, which is extremely useful for planning staffing levels.

Computer-telephony integration (CTI) is another powerful ingredient in most call centre installations. Account, product and client information can be called in an instant from the computer, with powerful linking functions between the computer and the telephone switch. In a telesales operation dealing with inbound calls, caller identity information can be arranged to

It is essential that the supplier of the call centre technology has a thorough grasp of the client's business

bring the caller's account details to the agent's screen before he or she even takes the call. This enables the agent to greet the caller personally ('Is that Mr Smith from Worcester? Yes, I thought it was. Did you receive the literature we sent you last Wednesday?') and avoids the need to take down address, telephone number and credit card details which may have been taken on a previous occasion.

On outbound calls, an application involving debt recovery uploads and then dials numbers from BT's Phonebase service, as well as presenting the necessary information on-screen to the agent dealing with the call. Links to computers located elsewhere are commonplace, such as to check stock levels in the warehouse, requisition goods for despatch and produce address labels, packing notes and invoices. Until recently, information technology of this sophistication used to demand expensive, individually designed bespoke solutions; it is only now that they can be provided at acceptable cost as an adjunct to the voice switch.

One-Stop Shopping

Integrating computer systems with the telephone is not new but most applications up to now relied on the human operator to supply most of the intelligence. Even in companies which improved their customer-service operation by providing agents with client data on screen, the agent often had to consult more than one terminal and copy data from one system to another to complete the transaction. Moreover, the perceived improvement and feeling of goodwill were lost if the call had to be transferred to another department. Even if the telephone call was not lost in the system, the customer who had waited patiently now had to repeat his or her enquiry to a different agent.

With computer-telephony integration this is no longer the case. In most cases the agent can deal with all kinds of enquiry and if a call is transferred to someone else, the information on screen goes with it because the complete voice and data operation is being handled by one single processor. Less time is wasted, the agent works more effectively and the customer is left with a better perception of that organisation.

It is of course essential that the supplier of the call centre technology has a thorough grasp of the client's business. The integration implicit in computer-telephony integration means that it is vital that the organisation's IT strategy flows directly from overall business planning in general and marketing strategy in particular.

Components of Call Centres

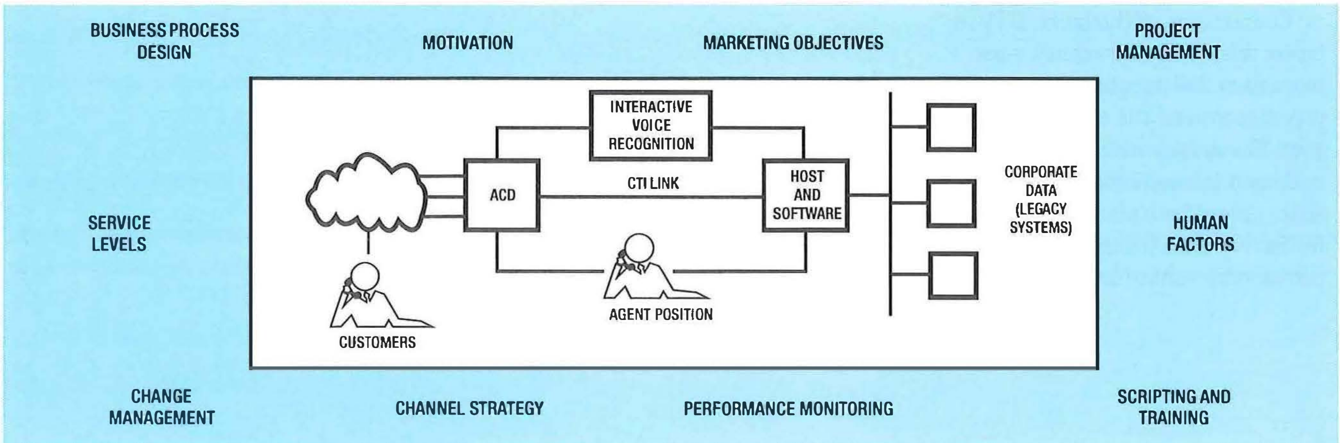
Call centres are complex. There is a vast range of considerations, both organisational and technical, that must be addressed before operations can begin. But regardless of the complexity, the technology is always less important than the implementation.

Successful call centre operations are derived from the organisation's overall goals and objectives aimed at maximising their gain from customer contact. Once these have been agreed, thoughts can progress to the underlying technologies. The four basic components of a call centre (see panel below) are:

- the delivery of calls (Freefone, Lo-call and alike),
- the automatic call distributor (ACD),
- the contact management software and computer system it runs on, and
- the access to existing corporate computer systems containing relevant customer and product details.

To this we can add cognate technologies such as voice processing and computer-telephony integration. But the list need not stop there. Workflow and image-processing systems, workforce management, billings systems, fax and e-mail and much more can be integrated to improve the efficiency of the operation.

Call centres—application is more important than technology



To the technical components of the call centre, Syntegra adds a keen understanding of the business issues that are central to 'doing business by telephone'. The emphasis is on setting objectives, understanding customer expectations and redesigning business processes to optimise the efficiency of the new installation.

Syntegra does not aim to supply all of the components of the call centre from within its own or BT's product range. Many of the elements are bought in as required from leading suppliers and then combined with BT products to form a composite installation. In this process, Syntegra acts as prime contractor taking on and managing the risk associated with the new investment on behalf of the customer.

Call Centre Applications Within BT

BT itself was an early adopter of call centre technology and it is this experience gained from its own operations which puts Syntegra in good standing for advising external clients. Indeed, it can be argued that BT's experience of managing telephone-based customer-service desks is unrivalled. Every day BT schedules around 6000 telephone agents to handle the calls received by directory enquiries, operator assistance, the emergency services, the customer-service centres and many other services. The services together handle some 750 million incoming calls every year and are available 24 hours a day.

Connections in Business, BT's in-house telemarketing agency, runs more than 200 agent positions and operates around the clock, 365 days a year. The agency runs inbound and outbound telemarketing campaigns that support both client and BT initiatives alike (ranging from the BT privatisation share issue to telesales).

Conclusion

Call centres, although relatively new, are based on mature technology and represent a major growth area. Syntegra is a leading provider of these systems, which enable BT's clients to improve the efficiency of their operations, making significant cost savings and improving customer service simultaneously. This can only be achieved, however, if organisations incorporate call centres, and the systems within them (both computing and communications), into their business planning. The solutions that Syntegra delivers must be an integral part of clients' total business strategy.

Biography

Martin Bonner
Syntegra

Martin Bonner has been with Syntegra, the systems integration business of BT, since 1992. In that time, he has been influential in establishing Syntegra's approach to contact management solutions—an approach that sees Syntegra working closely with many of the leading call centre suppliers. Prior to joining Syntegra, Martin worked for IBM, Tandem and Olivetti in a wide range of marketing roles. Martin has a degree from Leicester University and is a member of the Chartered Institute of Marketing.

David Preston

Frame Relay in Theory and Practice

In 1992, BT's Computing Service and Operations commenced the installation of a frame relay network. This article describes the technology, how it is used and results being obtained.

Introduction

Frame relay is an access protocol only. The standards defining it are concerned principally with the interface between the frame relay network and the attached devices. The standards do not define how frames are handled once they are inside the network.

The frame relay network operated by BT's Computing Service and Operations (CSO), in common with the public network, is based around Stratacom's IPX equipment. This is a fast packet switch in which variable length frames are broken up into small fixed length cells of 24 bytes. It is not essential to break frames into cells—whole frames could be transmitted. However, by using small fixed length cells, acceptable trunk utilisation can be attained without unacceptably high serialisation delays. This is the method used in the Stratacom IPX and described in this article.

Frame Relay and Other Data Communication Services

BT's (OSI) layer 1 services include MegaStream, KiloStream and Datel. Layer 3 services are provided by X.25. Frame relay lies in between and provides addressing, error detection and network management facilities between network and attached devices.

With permanent circuits (sometimes described as *circuit switching*), time-slots have to be reserved and paid for whether data is being sent or not. In addition, sufficient time-slots have to be permanently available for end devices to communicate at their maximum speed. Such a method is ideal for constant bit-rate services, but makes inefficient use of bandwidth if the source of data is not continuous.

With frame relay, frames of varying size are transmitted across the core/trunk as fixed length cells. Cells are transmitted (that is, relayed) only when data is to be sent. This results in two main advantages:

- the 'silence overhead' is eliminated as cells only occupy trunk bandwidth when there is something to send; and
- by using fixed length cells, a highly efficient cell packing density on the trunk can be obtained.

Conversely, there is a disadvantage—a cell overhead of 12·5%. If a network was dominated by continuous data rate services, this overhead could make circuit switching more efficient. With 'bursty' applications producing variable data rates, this overhead is offset by eliminating the silence overhead.

Frame relay offers the opportunity to reduce expenditure on core bandwidth and ports plus a reduction in network-added delays.

Origins of Frame Relay

Frame relay's predecessor X.25 is also a packet technique and was defined in an era when transmission media were prone to introduce errors and terminals were not intelligent. It was therefore necessary to ensure that only correct frames were delivered by the network—a dumb terminal cannot request a re-transmission.

Frame relay, however, belongs to a different era in which transmission media are less likely to introduce errors and end devices are often 'intelligent' and therefore able to cope with errors and re-transmissions.

Frame relay, unlike X.25, although it detects frames that contain errors, performs no error recovery— frames that contain an error are discarded. Frame relay assumes that the higher layers will cope with such losses.

Not guaranteeing delivery of frames produces another advantage—reduced serialisation delay. It is not necessary for an intermediate switch in a frame relay network to wait for the complete frame to be read in before it can be checked and if necessary request the previous switch to send it again. Round trip delays of around 15 ms are being achieved in the CSO frame relay network.

New demands for wide area networking facilities have also given rise to frame relay—in particular, the data requirements for local area network (LAN) communications. Inter-LAN communications have two predominant characteristics:

- *They are bursty.* As a result, if inter-LAN communications are sized to cope with average traffic levels, bursty applications will be starved when their bandwidth demands peak. Conversely, sizing for peak demands will result in wasted bandwidth. Frame relay overcomes this problem by only requiring a small amount of bandwidth to be permanently reserved for a permanent virtual circuit (PVC) and dynamically increasing the speed of a PVC when an application requires it (providing the bandwidth is available).

- *Meshing.* Fully interconnecting all LANs in a network becomes impractical and very expensive, lots of trunks and ports being required. As a result, the routers via which the interconnection is achieved are frequently used as intermediate hops. An example is shown in Figure 1.

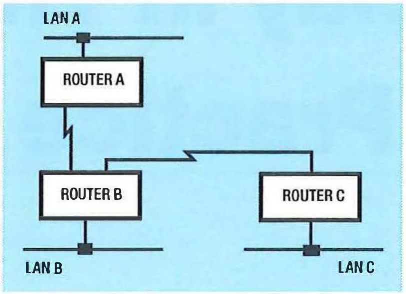


Figure 1—Passing data through an intermediate router

To prevent having to buy a trunk such as a MegaStream from router A to C, frames from LAN A to LAN C will be routed via router B. There will not be a virtual circuit between A and C via B. As a result, at router B, frames passing through router B will incur a delay as the frame is read in, queued for processing, processed, re-queued and re-transmitted over the trunk between B and C. As frames are commonly 1500 bytes long, such a serialisation delay can be significant. Frame relay overcomes the need for lots of trunks by permitting PVCs to many destination routers with one input port (see Figure 2). It also reduces serialisation delay as frames that can be thousands of bytes long are carried through intermediate IPX nodes as 24 byte cells.

How Frame Relay Works

When a frame is received from an attached device (described here as *data terminal equipment* (DTE)), the frame relay network identifies the destination, segments the frame into cells and relays the cells through the network (see Figure 3). The cells are reassembled into frames at the destination, the cyclic redundancy code (CRC) is checked and, if correct, the frame is relayed to the destination DTE.

Up to 251 PVCs can be present on a single physical port. Each PVC can be individually sized to reflect the requirements of the attached device or protocol. Having multiple PVCs on a single physical port results in another advantage. A high degree of meshing becomes possible from a single port; that is, it is not necessary to buy two ports for every two devices connected.

Figure 2—PVCs from one frame relay port to three destination ports

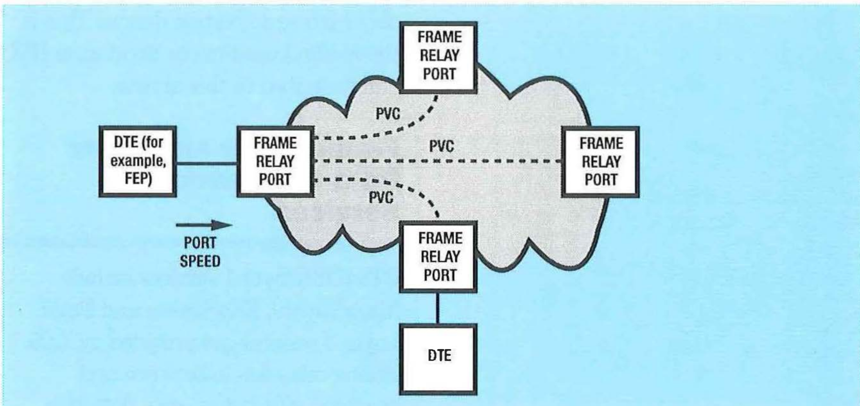


Figure 3—Conversion of variable length frames into fixed length cells

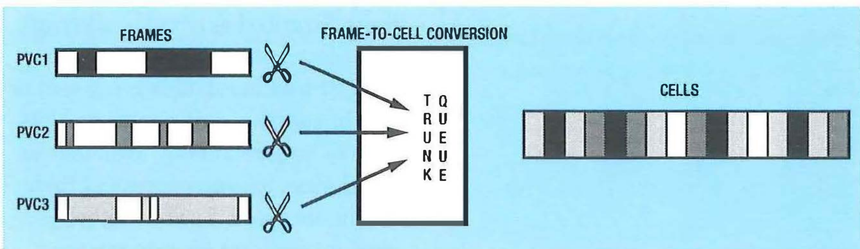
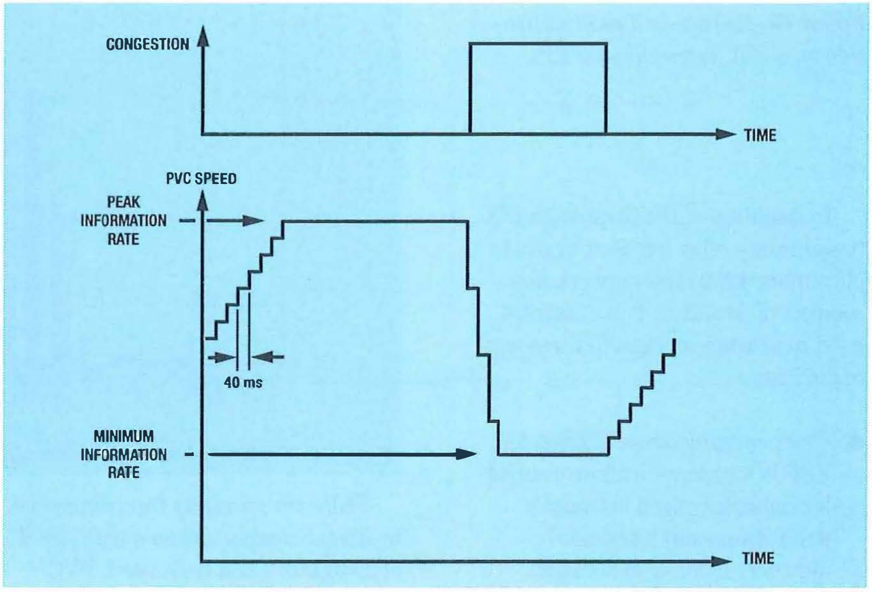


Figure 4—PVC speed variations



Metrics Defining a Frame Relay Connection

Three basic values are used (see Figure 4):

- *Port speed*—the speed at which bits are clocked into the frame relay port;
- *PVC minimum information rate*—the lowest PVC speed that is guaranteed to be available at all times; and
- *PVC peak information rate*—the maximum speed that the PVC can attain should there be sufficient trunk capacity.

Once inside the frame relay network, the speed at which a PVC runs is controlled by a congestion control mechanism known as *Foresight*™. When spare bandwidth is available on the trunks carrying a PVC, the speed at which the PVC is running is allowed to increase about every 40 ms. This continues providing there is data to be sent until the peak information rate is reached. Although the PVC speed will vary, the port speed remains unaltered.

Should congestion occur on a trunk, the speed at which the PVC is running is reduced by an amount that depends upon the degree of congestion. Reductions continue until the minimum information rate (MIR) is reached—reductions will not reduce the PVC speed below this level.

There are two main benefits of this process:

- providing data is present, trunks are always heavily loaded almost irrespective of how the PVCs are configured; and
- if two or more PVCs share the same trunk and one or more is not carrying data, other PVCs on that trunk can occupy unused trunk capacity. However, when the PVC(s) previously not transmitting commence transmitting, they get their minimum information rate at the expense of those that were running above their MIR. Any unused trunk capacity (that is, not required to provide the MIR) is available to all PVCs if/when required and is shared in the ratio of the MIRs.

Congestion

Frames enter the network at the port speed and are carried across the core at the Foresight™ controlled PVC speed. Clearly, it is possible for the rate at which data is generated by the DTE to exceed the PVC speed. When this occurs, congestion can build up and, unless the data source slows down, frames can be discarded. To cope with this, if congestion builds up when data is being sent from DTE A to DTE B, a *slow down* message is sent by the frame network to DTE A as part of the normal frames from DTE B to A; this is known as *backward explicit congestion notification* (BECN).

Unfortunately, not all DTEs connected to a frame relay network are able to control the root source of data; for example, routers connecting file servers on different LAN segments. Tests carried out have shown that, under congestion conditions, the frame-generation rate reduces, and that only when acute congestion arises is any data lost. This is probably due to the build up of congestion causing network delays to increase and acknowledgements to the sending DTE taking longer to arrive.

CSO Use of Frame Relay

CSO is using frame relay as part of its networking strategy (see Figure 5).

Figure 5—Layers within CSO network strategy

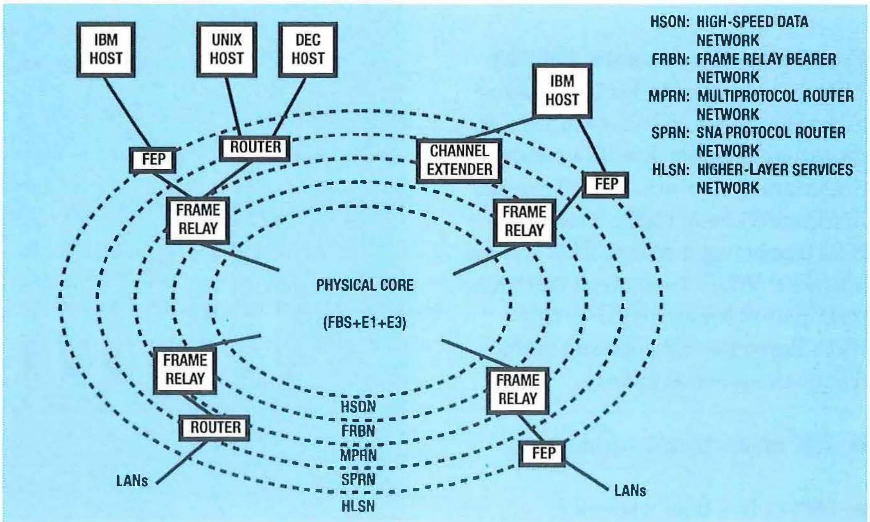
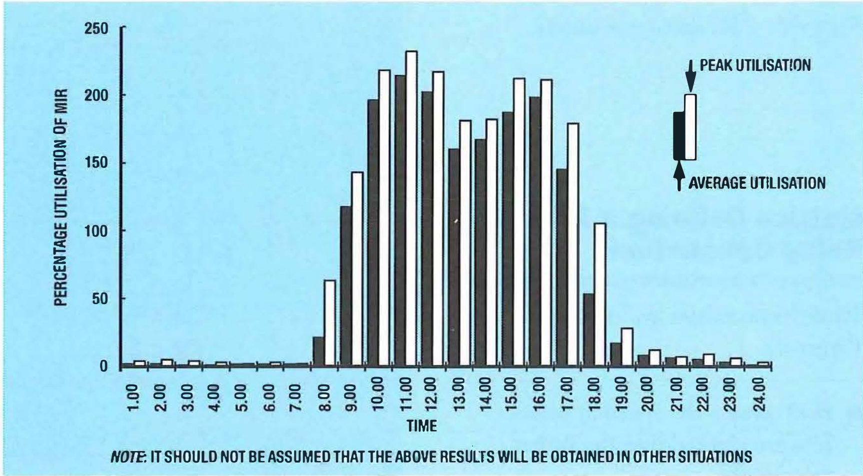


Figure 6—Average and peak utilisation on a PVC between two FEPs



Installation of the Stratacom IPX-based frame relay network began in November 1992. The network now consists of 31 nodes. It is currently used to provide services to three sets of devices:

- *Front-end processors (FEPs)*. An FEP is a communications controller commonly used between a main frame and a terminal network. It can also act as a switch for SNA traffic.
- *Main frame channel extenders*. Channel extenders are used to extend 4·5 Mbyte/s parallel IBM mainframe interfaces to remote channel-attached devices including mainframe printers.
- *Routers*. Routers are being used for internetworking remotely located LANs. Routers operate at the OSI network layer.

In selecting a frame relay network, cost savings were being sought in two ways:

- reduced trunk bandwidth costs; that is, no need to have trunk bandwidth permanently available to meet the maximum speeds required irrespective of whether data was being transmitted; and
- reduced port costs; that is, a high degree of meshing from a single access port.

Front-end processors (FEPs)

FEPs are being installed in telephone exchanges to act as remote interfaces for terminal traffic into BT's national SNA routing network. The histogram in Figure 6 shows traffic levels on a PVC connecting a remote FEP and its host FEP. Traffic from about 600 SNA terminals is being carried by this PVC. Response (network and mainframe) times are as follows:

- 78% in less than 1 second, and
- 98% in less than 2 seconds.

FEPs are normally interconnected by MegaStreams and so a port speed of 2048 kbit/s has been used. PVC minimum information rate is 128 kbit/s with a peak of 1536 kbit/s.

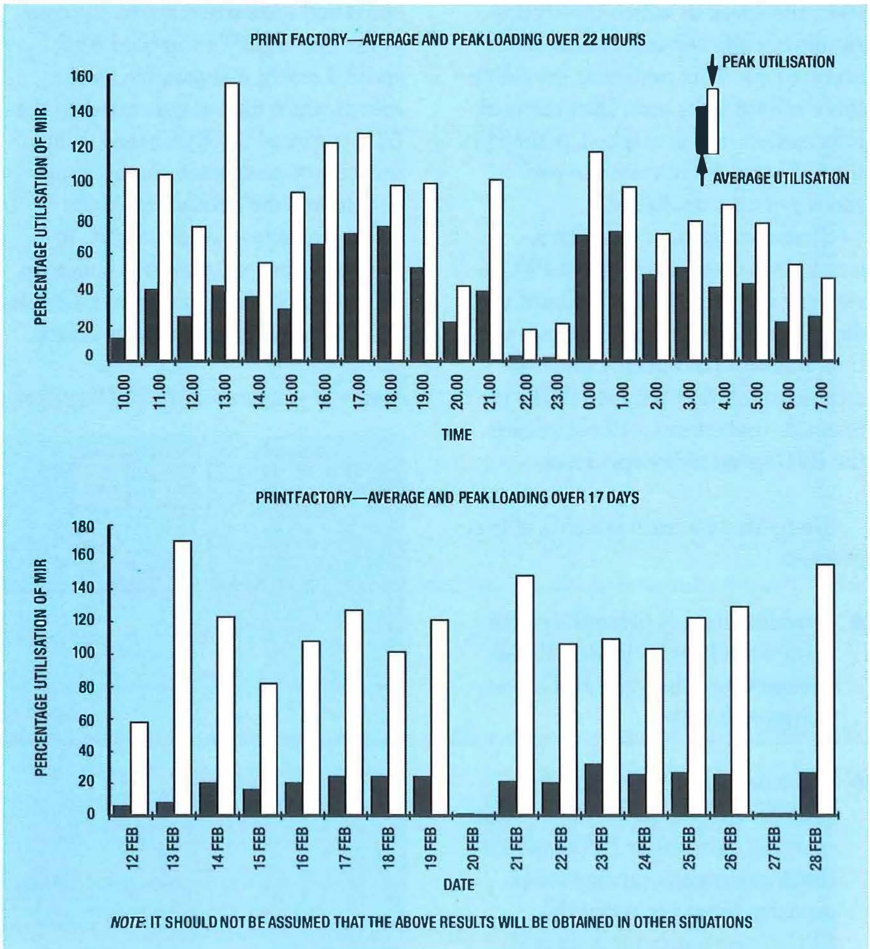
Significant bandwidth savings are being achieved without impacting upon response times. The FEPs, having a port speed of 2048 kbit/s, believe this is the speed they are interconnected at, but in fact only 128 kbit/s is reserved.

Channel extenders

BT is concentrating all bulk print work (for example, telephone bill printing) at a small number of locations. A modified form of frame relay known as *frame forwarding* is being used to interconnect channel extenders as they do not yet support frame relay.

Frame forwarding provides a subset of frame relay facilities for DTEs that do not support frame relay. However, the attached device must adhere to some simple restrictions; for example, not exceeding 4506 bytes per frame. Frame forwarding is supported via the normal frame relay packet assembler/dissassemblers

Figure 7—Average and peak utilisation on a PVC connecting two channel extenders



(PADs). As DTEs using the frame forwarding facility do not conform to any of the frame relay standards, some frame relay facilities are lost:

- multiple PVCs (point-to-point only), and
- congestion feedback (the DTE cannot use the information).

Once inside the IPX network, frame forwarding cells are treated as if they were frame relay cells.

The channel extenders, although normally interconnected with a MegaStream, have a port speed of 1024 kbit/s, a minimum information rate of 512 kbit/s and a peak of 1024 kbit/s.

Figure 7 shows an example where in a 22 hour period 117 000 telephone bills, 100 000 reminders and 4 000 000 lines of plain stationery were printed. Instead of interconnecting at the normal 2048 kbit/s, only 512 kbit/s is reserved.

Acknowledgements

Foresight is a trademark of Stratacom Inc.

Biography

David Preston
BT Development and Procurement

David Preston is a senior network analyst in Computing Services and Operations. He joined the company as an apprentice and, after gaining an honours degree in Electrical Engineering, worked on the development of computer systems used in network planning and design. Responsibility for the planning and implementation of the computer network for BT's Directory Enquiry Service followed. Since then he has been involved in the multiplexers used by BT's internal computer networks. He is a member of the Institution of Electrical Engineers and a Chartered Engineer.

Glossary

CSO Customer Service and Operations
DTE Data terminal equipment
FEP Front-end processor
LAN Local area network
MIR Minimum information rate
OSI Open Systems Interconnection
PVC Permanent virtual circuit

Tony Mullee and Bob Faulkner

Planning for a Customer-Responsive Network

Customers' needs are changing at a rate never before experienced in the telecommunications industry. Customers want new, innovative services tailored to their needs and delivered to increasingly compressed time-scales. In order to meet this challenge, the orientation of network planning must be towards fulfilment of customer needs and not towards the unfocused exploitation of new technology. This article seeks to consider some of the issues facing telecommunications companies in the planning and implementation of customer-oriented networks and services.

Need for a Customer-Responsive Network

'We have always had a customer-responsive network. We plan to meet forecast demand and functionality, we aim to keep pace with technology, our customer complaint levels are steady and our call volumes and customer base are stable. We like to think we do our best to keep our customers happy (within certain constraints). So what's likely to be so different from now on?'

This statement could be typical of a moderately progressive but non-competitive telco. On the face of it they are well intentioned, keen to progress and adequately aware of their customers' needs. However, what our typical telco lacks is an 'obsessive' preoccupation with the customer, determination to understand their requirements, and a proactive attitude towards meeting those requirements **and** the effects of impending competition head on.

In order to prosper, indeed, in order to survive, the cosy, introspective world of the technology-led telco must give way to the lean, sharp, outwardly focused and customer-led company of the future.

Competition and Regulation

In the current climate, the interrelationship between regulation and competition is so tightly coupled that it is impossible to consider either one in isolation.

This article is based on a paper originally prepared for the 6th International Network Planning Symposium

Regulation, or rather liberalisation, of telecommunications has clearly been the main enabler to competition. The UK Telecommunications Acts of 1981 and 1984 opened the doors for limited competition to improve customer choice of products and services on a national basis. Since that time, the pace and scale of regulatory change has increased and extended to encompass both European and global communications.

Open Network Provision (ONP) will require operators to meet, measure and declare performance against targets which until now were internally set and accounted against. It will also stipulate the frameworks for open systems by which to ensure effective interconnection between operators and service providers, in pursuance of the vision of truly seamless pan-European and global communications. But how should today's operators plan to take account of the dazzling array of contradictory regulatory, competitive and economic pressures with which they are now faced?

If regulation seeks to promote competition and customer responsiveness, then to achieve this aim, even on a national scale, it needs the support of standards for effective interworking between dominant and emergent networks. However, standards take time and are often expensive to implement for manufacturers, service providers and network operators alike. Customers (and potential service providers) become frustrated at the apparent lack of progress and the fact that both regulation and standards often fail to address adequately those aspects of service of greatest importance to them.

Figure 1 seeks to reflect this confusing array of influences, and

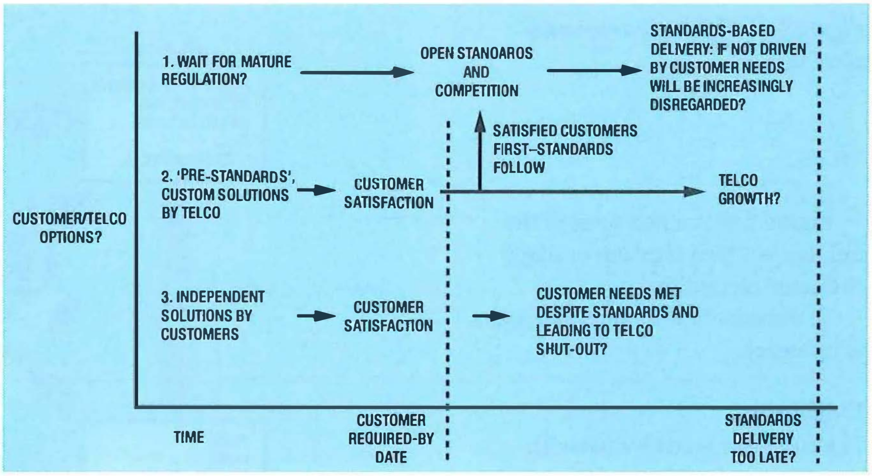


Figure 1—Regulation and standards: lead, follow or get out of the way?

postulate upon the preferred route for the telco determined to survive.

The telco that invests effort into understanding and meeting customer requirements can have full confidence in its own market-led strategy and be assured that, in time, customer-oriented regulation and standards **must** follow. They will also have the confidence to embrace and develop relationships with service providers and other operators based on a sound, customer-focused vision. This assumption is borne out by a recent article in an international communications publication[1], which stated, on the subject of international regulation:

'Faced with the same old operator offerings, telecoms users and service providers are bending the (regulatory) rules. Last year's (1993) telecoms services review by the European Commission did not produce the results many business users wanted—immediate liberalisation of **all** network infrastructure and services markets. Now European telecommunications users and service providers are finding ways to stretch the boundaries of regulation in order to meet their needs.'

Clearly, if the users themselves are prepared to challenge and shape European regulation and standards through innovative, often corporately-based solutions, then the telcos must be similarly disposed if they are to survive.

Cost Control: Impact on Pricing and Profit

In the pre-competitive environment, network infrastructure investment is a fairly introspective process.

For the newly privatised company comes a fundamental switch of focus from what was essentially a sociological orientation to one of profit, share value, efficiency and competitiveness. This is a paradigm shift in emphasis requiring nothing less than a total culture overhaul.

The privatised telco perceives the need to modernise, to exploit technology, under the misconception that success in the competitive arena will go to the operator with the most technologically advanced network infrastructure. This is one of the most expensive lessons on the road to competitiveness. Customers are given (what the telco believes to be) the best of everything that they could possibly want. However, in this well intentioned but misguided drive for universal technological excellence, very few customers are actually given what they **need**.

Over time, the effects of competition and regulation will become more acute. Pressure on tariffs pushes towards a bottoming-out effect, and opportunities for differentiation on price recede. During this phase the drive is for increased internal efficiency, operational rationalisation and any cost savings which can result. The eye of the company accountant is firmly focused on the diminishing margin between operating costs and revenues.

The clock runs on remorselessly and the evolutionary principles of natural selection and survival of the fittest continue to apply.

The way for the telco to differentiate itself and survive the rigours of the competitive evolutionary process is to be among the first and the best at focusing upon, understanding and satisfying the real needs of its customers.

The telco can then begin to offset the effect of diminishing returns by focusing capital spend where it will result in the greatest return; that is, where the beneficial effects of the spend most closely align with the needs of the customers affected by them. (An example would be the selective upgrade of only that part of the digital transmission infrastructure, to say synchronous digital hierarchy (SDH) or asynchronous transfer mode (ATM), which will serve the customers or customer segments with a specific requirement, for high-grade speech or data links, with higher than 'normal' availability and error performance.)

In this way, the impact on customer perception and the benefits in terms of customer loyalty and resultant revenues will be most positive.

Customer Awareness: Opportunities for Choice

Clearly there are many elements in the complex array of influences enabling increased customer awareness.

Business customers have well developed, clearly defined needs and their choice of operator will be principally governed by their level of confidence in an operator to satisfy them and hence support their business activities.

Conversely, personal customers may be less discerning and accept more basic, reliable communications services. They are more influenced by softer 'service surround' issues such as courteous customer service and an efficient, effective repair service. Price is an issue for them but few personal customers would go to the inconvenience of switching operators on the basis of a nominal price differential alone. Any such move would be either opportunistic, such as during a house move, or be prompted by an emotional reaction to a negative, unsatisfactory experience with the operator.

Figure 2—Influences on customer perception

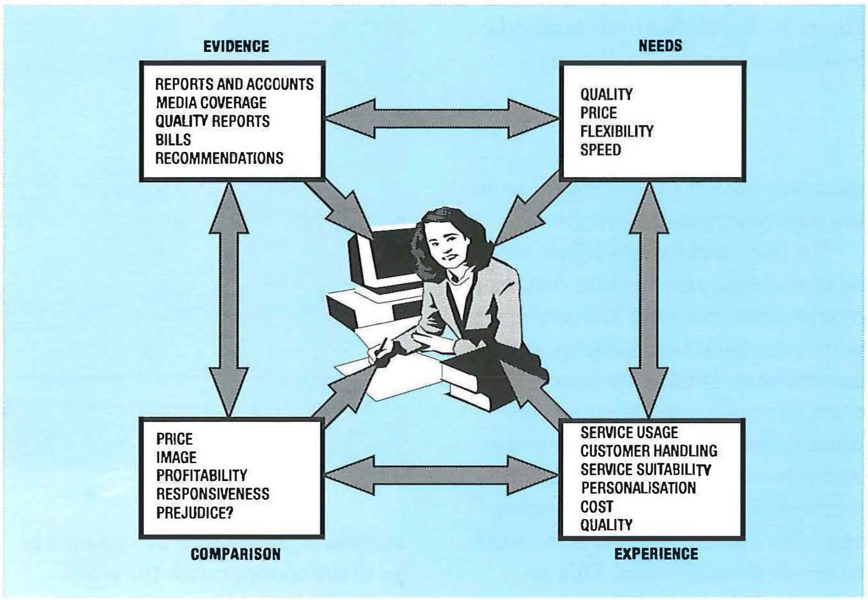


Figure 2 shows just some of the influences which combine to shape customer perception.

To consider briefly the major areas of influence:

Evidence

The customer is confronted with many forms of evidence, from external publications and media, to publicity and management reports produced by the operators themselves.

Badly presented bills or management reports awash with meaningless information will evoke a negative reaction from the customer. By paying care to those aspects of the customer experience concerned with the generation and consumption of 'evidence' the telco has a powerful tool through which to influence, win and retain customers in a competitive environment.

Comparison

The criteria by which one operator is compared with another will vary over time and with the needs of the customer in question.

Price is still a major benchmarking issue, although new market entrants now frequently cite quality-of-service and, in particular, service availability as prime bases for differentiation.

Needs

A personal customer may need to receive separate bills for a number of individuals (for example, teenage children) who all make calls from a single telephone line. During certain periods in their life, such as family illness, it may be important for them to have 100% service availability. Hence their needs will vary with their circumstances.

A business customer may need a structured, customised communications plan, requiring services and service combinations outside of the telco's portfolio or technical capability. They may need a particular style, frequency or delivery of billing information to assist in dynamic business accounting.

Marketing, product and, therefore, network plans based upon wide generalisations about customer needs are no longer adequate. Market segmentation is tending towards individual customisation and context marketing, which understands that the needs of even a single customer will vary widely over time. In order to be responsive to those needs the telco must adopt a dynamic, attentive culture, attuned to changing customer needs and driving processes and systems equipped to respond flexibly and rapidly.

Experience

The most direct, and therefore the most potent form of influence on customer perception, is that of personal experience.

Each customer will have certain expectations against which they will be most sensitive to the quality of their experience with the telecommunications operator. If the telco is not adequately attuned to those expectations then the quality of the experience for the customer will be low.

Understanding Customers' Needs

The telco's largest customers are characterised by increasing expectations, greater sophistication, computer literacy and now global coverage. Telcos need to provide their business customers with services that enhance and add value to their customers' own business activities. Major customers' requirements can be summarised as follows:

- the ability to grow organically or by acquisition without system or communication constraint;
- the ability to create and maintain competitive edge;
- the ability to drive and control their supply chain; and
- the ability to offer any product through any fascia without loss of control.

In order to support these requirements, telcos will need to offer their customers the following service characteristics:

Interoperability: The ability to interconnect all their equipment and sites to form an integrated network. The ability also to connect any customer premises equipment into the network without complicated interfaces or operating procedures.

Flexibility: The ability to offer a wide range of customised features—such as bandwidth on demand, intelligent network services and a range of guaranteed volume/duration options at an accurate costed price, with tariff options.

Multimedia services: The ability to combine voice, data and value-added services over the same network, through one access point with the minimum amount of customer premises equipment.

Figure 3—Framework for capturing customers' requirements

SERVICE FUNCTION \ SERVICE QUALITY CRITERIA		SPEED	ACCURACY	AVAILABILITY	RELIABILITY	SECURITY	SIMPLICITY	FLEXIBILITY
SALES								
SERVICE MANAGE-MENT	PROVISION							
	ALTERATION							
	SERVICE SUPPORT							
	REPAIR							
	CESSATION							
CALL TECHNICAL QUALITY	CONNECTION ESTABLISHMENT							
	INFORMATION TRANSFER							
	CONNECTION RELEASE							
BILLING								
NETWORK/SERVICE MANAGEMENT BY CUSTOMER								

Network management: The ability to control and re-configure the network to suit their individual requirements. Request up-to-the-minute information on the status of their network and all telco action taken on their behalf.

Single point of contact: The ability to access one point within a telco which would then take total responsibility for customer interface planning, implementation and operation of their networks.

Customised billing: The ability to provide customers with whatever type of bill they require whenever they require it.

Value for money: The ability to provide customers with the correct quality of service for the price paid. Customers expect a lot more from the network than they did 10 years ago; their requirements today are based on cost, quality and, for business customers, global reach. Therefore, the challenge of tomorrow's network planners is how to provide global networks with the required quality at the required price. This is a simple challenge to articulate but extremely difficult to implement.

Telcos are faced with the apparently contradictory objectives of improving quality to meet customers' expectations and reducing costs. The telco must therefore establish and maintain a balance between cost and quality. The challenge of the 1990s is how to build networks which reduce costs, and improve quality within an overall framework which meets customers' requirements.

Interpreting Customer Requirements and Responding Effectively

In order to plan and implement minimum-cost networks and support facilities which meet customers' requirements, telcos need structured tools and processes with which to interpret customers needs with speed and consistency. Any future network plan will have an auditable trail back to customer requirements. An example of such a tool[2,3], developed by the Federation of Telecommunications Engineers of the European Community, and its use are now explored as a front end to the quality-of-service component of the network planning process.

The framework shown in Figure 3 represents a generic matrix which enables the user to take a consistent and methodological approach to obtaining customers' quality-of-service requirements and fixing performance targets with consistent definitions and, therefore, consistent deliverables.

An example, again referring to Figure 3, is the service quality criteria 'speed' and service function 'provision'. Definition: 'The time taken from a contract being signed to the time the customer can use and be billed for the service'. Measure: Parameter expressed in days.

Each cell on the matrix has been defined in this way, thus creating a 7 × 11 = 77 cell matrix for any telco product or service. This, together with other criteria, will form the foundation of the customer requirements for any given service which will then need to be split into network specific and non network specific requirements as shown in Figure 4. The network-specific customer requirements can then be converted to technology capability and a minimum-cost network built which meets customer needs.

The major benefits of this approach can be summarised as follows:

Figure 4—Minimum-cost networks from customer requirements

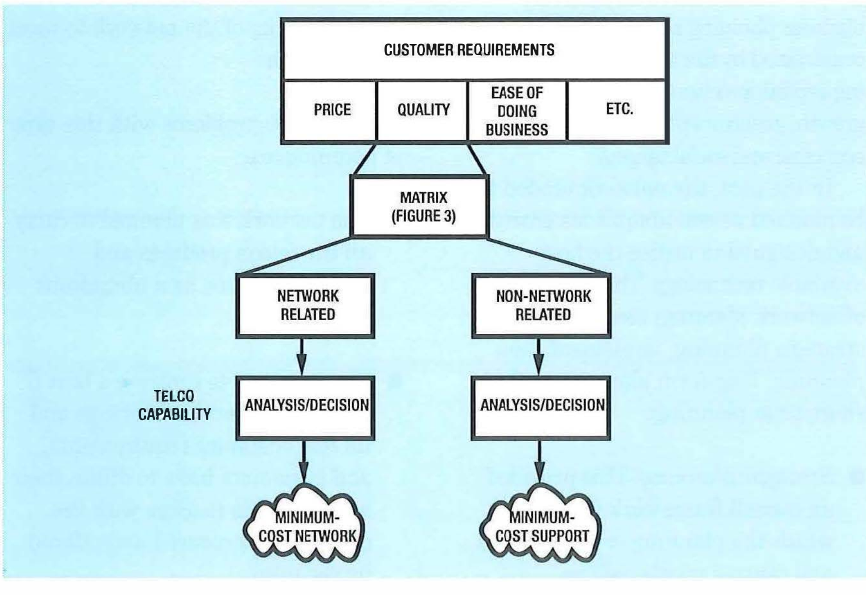
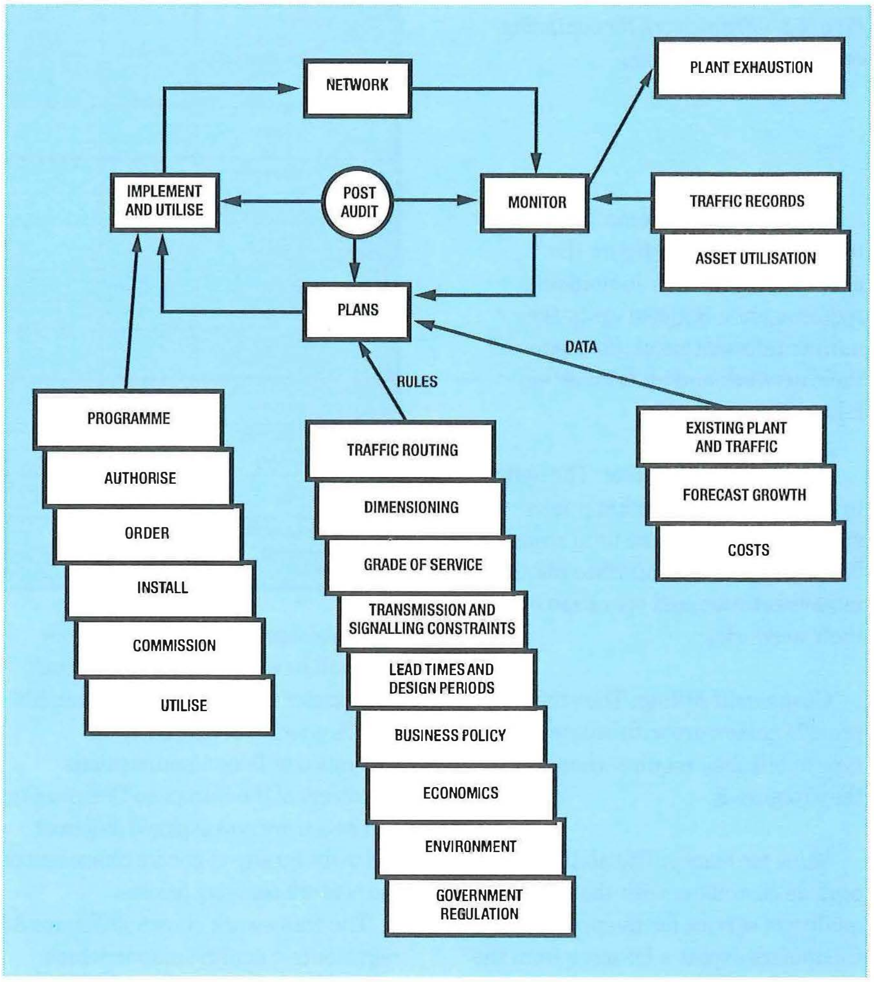


Figure 5—Traditional network planning process



- methodical approach to obtaining customer requirements for any product or service;
- methodology to build networks to meet customer requirements using cost-effective technology; that is, planning and building customer-responsive networks;
- basis for customer-oriented network performance measures and targets;
- change of culture within a telco away from technology for technology sake to technology to meet customers' requirements; and
- all network infrastructure spend auditable back to a proven customer requirement.

Traditional Network Planning

In the past, network planning consisted of a continuous process of monitoring the state of the network, producing plans to meet the requirements for growth and enhancement, implementing those plans and auditing the outcome against expectations as depicted in Figure 5[4].

The end product of network planning is a heavy capital investment and the use of resources which in turn result in revenue and profit. Network planning was, and still is, inextricably tied to business planning and therefore constrained by the telco's policy including capital, procurement, manpower, growth, government regulation and economic and social targets.

In the past, the network tended to be planned as one ubiquitous entity and designed to utilise the best available technology. The main stages of network planning consisted of strategic planning, implementation planning, long-term planning and short-term planning:

- *Strategic planning*: This provided an overall framework within which the planning of the network and centres must conform.
- *Implementation planning*: This provides the evolution plans from the existing network to the future network.
- *Long-term planning*: These were broad-brush plans to take account of the more detailed requirements of the individual centres.
- *Short-term planning*: This comprised the continuous review and re-planning of the network to meet its growth.

The major problems with this type of planning are:

- The network was planned to carry all the telco's products and services; that is, as a ubiquitous whole.
- The telco has to contrive a best fit between its service offerings and its real customer requirements, and customers have to dilute their requirements to align with the capabilities procured and offered by the telco.
- New technology was provided on a 'blanket uplift' basis—that is, to the entire network—which led to uncontrolled technology costs, surpluses of obsolescent technology and expensive research and development designed to find ways of adapting existing technology to support emergent services.
- The growth of the network was based mainly on the historical growth.
- Response to evolving customer needs was sluggish owing to the extremely long network planning lead-times (see Figure 5).
- Quality of service was reduced to the lowest common denominator and based upon technical capability and not customer requirements.

Within a telecommunication business, network planning gives rise to the largest proportion of capital expenditure and probably the largest current cost in terms of supporting the network. It is now evident with the advent of competition and the focus of telcos towards the customer that this traditional method of planning the network, with its extremely long lead-times, is no longer sustainable.

Responding to the Future

The right culture

Most competitive, progressive telcos are trying to become more attuned to the needs of their customers. Customers will express their needs in every way from simple phrases of common language to highly specific communications plans including service capability, numbering, billing, management and quality.

The challenge for the telco is to be equipped to understand the full range of customer needs, and to identify from the total set a rationalised range of service solutions through which to satisfy economically the largest number of 'individual' customer needs.

For all but the most enlightened telco, the move to true customer orientation requires nothing less than a total culture change on the scale of that required for total quality management (TQM). Further developing the synergy with TQM, it would be naïve and misguided to assume that polishing up the customer-facing part of the business is all that is required. Indeed, customer responsiveness in a customer-facing department **should** be the easy bit and the least part of the challenge. The really tricky bit is ensuring the following:

- Customer orientation must become the very lifeblood of the business, the baseline against which all other activities are referenced.
- Customer requirements must not become diluted, modified or lost on their journey through the business. Through structured processes and systems it should be possible to track a single customer requirement through all stages of realisation from raw requirements, through feasibility, network planning, implementation and service launch.
- The telco must not merely map customer requirements to existing portfolio offerings, providing the customer with a 'near enough' attitude.
- The telco must be able to track every item of infrastructure spend to a verified and proven customer requirement.
- Customers' needs must be viewed by the telco from a customer perspective and network plans must take full account of those needs on an end-to-end basis and not a traditional access, switch, core network type split.

The right structure and processes

Faced with such a diverse range of customer requirements it would be impossible to expect to capture and progress each of them on an individual, ad hoc basis. Indeed, without the necessary structure and consistency, any such approach would be doomed to failure in all but the smallest of telcos. For the effective, consistent capture and satisfaction of customer requirements the telco or service provider must adopt well structured, consistent and widely understood mechanisms for the following:

Capturing customer requirements

This requires a subtle balance between encouraging the customer to engage in an unconstrained expression of requirements to guiding them, perhaps through structured interviewing, to supply the full breadth of information required for meaningful analysis.

Interpretation of requirements

This is the first stage in converting raw requirements into terms which are meaningful for engineering feasibility, costing and network planning. At this stage the use of a tick-list or template is essential to ensure that the full spectrum of issues are captured including quality, price, volume potential, ergonomics and functionality. In respect of intelligent network services this could mean the expression of functionality requirements in terms of service independent building blocks (SIBs), service management objects etc.

Rationalisation and verification

Having determined a structured interpretation of customer requirements we can check to see which of the required capabilities we already have and which we may need to build. This is essentially a *gap-analysis* process.

Evaluation and costing

How much will it cost to 'bridge the gap' in our capabilities. Is it a market we wish to enter, is it technically feasible, how much would we have to charge to make it viable and would the customer pay?

Detailed planning

Having decided to embark on a course of action we need to quickly develop a suitable architecture, identify procurement needs, establish an end-to-end quality plan.

Procurement

By this stage we are equipped with some powerful information. We know exactly what our customers want, we know exactly what it means in network engineering terms, we know our margins and we can put together an extremely concise procurement specification covering the full spectrum of our needs. Previously we would probably have only specified the bits we knew about only to find later that our 'off-the-shelf' purchase did not entirely fit the bill.

Installation and testing

No problem! We have an end-to-end performance specification for our

Telcos will need customer-responsive networks in order to survive in a truly competitive environment and to retain and increase market share.

network components and service delivery mechanisms. We can test against a customer perspective and confidently embark on a beta trail involving our most discerning customers.

Launch
Our slick processes and comprehensive understanding of customer requirements mean rapid launch, no delays for costly reworking and no dissatisfied pilot customers to spread 'bad vibes' about our emergent service.

Process characteristics
At all of the previously mentioned stages we need the ability to track progress against the original customer requirements. The template which we used at the interpretation stage will have carried right through from concept, inception etc. to launch. A powerful tool which can aid the establishment and verification of these threads throughout the programme is the quality function deployment method (QFD) or 'house of quality'. Unfortunately limitations on space preclude further discussion of this methodology which is nevertheless well worth exploring 'off-line'.

Need for a Customer-Responsive Network?

Telcos will need customer-responsive networks in order to survive in a truly competitive environment and to retain and increase market share. This can only be achieved as follows:

- Ensure that customers' requirements are fully identified and properly understood.
- Give customers exactly what they want, **not more**. Blanket modernisation has proven to be an expensive indulgence which modern telcos can ill afford.
- Specify, plan, procure and build our networks to meet customers' requirements; that is, build minimum-cost network(s) in terms of infrastructure, management and support with focused network investment only when justified in terms of customer requirements.
- Ensure that each and every element of network expenditure is auditable back to a clearly identified and understood customer requirement.
- Deploy new technology according to a customer-focused, market-led strategy.

References

1 *Public Network Europe*, Oct. 1993.

2 Quality of Service Considering Customer Requirements, FITCE Study Report, 1994.

3 MULLEE, TONY, and OODAN, ANTONY. Telecommunications Quality of Service: Principles and Management. *Br. Telecommun. Eng.*, April 1994, **13**, p. 53.

4 WARD, K. BT Network Planning Guide.

Biographies

Tony Mullee
BT Global Communications

Since joining BT in 1972, Tony Mullee has held a number of management posts within both the technical and the commercial sides of the business. He was also one of the triad responsible for the development of BT's Master's Programme, designed to develop BT's future senior managers. Tony chaired the two-year FITCE European Study Commission 'Study of Network Performance Considering Customer Requirements'. Currently, Tony is employed as BT's Strategic Global Marketing Manager, with responsibility for BT's Advanced Managed Services Portfolio including ServiceView and global billing. Tony has a B.Sc. in Applied Physics, an M.Sc. in Telecommunications Management and an M.B.A. from Essex University.

Bob Faulkner
BT Development and Procurement

Bob Faulkner joined the British Post Office in 1972 as a Youth-in-Training in the Oxford Telephone Area. After completing an HNC in Computer Studies in 1986, he joined the Systems X Product Support Group in London and in 1989 moved to the Central Operations Unit in Oswestry where he was responsible for network traffic management support and development. He went on to establish and manage the Oswestry interconnect fault reporting point, working with a team to provide 365-day, 24-hour fault management services to other licensed operators, a period which greatly reinforced his awareness of the critical role of customer service in BT's continuing success. In 1991, he joined the newly formed Network Performance Planning Department which declared as a key part of its mission 'to lead a culture change within Worldwide Networks towards a more customer-focused perspective'. Currently, Bob is employed as Manager, Intelligent Network Services Operations Development, with responsibility for ensuring that BT's intelligent network platforms and services are supported by operational processes and systems which will lead BT's drive for improved speed to market, service flexibility and customer-service excellence.

Denis Johnston

Speech Recognition for Speech Services

The benefits of machines able to recognise and understand speech are enormous. While speech and language processing are not at the stage where we can have casual yet meaningful conversations with a computer, the past five years have seen an increase in systems which use simple speech recognition, and the biggest single growth area has been in telecommunications. This article provides a background to the technology of speech recognition, looking in particular at how it is used in telephony applications and where it may appear in the future.

Introduction

If we could simply talk to machines a major barrier to the application of computers would be overcome. Not only would speech input make computer-based systems simpler for the vast majority of users, but the telephone network would provide low-cost access for an almost unlimited range of information-based services.

Speech is the quickest, easiest and most expressive means we have for communicating with each other. However, when we interact with computers we normally have to resort to more artificial methods: keyboards, tonepads, touch-screens or mouse-driven menu systems.

So, why are we not yet surrounded by machines that listen to us and can understand what we say to them? Why is it that such an obvious solution is not in widespread use? What is so very difficult about making computers that can recognise speech?

This article aims to answer these questions by providing an overview of today's automatic speech-recognition technology, how the problem is being tackled and how existing solutions are being applied to systems within BT.

It is hoped that this article will help potential users of such equipment to become familiar with the basic ideas behind the technology, understand the terms used and be aware of the opportunities today's technology can offer. Furthermore, it is hoped that it will help designers appreciate the system requirements that must be met if today's automatic speech recognition is to be successfully integrated into applications.

The article is in three parts. The first is a tutorial that describes the nature of speech including how it is produced and perceived. The second provides a 'state of the art' summary of current speech-recognition techniques. The final part addresses technology implementation issues and applications.

The Nature of Speech

Speech, speakers and language

Although most of this article is concerned with automatic speech recognition, the most ubiquitous form of speech recogniser is of course the human being. It is therefore appropriate to preface any discussion of automatic recognition by considering the nature of speech itself, how it is produced, how it is perceived and how it is linked with language.

As children we develop speech skills so effortlessly that it appears obvious that it must be a very simple process. This notion is reinforced when we listen to speech as we seem to hear clearly separated words each of which consists of a series of clear, unique sounds. What is more, the fact that we all can communicate so easily with each other so effortlessly strongly suggests that the underlying signals must have something in common and be part of a universal coding system. Finally, we know that we can write down speech using only the 26 letters of the alphabet, again supporting the view that speech must be very simple.

On the other hand, when we think about it more deeply, our experience also suggests that perhaps speech may be more complicated. For

Figure 1 (a) Figure 1 (b)

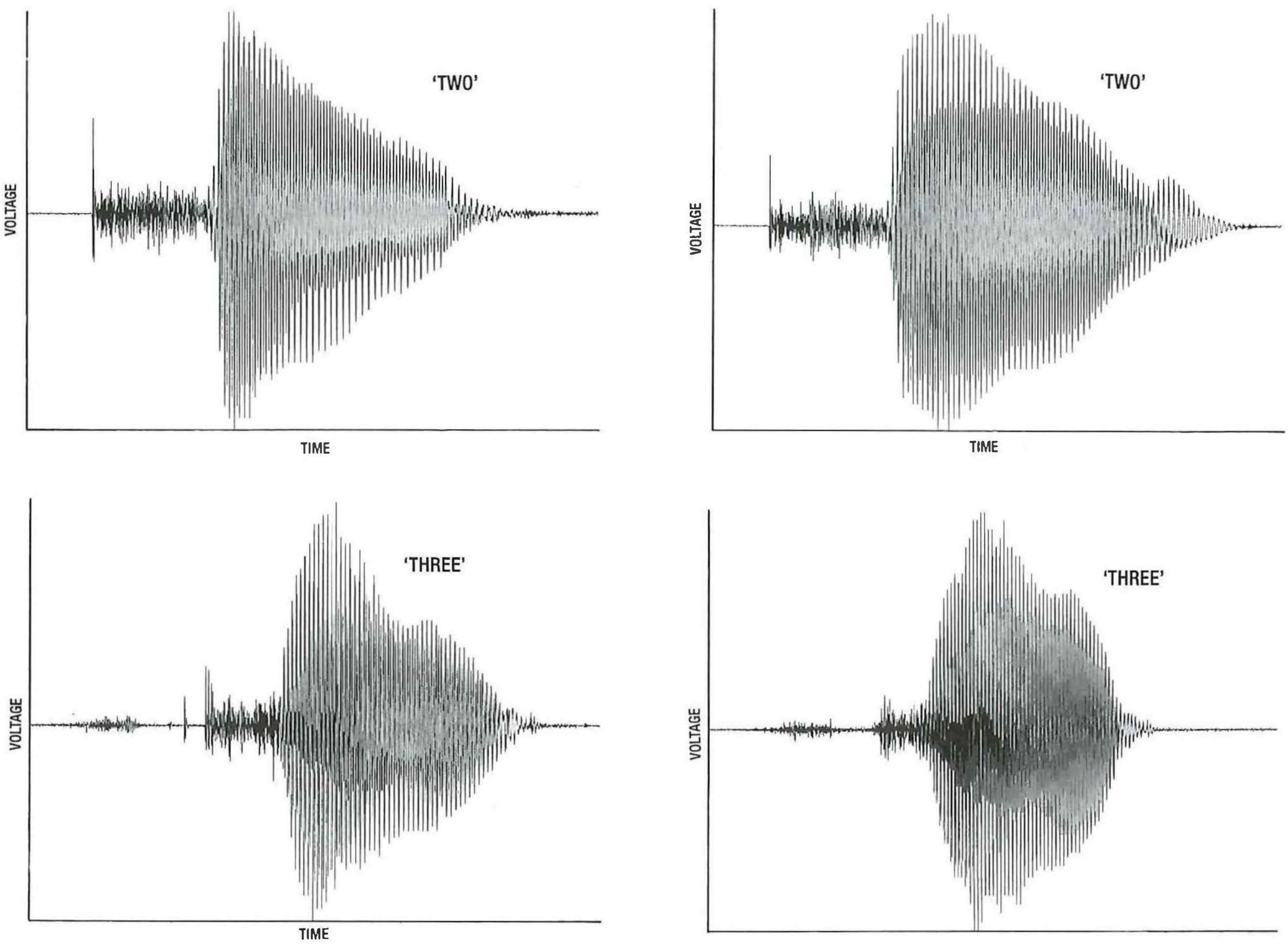


Figure 1—Single speaker, several words (top, two speakers saying 'two'; bottom, two speakers saying 'three')

example, when we listen to someone talking, we can usually tell something about the importance of what is being said from the 'tone of voice'. Likewise, we can often recognise the speaker and appreciate what their mood is at that instant. Even when we do not know them, we can often tell their approximate age and their sex, and make a good guess as to where they come from by their accent. Sometimes we can judge something about their educational or social background and may even be able to say something about to whom they are talking: a friend, a child, a crowd, a pet, etc.

The evidence that speech can simultaneously contain all these different messages suggests that there is much more to speech than 'mere words' and hints that it is much more than a simple linear coding scheme. The truth seems to be that speech is actually a very complicated phenomenon. So much so that much of how we perceive and understand it remains unknown.

It is also clear that it is impossible to discuss the nature of speech without taking into consideration many other factors. For example, our ability to 'hear' words is not just an auditory process—it is linked strongly to our language-understanding capability. As we listen we use our knowledge of language to 'make sense' of what we hear: indeed it is generally accepted that it is higher-level linguistic processing (as opposed to our auditory mechanism) which allows us to separate the flow of sounds into our ears into separate words and sounds. Anyone who has listened to an unfamiliar foreign language is aware that it is not generally possible to identify where words begin and end. Indeed, without understanding what has been spoken, it is very difficult to remember and repeat a simple sequence of sounds beyond a span of a few syllables.

But while the nature of speech and language remains elusive from a philosophical point of view, over recent years a lot more has become known about the acoustic properties of the speech signal. In particular, our ability to store and analyse speech using digital computers has meant that we can now examine the 'raw material' more effectively than ever before.

This approach (sometimes disparagingly referred to as *ignorance engineering*) has meant that, even without knowing a great deal about the nature of speech, it has been possible to provide engineering solutions to the problem of word recognition. This is to the extent that it is now possible to apply this knowledge to the design and construction of reliable, practical automatic speech-recognition systems.

The speech signal

Let us first look at some examples of real speech. Signals captured by a microphone and turned into an electrical signal are shown in Figure 1.

The four fragments of speech are from the same speaker. The first two traces are of the word 'two' and the second two are of the word 'three'.

There are several key points:

- each utterance of the 'same' word, while similar in overall appearance, differs significantly in detail; and

Figure 2 (a)

Figure 2 (b)

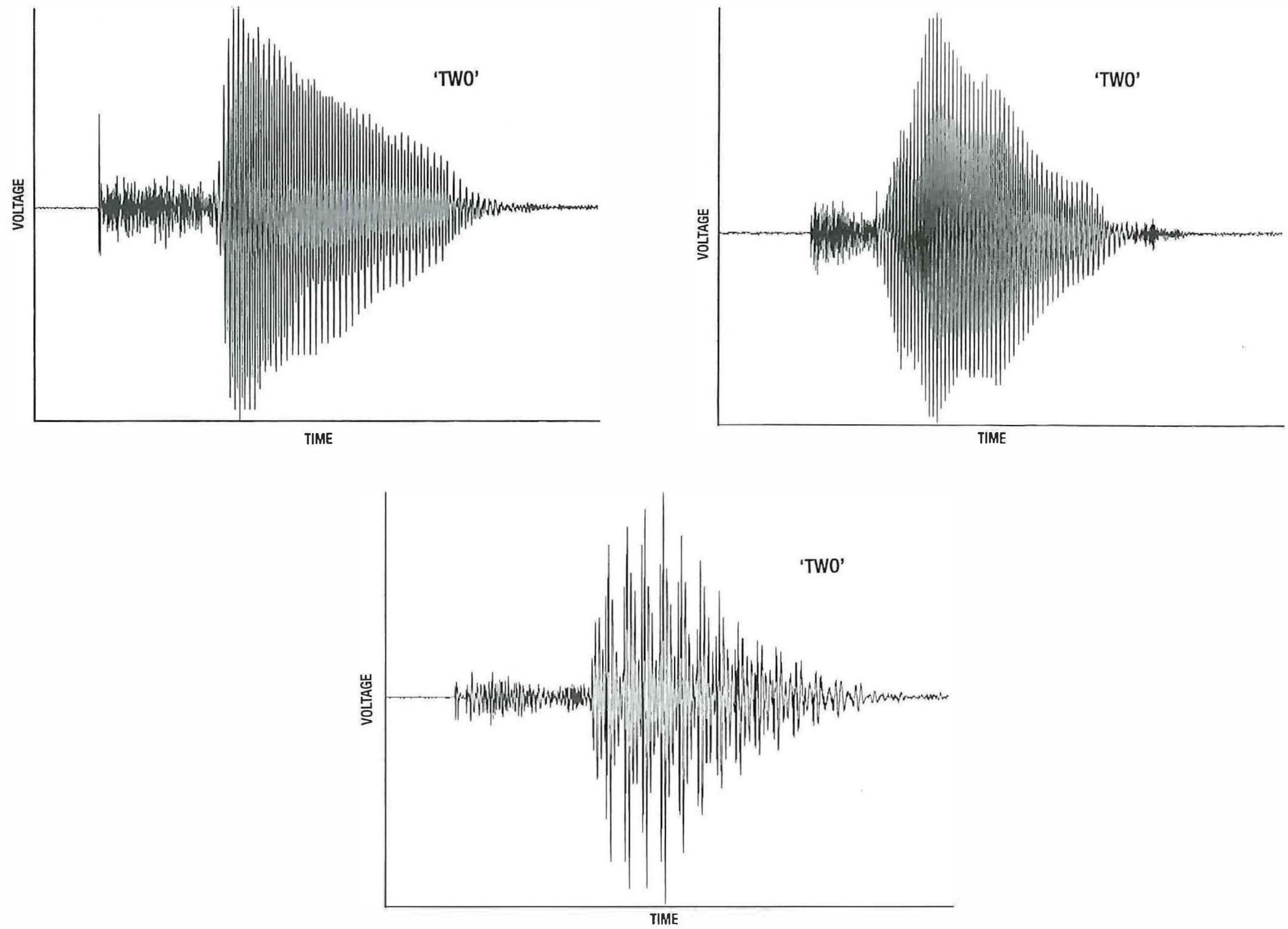


Figure 2—Three speakers saying the same word 'two'

- 'different' words (here 'two' and 'three') do not appear to have many more differences than repeated utterances of the same word ('two' and 'two').

So, even with the same talker speaking the same word, the signal is different and with different speakers the differences are even greater. Figure 2 shows waveforms of three people saying exactly the same word. Tantalisingly, although again there are many 'gross' similarities clearly at the level of fine detail the signals are very different. Indeed, it is almost impossible to see any parts of the similar utterances that are strictly identical in any way.

It is this intrinsic variability and non-repeatability of the signal which makes automatic speech recognition so very difficult.

Connected/continuous/fluent speech

The waveform in Figure 3 shows a sequence of words—the number 'two-six-two'—spoken in a normal speaking voice.

Possibly because of our familiarity with written text, we tend to imagine that speech consists of a series of words clearly separated by silence. Unfortunately, as this diagram shows, this is an illusion. It is clear that the words are not separated by silence and, just to add to the problem, there are often very clear silence regions right in the middle of words. In this example the most noticeable silence (in the middle of the word 'six') is particularly spectacular.

It may also be noted how very different are the signals of the word 'two' at the beginning and end of the utterance even though they have been

Figure 3—Speaker saying 'two-six-two'

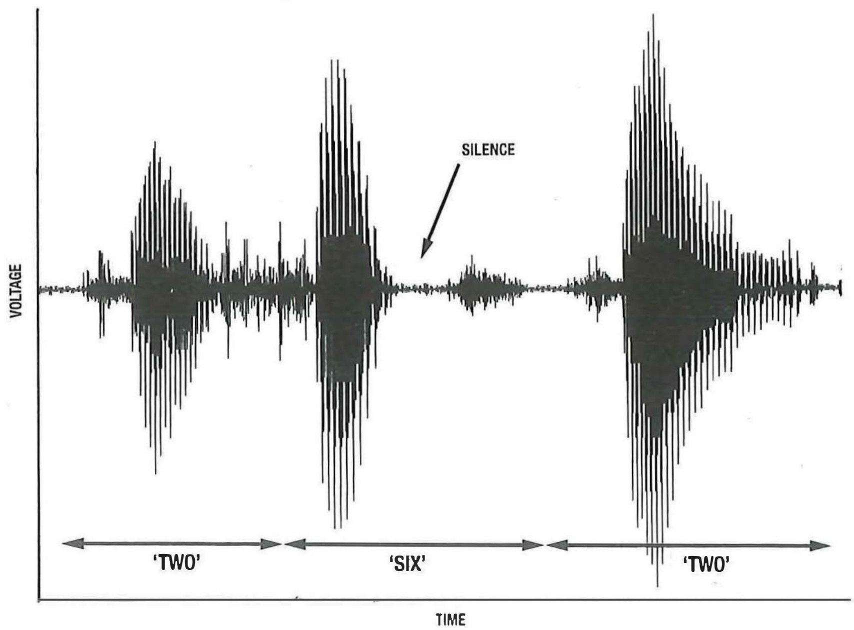
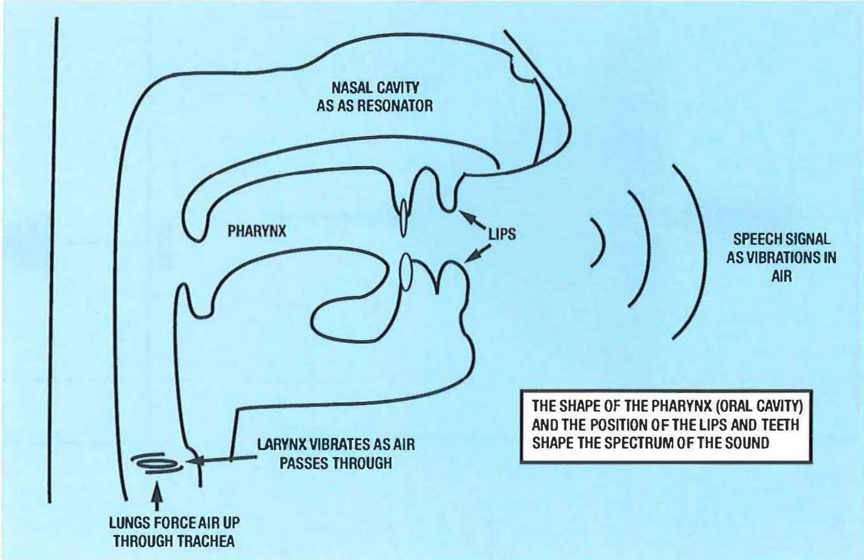


Figure 4—Speech production



spoken by the same speaker in the same breath. This illustrates the effect of context upon an utterance: adjacent words and the position of a word within a longer utterance are factors which affect the particular realisation.

Overall, it is clear, and rather disappointing, that unlike written text, speech does not consist of strings of clear letter-like symbols separated by spaces. In addition, although patterns and trends can be seen within the signal, it is apparent that these are difficult to describe and analyse in a rigorous manner.

Other factors

So far we have seen that words which we perceive as being the same may actually be very different signals. Other factors such as accents also introduce gross differences. This is to the extent that a word pronounced in one way in one accent may be a different word in another accent. For example, the word 'cart' is often pronounced 'cat' in Southern English.

In addition to these 'natural' sources of variability, there are also those introduced 'unnaturally'. In particular, a telephone channel introduces more variability due to:

- attenuation,
- non-linear coding methods (including non-linear microphones),
- bandwidth restrictions,
- effects of 'close speaking' microphones,
- line noise, and
- pyschoacoustic effects.

Human speech and hearing

Regardless of the great variability in speech signals, there must be something in the signal which remains substantially unchanged. We know this because we, as humans, are able to make sense of these signals quite easily. Even after they have been grossly distorted by transmission systems we can still understand the meaning of the words and the characteristics of the speaker's voice. To gain an insight into what these mechanisms might be, the vocal tract, the ear and their associated nervous systems have been extensively studied and it should not be surprising to learn that many of the techniques used in speech recognition have been inspired by our attempts to understand the mechanisms of human speech.

Next we examine how humans produce and analyse speech.

Speech production

Figure 4 shows how speech is produced by the lungs, larynx and vocal tract. The lungs force air through the larynx causing it to vibrate. This process 'chops up' the air flow in a manner similar to the way the reed on a musical wind-instrument produces a note.

This strong signal not only contains the basic frequency of vibration, which we normally perceive as the 'pitch' of speech, but also its harmonics.

It is the job of the articulators—the jaw; tongue; lips, etc—to shape the range of frequencies produced by this source signal by variously adding resonances and damping out frequencies. Acting together, all these parts of the system produce a signal of high acoustic energy capable of carrying a lot of information. This is the speech signal radiated from the lips.

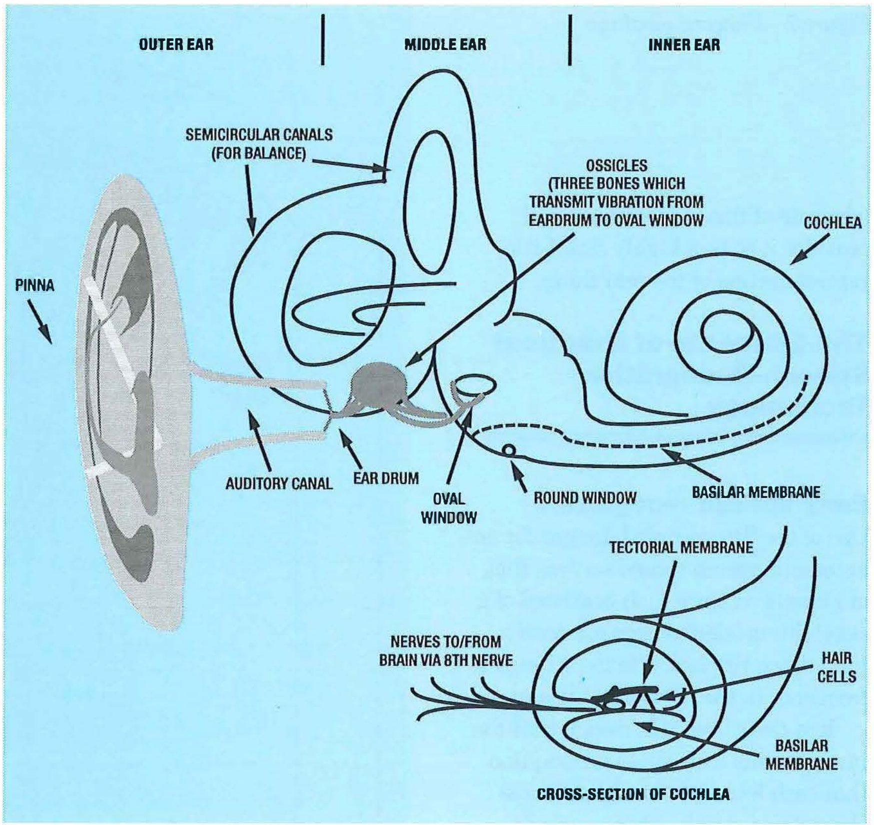
The dimensions of the articulators, vocal tract and larynx vary from person to person and it is these variations which combine to give each individual his or her own voice. In addition to these physiological constraints, the manner in which speakers control their articulators contributes characteristics that we acknowledge as accent, tone and personality. All of this means that, although the basic mechanism appears simple, there is only a remote possibility of a particular signal ever being identically reproduced, even by a single speaker repeating a word.

It is useful here to consider the parallels between human speech and musical tunes. Just as a melody can be played on a wide range of musical instruments, in many different keys with different tempos and loudnesses and yet retain its uniqueness, so spoken words retain their meaning no matter with which vocal instrument they are produced, how loud they are or how quickly or slowly they are uttered. It is the relationships between the notes, rather than values of absolute pitch, amplitude or frequency, which are important. Similarly in speech it is the relationships and contrasts between and within sounds, rather than their absolute values, which carry the message.

The ear

Whereas the mechanics of speech production are reasonably well understood, because the vocal tract is both large and accessible, the same cannot be said for the ear. Figure 5 shows the main parts of the ear and shows how sound is funnelled from the outer ear to the eardrum. The vibrations on the eardrum are transmitted to the inner ear (the cochlea) via three small bones

Figure 5—The ear



the last of which is fixed to a small membrane (the oval window).

In signal processing terms the mechanisms of the outer and middle ear are fairly well understood. Their function is to channel as much sound as possible by collecting it and then impedance matching it to the liquid medium that transmits sounds through the inner ear. Much less however is known about the mechanism of the inner ear, the cochlea, itself.

Not only is the cochlea extremely inaccessible, being a very small organ buried deep within the skull, but it is extremely delicate to the extent that it is capable of detecting movements approaching atomic scales. Indeed, the ear is sensitive to an amplitude of vibration which is the size of a hydrogen molecule[1].

The main transducers which detect the movement within the inner ear are rows of tiny hairs (stereocilia) arranged neatly in 'W' shaped ranks along the basilar membrane. Each group of hairs (typically 40–70 in a group) is associated with a cell and each of these hair-cell units seems capable of translating movement of surrounding fluid into pulse-frequency-coded nerve impulses which are transmitted to the brain along the eighth nerve. Studies regarding the responses in these nerves generally suggest that the cochlea splits the signal into frequency components. However, there is evidence that particular features in sounds (for example, rising or falling tones) may also trigger particular nerve firing patterns, and it appears oversimplistic to model the cochlea as a bank of linear filters[1, 2].

Psycho-acoustic experiments

Although physiological studies of the ear are likely in the long term to provide the most comprehensive theories of hearing, many of the most useful theories of hearing have actually come from a related but very different area—psycho-acoustics. Here, carefully controlled sounds are introduced into the ear, and responses (normally subjective) are recorded.

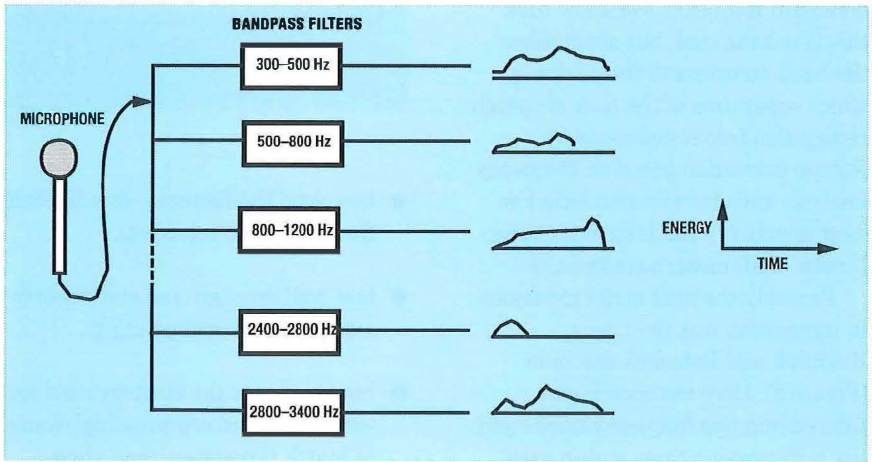


Figure 6—Electrical analogue of the ear

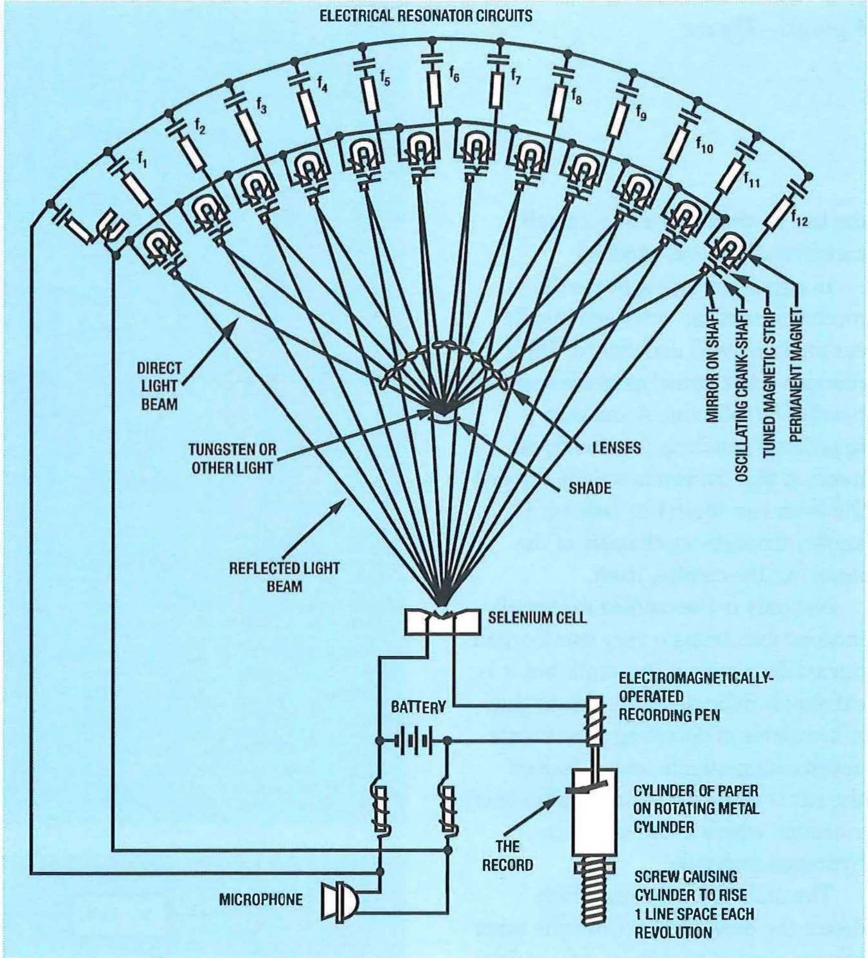
From such experiments we can infer much about how we perceive signals. For example:

- loudness is perceived as logarithmic,
- the ear is 'phase deaf' (at least over short time intervals),
- the ear is not uniformly sensitive over all frequency bands, and
- a variety of 'masking' effects occur. For example noise is masked in the presence of a signal.

However, even these apparent 'laws' do not appear to hold true at all times and it is possible to find many counter examples of all of these. For example, although for most purposes each ear appears to be 'phase deaf', we can resolve very short time differences between ears. This evidence indicates that the ear must be able to resolve very fine phase information.

Nonetheless, the general psycho-acoustic 'laws' outlined above have proved very useful for many practical purposes (for example, speech coding, vocoders, etc.). Figure 6 shows one commonly held view of how the ear operates and this model is the one that is generally used to justify the techniques of feature extraction processes used in many present day recognisers. However it should be stressed that even the most ardent

Figure 7—Flowers' machine



advocate of this model would still consider it to be a highly simplistic representation of the real thing.

The Elements of Practical Speech-Recognition Technology

Early speech recognisers

One of the first recorded designs for an automatic speech recogniser was that of Flowers[3] (Figure 7). It consisted of a set of filters (electromagnetic resonators) which responded to the strongest frequency in the particular utterance.

It is clear from this design that the inventor was making an assumption that each letter of the alphabet was characterised uniquely by a single dominant frequency. We know that this is not the case, but nonetheless the basic structure of the device in which separation of the task of speech recognition into components by feature extraction based on frequency analysis and classification based on 'best match' is a fundamental characteristic of all modern recognisers.

Probably the next major milestone in recognition was the Davis, Biddulph and Balashek machine (Figure 8). Here the speech was filtered into two frequency bands and the zero crossing rates within each band used as the primary features[4].

The first implementations using digital computers typically had configurations as shown in Figure 9.

The first speech recognisers operating along these lines appeared in the 1960s and were made possible by the advent of the digital computer as only digital technology provided the stable and accessible memory which was necessary.

In the recognition process the 'unknown' sequence of spectral components was aligned in turn against each known sequence in the 'template' library. That which fitted best was selected as being the 'correct' example.

The performance of these machines was determined mainly by the following factors:

- how good the features were in each frame (feature selection),
- how well the start and end of words were detected (end pointing),
- how good was the strategy used for 'stretching and compressing' words to match templates (time alignment),
- how representative were the templates used for training, and
- computational power available.

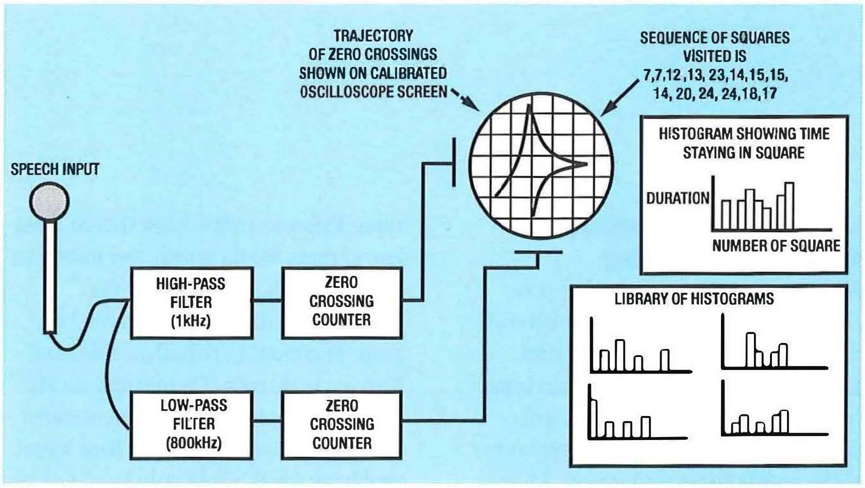
Computational restrictions meant that only very simple processing could be used at almost every stage, but an even greater problem was due to the sequence of operation. The following was a typical sequence:

1. Detect the start of speech.
2. Collect speech until silence was detected.
3. Normalise the speech to a fixed length.
4. Start the matching process against all the reference templates in turn.
5. Report the result.

The need to wait until the end of the speech before any normalisation could be done meant that the matching process, which represented the greatest computational load, could not even start until the end of speech was detected. Typically, 200 ms of silence had to have elapsed to be certain of the word end. This meant that the processing could not start until well after the word had finished and the final result would be much later still. Indeed much of the development effort directed at improving these machines was concerned with improving the response time.

However, the application of a new mathematical approach to non-linear pattern matching was to revolutionise the whole process to the extent that post-utterance processing delay could be completely eliminated. The technique was known as *dynamic programming*.

Figure 8—The Davis, Biddulph and Balashek analogue recogniser (1952)



Dynamic programming

Dynamic programming (popularly referred to as *dynamic time warping*) simultaneously overcame almost all of the theoretical, although not all of the practical computational problems. Figure 10 shows how the process works. For clarity this example shows spectral frames derived from a four-channel filter bank, but in practice many more channels, typically 16–20, would be used.

Instead of having to wait for the end of an utterance, the distance of every spectral frame in the input from every frame in every template was calculated as it arrived and the accumulated difference along the best path maintained.

It offered the following advantages:

- Matching could take place as the signal arrived.
- Processing could be fully pipelined.
- The very best time alignment possible between the unknown signal and the known templates was obtained.
- Recognition was complete as soon as the word was finished. Indeed, if there was sufficient discriminative information in the initial sounds of the vocabulary words then the word could be recognised even **before** the speaker had finished uttering it.
- Connected word recognition was achieved as a natural extension to the process, almost as a by-product.
- Connected word recognition also meant that word spotting was possible.

It is not surprising that dynamic programming was hailed as the greatest single breakthrough in speech recognition. This single theoretical technique appeared to provide a solution for almost all the technical problems in speech recognition and formed the basis for a new generation of practical speech recognisers.

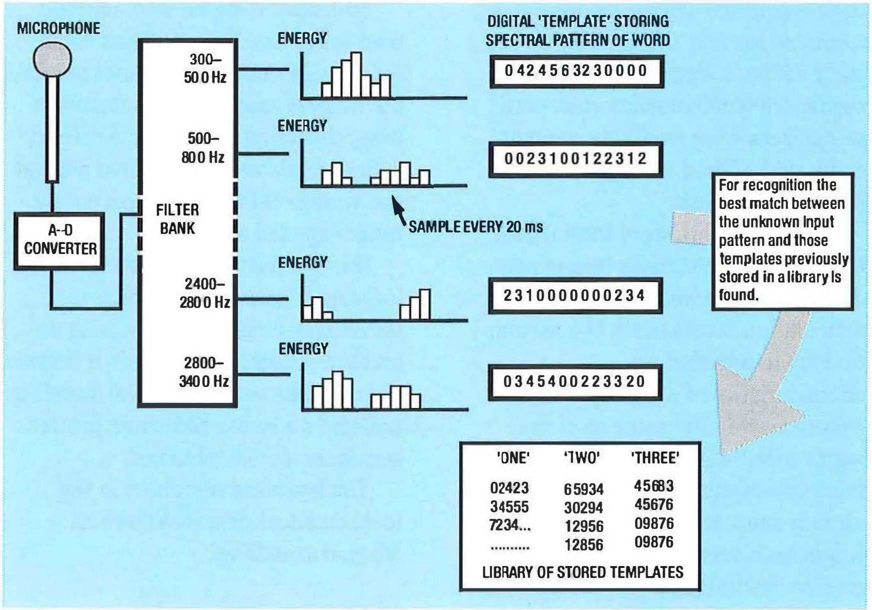
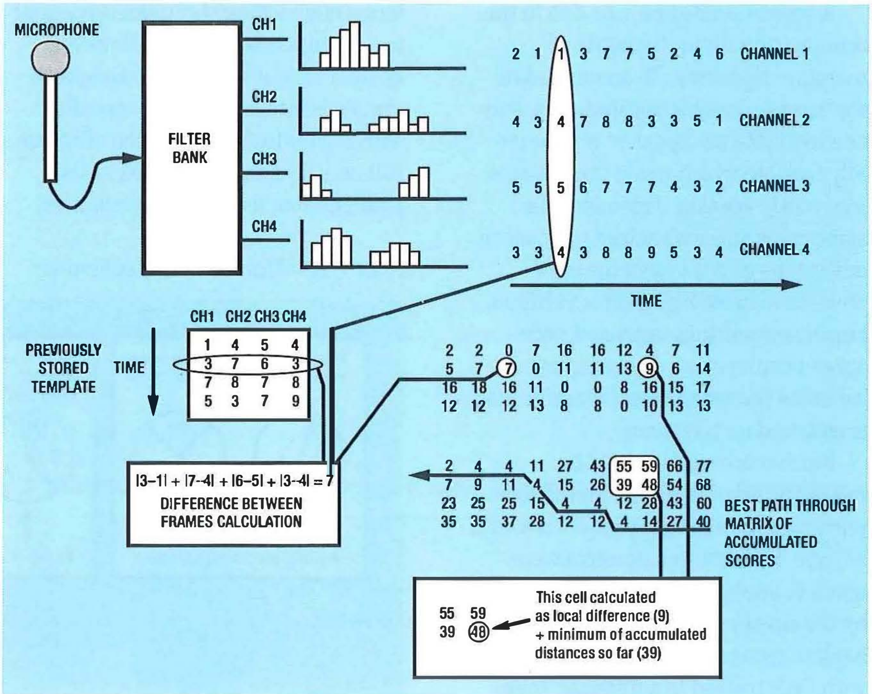


Figure 9—Recogniser using digital storage

Figure 10—Dynamic programming principles



Limits to template-based dynamic processing

While dynamic programming provided the solution to the problems of time alignment, endpointing and concurrent processing, most systems relied upon templates being sufficiently representative of all variants of the words to be recognised. This approach worked reasonably well for those situations where the user was available to train the system. By the early 1980s, a significant number of speaker-dependent connected word recognisers were available commercially that offered vocabularies of hundreds of words.

However, there were limitations. One problem that these recognisers did not address was that of co-articulation. Essentially, the assumption made was that the characteristics of words spoken in sequence were the same as of the words uttered in isolation. This was a false assumption because, as we have already seen, the nature of a speech fragment is strongly influenced by its context. Techniques such as 'training in context' or 'embedded' training, where the templates were made from words excised from strings of words, were developed to allow for this with limited degrees of success.

A second limitation was due to the computational requirements of template matching. To accommodate the largest possible vocabularies, only one template per speaker was generally feasible which made the systems inherently speaker dependent. In some cases this was seized upon as an advantage and the systems were advertised as being speaker verifiers, but in general this restricted widespread deployment. In particular, it excluded the vast potential application area based on telephony.

Much more useful would be recognisers that were not 'tuned' to a particular speaker but would work with anyone. For a period attempts were made to provide speaker independence by the simple expedient of having a bank of speaker-dependent recognisers with each trained to a different voice type. This was in the hope that at least one of them would match any unknown speaker. Such an approach was computationally very expensive, but, more importantly, raised questions of how to create sets of templates which could best represent large populations of speaker characteristics. These were problems which could only be solved by turning to statistical methods.

Statistical methods were initially used to improve the templates: sometimes using multiple templates to span the range of speakers and sometimes using clustering methods to form and define templates that captured not just the average values of features but the range expected as well.

However, just as the mathematical technique dynamic programming provided an elegant way of solving the problem of aligning sequences, it became clear that the use of statistical modelling provided a way in which much greater generality could be obtained.

The two were combined in the mathematical process known as *Markov modelling*.

Markov models as a mechanism for speech production

So far the idea behind speech recognition has been described mainly in terms of matching the 'unknown signal' to a set of 'known' signals. However, another way of thinking about speech signals is intuitively more appealing with regard to the production of speech. Not only does this let us look at the problem from a different perspective, but it also provides a more general mathematical framework.

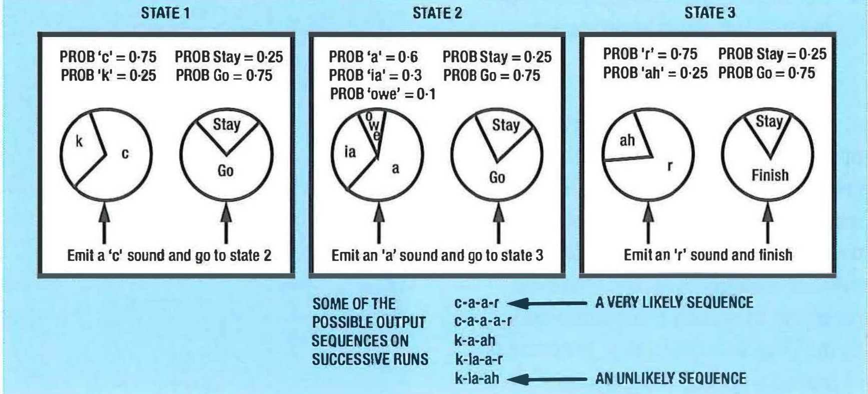
Rather than thinking in terms of matching sound patterns directly, it is necessary to think in terms of 'What production model was most likely to have generated this sequence of sounds?'.

Previously in this article the mechanism of human speech production was shown to be a process in which sound was produced by virtue of a set of articulators moving from position to position in a stepwise manner. One way of modelling this is to model speech as a sequence of sounds produced by a state model. For example, in pronouncing the word 'car' the first state is a 'c' sound generator, the second state is a vowel generator normally producing the 'a' sound (or possibly an 'aw', 'ia' or 'owe' sound) and the final state is the 'r' or 'ah' sound generator.

What is interesting about this representation is that by adding probabilities to the various state transitions we can create a machine which is capable of producing sound strings with variabilities in sound quality and duration that look plausibly similar to those of human speakers.

One very simple way to imagine this is to think of each 'state' as containing a pair of wheels similar to roulette wheels. One of these wheels contains a range of possible outputs in the ratio in which they actually occur in the real world. The second wheel simply has two segmentations: 'stay' or 'move'. Figure 11 shows a three-state model which is

Figure 11—Markov model generator



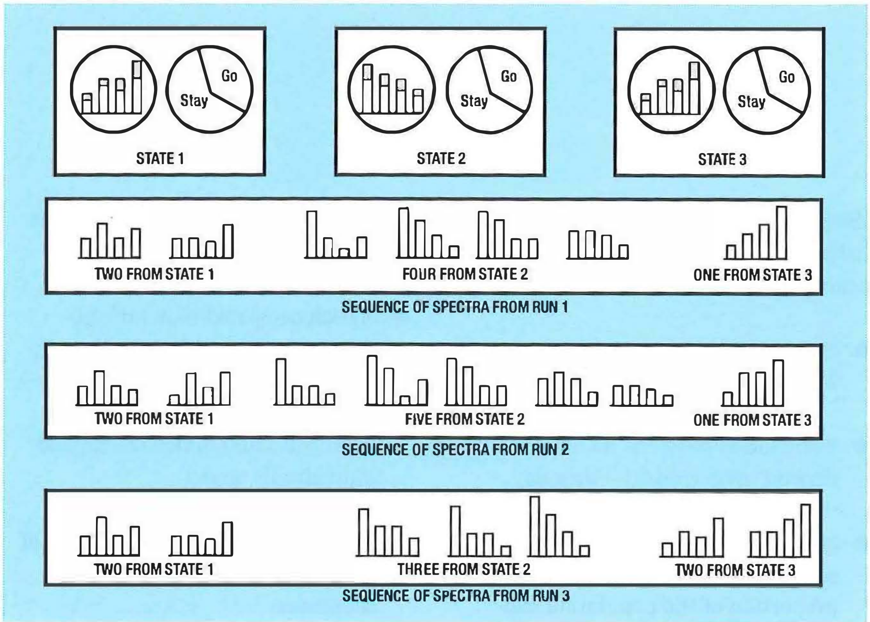


Figure 12—Markov model using spectral features

capable of producing a vast number of different, yet similar, sequences.

The operation is as follows. A 'run' starts by spinning the first wheel and it stops at a segment marked 'c', which is the first symbol to be emitted. Then spin the second wheel associated with state 1; this says 'go', so we move to state 2. Had it stopped at 'stay' we would have remained in state 1 and respun the first wheel to obtain the next symbol. This process continues emitting 'symbols' according to where the first wheel stops and moving to the next state when the second wheel lands on a 'move' segment. In this case for example the sound sequences 'caar', 'kah', 'kiah', 'cower' are all plausible patterns which can be produced from the model.

If it were possible to find a way to select the model probabilities such that the range of output strings matched those spoken by the total population then clearly it could be argued that this simple model had captured the essence of each word.

The fact that such a simple model is capable of producing sequences with properties similar to those of real speech is very attractive. Not only does it deal with variability in time and sound type, but it also provides an extremely compact representation which, in turn, means that computation can be very fast indeed.

Speech recognition using Markov models

Real speech, of course, does not consist of discrete letter-like sounds, and most recognisers rely on the assumption that speech is actually a sequence of spectral frames that change fairly slowly.

Such a model is shown in Figure 12. This shows pictorially the means and variances associated with each state.

In a typical recogniser, each word in the vocabulary would have its own model associated with it and the recognition process works by finding which model best explains the actual sequence of sounds.

In practice, a search technique similar to that of the dynamic programming process above is used to simplify the recognition.

Training of hidden Markov models

The usual way to train a Markov Model recogniser is to use many examples of each word from as representative a set of speakers as possible. Typically, over 500 utterances from speakers with different accents, different moods, different telephone lines and instruments are used in training telephone-line-based recognisers.

The basis of training is now described.

For each word in the vocabulary:

1. Start with a set of states (say six).

2. Partition every example in the training set into six equal time segments. The corresponding segment of each word will be used to set the initial estimate of statistics for each state.

3. Average all the spectra for each segment and calculate the variances for each spectral feature within each segment. Allocate the values from segment 1 to state 1, values from segment 2 to state 2 etc.

4. Set the transitional probabilities for each state to be 0·5 stay and 0·5 move.

This process provides an initial guess as to what the values and distributions should be. The next stage is a 'hill climbing' exercise. Here the aim is to improve the allocation of frames to individual segments. One way of doing this is as follows:

1. By using dynamic programming, match every word in the training set against the existing model and record which frames fall into each state.

2. Use this new segmentation to update means and variances for each state. Also use it to re-estimate the transitional probabilities.

This process can be repeated several times either until a limit is reached or until no further improvement is found.

This process of re-estimating the values generates a model which captures the statistics of data presented to it. For more details of this and other training techniques see Reference 5, 6 and 7.

Limits to current technology

Although the technique of hidden Markov modelling is generally considered to be the most powerful technique available to date, it is important to examine where its limitations lie. This is particularly important from an implementation perspective as knowledge of these limitations will allow sensible choices to be made concerning the vocabulary used or the way in which training data is collected.

As it is these assumptions which, in the end, are going to set bounds to the performance that can be achieved, listing some of the more questionable assumptions is useful:

- real speech originates from a Markov process—it doesn't;
- word sound patterns do not depend upon context—they do;
- training data is always representative—in practice only a tiny proportion of the population can ever be sampled;
- training in a 'standalone' fashion is adequate—there is not usually any attempt to exploit those features which discriminate between words; and
- Markov models generalise appropriately—they don't.

Other types of speech recogniser

Most of this article has concentrated on the mainstream of progress in practical speech recognition and has followed a thread leading to the current methods used in telephony. However, there have been other approaches which for completeness are described here. With the arguable exception of artificial neural network approaches, most are now recognised as actually being only special cases of the statistical Markov model approach.

Knowledge-based systems

During the 1980s there was a lot of activity in the field of artificial intelligence and not surprisingly a great deal of effort was directed at the problems of speech (and vision) recognition, mainly because these were areas that had not proved tractable to other means. The assumption made was that speech was produced and understood through a finite series of simple logical rules. As computers were excellent at following such rules, it was only necessary to capture these rules and the problem would be solved.

Specific assumptions made about speech were:

- speech consisted of invariant 'atomic' components (phonemes),
- phonemes could be described by fixed rules and hence recognised individually, and
- rules determined the sequences of phonemes to make words and sentences.

The difficulty was that neither rules nor phonetic units that were simultaneously simple and general could be found.

To a large extent, Markov modelling provided the solution and bridged the gap by relaxing the need for invariant components. In Markov models the 'rules' are encapsulated as probabilities and the traditional 'phonemes' are modelled as distributions of sub-word units. Today a number of large-vocabulary recognisers combine phonetic representations of words with the probabilistic models of sounds to get the best of both worlds.

Artificial neural networks

As has been indicated, many of our speech-recognition techniques have been inspired by physiologically based models, and research into artificial neural networks represents an attempt to extend this type of modelling further. What is particularly interesting about neural-network-inspired models is that they 'naturally' have a certain degree of discrimination built into them: they use the difference between classes as well as the similarities within each class to characterise patterns. While this property is by no means unique in pattern recognition (and it is possible to apply it to Markov models as well), it is such an emergent property of the neural-network technique that it is difficult not to believe that it is one of the key processes in human speech recognition.

The appeal of neural networks is that they provide a different, possibly broader, perspective to the whole problem of pattern recognition, and, just as Markov models are now seen to have been a generalisation of dynamic programming, so certain neural network architectures may come to be seen as a generalisation of Markov models where the structural constraints imposed by the state model are relaxed.

In neural network models, the performance achieved is determined entirely by the architecture and learning strategy of the neural networks as it is this which finds the key recognition parameters. At present, it is by no means certain that our current models of neural networks are sufficiently close in either structure or scale to those of 'real' neural networks to justify any of the assumptions required. On the other hand, even very simple neural networks tend to exhibit useful pattern-recognition characteristics, and such circumstantial evidence strongly suggests that there may be a lot to be learned by following this route. More details can be obtained from Reference 8.

Implementation issues

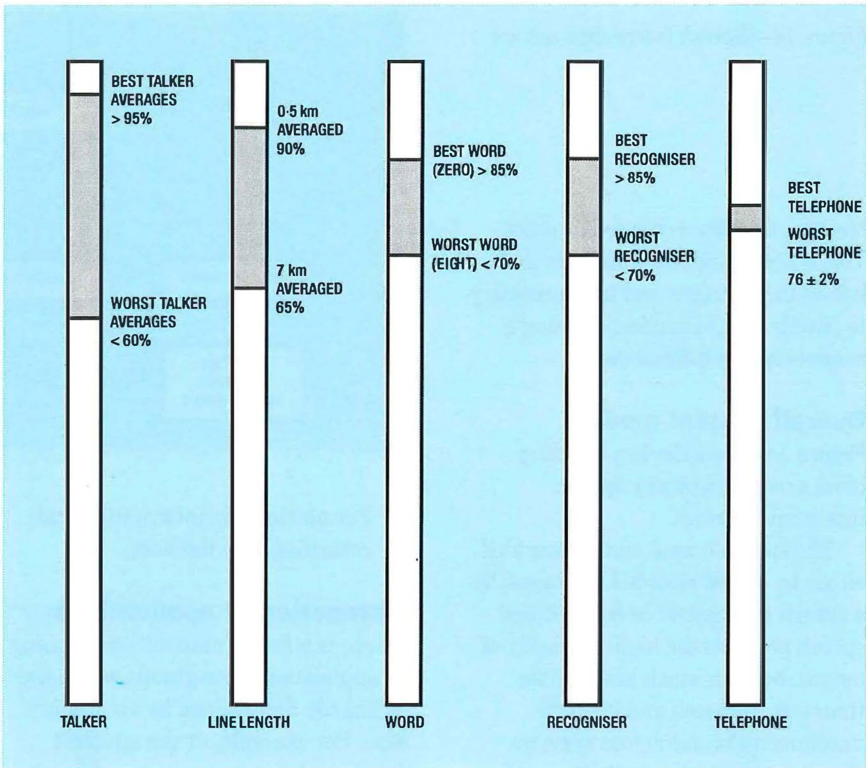
The performance of speech recognisers

The implementation of all speech-recognition systems is dominated by one single factor—the fact that speech recognisers make errors. What is not immediately obvious is that there are several types of errors:

- substitution errors—the recogniser responds to a valid input but gets it wrong;
- insertion errors—the recogniser responds to an <u>invalid</u> input (for example, noise, spurious sound or an invalid word); and
- deletion errors—the recogniser fails to respond to a valid input.

All of these error types can be traded off against each other in any

Figure 13—Experimental results showing range of accuracies for different conditions



application. For example, it is possible to reduce the number of substitution and insertion errors by accepting only those matches which are really good. The penalty for doing this is that many more correct responses will be falsely rejected.

It is particularly important for applications engineers to be aware of the interaction between these error types because it is possible to achieve an extremely good substitution error rate by judicious selection of the acceptance thresholds, only to discover later that this apparent improvement has been bought at the expense of an unacceptable deletion error rate.

Measurements

A question frequently asked of a speech-recogniser provider is 'How accurate is it?' Unfortunately, to specify the performance of a speech recogniser in this general way is really quite meaningless because any accuracy figure depends upon the speaker types and their skill, the vocabularies used, the test environment etc. A recogniser which delivers 100% accuracy on one vocabulary with one talker may deliver only 15% accuracy with a different vocabulary/talker combination.

This is not to say that measurements of accuracy cannot be used to make a meaningful comparison between two or more recognisers, and the normal way in which this is done is to use statistical methods which test to see if one recogniser is, on average, better than another over a wide range of representative speakers, telephones, networks and word types.

Figure 13 presents the results of an experiment in which different (digit) recognisers were compared within several environments. The recognisers tested included those which have been commercially available in recent years.

The results show that the single factor having the greatest impact upon the performance of a recogniser is the speaker's voice. This was the case for all recognisers, although some were less variable than others. The telephone line was the second biggest factor with performance deteriorating as lines got longer. Words and recognisers came next with the telephone instrument being less important. This experiment also used two (relatively quiet) levels of background noise and found that these were almost insignificant. This does not imply that background noise is unimportant, but merely indicates that in normal quiet environments a small variation in background noise (10 dB) is less important than the other factors.

Telephony-based services and applications

Telephony applications currently account for the greatest area of growth for speech-recognition applications. However, the integration of speech-recognition into networks and particular applications can pose formidable problems and the specification of a speech recogniser depends upon where in the communications link it is positioned and how it must interact both with a host computer (for example, a database) and the network infrastructure.

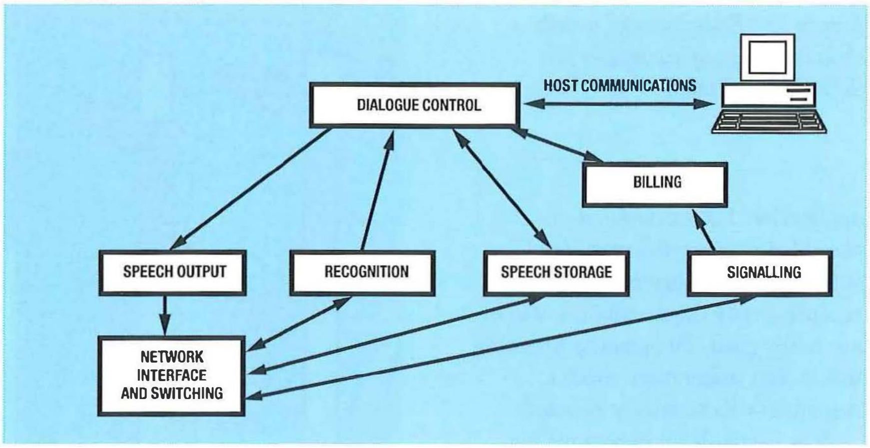
Typical services which employ speech technology are given in Table 1.

To address this range of applications, the BT Speech Applications Platform

Table 1 Typical Services Employing Speech Technology

Voice Messaging
Finance telephone banking stocks and shares insurance quotations credit card transactions
Entertainment betting horoscopes games competitions
Information Services timetable yellow pages
Telemarketing promotions market research
Teleshopping/Reservations theatres airlines mail order
Field Operations data entry and retrieval field personnel job dispatch voice access to electronic mail
Automatic Operator network services customer premises

Figure 14—Speech interactive service



has been developed over recent years. This is a network-based platform that has all the interfaces and tools necessary to provide a rapid application using a range of speech technologies.

Overall system model

Figure 14 shows the key building blocks required by any speech interactive service.

The speech output can be provided either by stored recorded messages, by a speech synthesiser or both. Stored speech provides the highest quality of output, but it is much less flexible than text to speech and it is not uncommon to come across systems that combine the best of both worlds by using recorded messages where possible and deploying text to speech to provide flexibility when required.

The role of the dialogue controller is central to the whole system. It has the following functions:

- Interrogating the host computer to find which item of information is required next.
- Deciding upon the best strategy for prompting the user. This will depend both upon the requirements of the application and its own experience of the current user and environment.
- Configuring the recogniser. This will involve sending messages defining the parameters to use; for example, vocabulary content, grammar linkages (for connected words), time-in and time-out durations and rejection thresholds.
- Monitoring and interpreting the messages returned from the recogniser. Typically the message will consist of an ordered list of words and associated probabilities.
- Entering a 'macro-dialogue' either to check that the word recognised is correct or to prompt for a repeat. This will generally require that the recogniser is reconfigured on each cycle.
- Formatting the information and returning it to the host.

Categories of applications

There is a fairly 'natural' partitioning of applications into groups which are primarily determined by vocabulary size. For example, at the simplest level (such as answering machines) a vocabulary of a few words—for example, yes/no—is sufficient.

At the next level, a wide range of applications requires digit entries. Many modern telephones provide TouchTone® signalling, a fast and reliable method for entering digits which to a large extent removes the need for speaker-independent isolated digit entry for simple systems. However, reliable spoken connected digit input allows much more flexibility and gets over the 'formatting' problem which arises when there is a need to use TouchTone® to input more complicated strings of numbers such as amounts of money. To ensure that a figure like 'two hundred and thirty five pounds twenty seven pence' is correctly entered using TouchTone® is much more difficult. For such tasks, speech recognition becomes a much more attractive proposition.

There is also a wide range of applications where other (non-digit) small vocabularies can be used. In general, however, the problem in deploying these is in finding ways of ensuring that the user will know (preferably without requiring any guidance) what to say.

Examples of the types of 'spontaneous' replies that will work are:

- yes/no,
- days of the week,
- months of the year,
- place names, and
- types of service ('divert all calls', 'transfer'..),

However, care must be taken in phrasing the prompt. For example, imagine a railway timetable service. A vague prompt such as 'When do you want to travel?', will generally elicit a response like 'Can I go the day after tomorrow and come back two days later'; whereas 'On what day do you wish to start your journey?' is more likely to elicit a single reply such as 'Monday'. Even more effective, although less friendly, is to say 'Please state on which day you need to start your journey'; this may help avoid replies such as 'Monday please'. BT has been developing its own style guide to give its services a unique 'look and feel' and which encourages good practice in this area[9].

Location of the recogniser

Depending upon the application, recognisers may be located anywhere in the network. However, the signal characteristics, particularly in terms of bandwidth and the range of levels present, are so highly variable that different recognition strategies and training regimes need to be employed to ensure the best performance. Table 2 gives the main areas where recognisers can be deployed.

A Glimpse of the Future

The history of speech recognition is littered with false starts and unfulfilled promises. However, there is now substantial evidence that applications are appearing in significant numbers.

Table 2 **Main Deployment Areas for Telephony-Based Speech Recognisers**

Recogniser	Performance	Signalling Issues	Cost	Functionality	Feedback Form	Comments
At telephone or cellphone	For repertory dialling. Cost/power issues may compromise performance	Limited to standard network interfaces	High per unit cost with no sharing of capacity	Digit dialling and repertory dialling	Audio or LCD display on telephone	A number of products on market
On office workstation	No network transmission loss and opportunity to use high-quality local microphone gives good performance. However, incoming speech quality will be worst case	Network control limited to those available over ISDN or MF4	Per-unit costs are not spread over several users and are therefore high. However, hardware may be used for other things as well	May include non-telephony applications and local end speaker adaptation. Answering machine functionality possible	Visual or spoken	Likely to grow with ISDN penetration
At PABX	Generally insignificant local network losses give good near-end recognition. Incoming speech quality will be worst case	Integration with switch allows local control of telephony system	Spread over many users	Speaker dependent or adaptive recognition possible	Speech or fax only at present	Includes Centrex types of applications
Network switch	Allows many users to share the same unit. Near and far ends subject to similar recognition performance	Signalling can be fully integrated. All functions available	Interfacing must be carefully controlled to ensure no danger to switch	Many centrally managed applications possible. 'Pay as you use' opportunity	Speech or fax only at present	CALLMINDER is an example

The first most significant change in accelerating the spread of services will be user familiarisation. Just as the initial trepidation of bank customers in interacting with cash dispensing machines instead of a human cashier has now reached the stage where many customers prefer the machine-mediated interaction, it is certain that a similar phenomenon will occur with speech input. Not only will customers become familiar with the technology, but they will develop their own styles of interaction.

When this happens the applications base for speech recognition will change radically. A secondary effect is likely to be the emergence of sub-languages for interacting with machines. These will be languages specifically tailored for people–machine interactions.

The second change will be in the area of cost reduction. A decade ago even the cheapest marginally-usable speaker-dependent recogniser cost more than £1000. Today, it is possible to buy very-high-performance limited-vocabulary speaker-independent speech recognition in a chip-set for a few tens of pounds. This trend will continue with the price/performance ratio continuing to improve.

The third change will be the deployment of language understanding methods to allow interactions with large databases. At present, most database query systems are accessed using proprietary query languages. In such cases, the recogniser must be constrained to provide outputs which fit the format of this sub-language. A really significant breakthrough would be the freeing of the database machine from the constraints imposed by the query language. However, it is difficult to see major progress here in the immediate future within the fairly rigid constraints of existing relational database structures.

Finally, despite the fact that speech recognition is moving into applications, the performance of even the best speech recogniser still falls short of anything that can be achieved by humans. Not only can we recognise many words in and out of context but we can hear and understand words in noisy environments, can extract a single thread of conversation even when surrounded by many other people talking simultaneously (the cocktail party effect) and can understand what is meant even if it does not correspond to what is said.

In the past few decades, the digital computer has meant that it has been possible to put to the test many traditional theories of speech and language. Many have been shown to be either erroneous or ineffective but we have now reached a stage where we know which ones work best and lead to practical solutions.

We are now in a situation where the barriers to better speech recognition are no longer those of inadequate supporting technology. Our greatest single need is better knowledge. For speech recognition to move forward again we must have a better insight into the fundamental processes of human listening and reasoning. In the meantime, many opportunities exist for exploiting applications of the technology for the benefit of customers.

References

1 GELFAND, S. A. Hearing—An Introduction to Psychological and Physiological Acoustics. Dekker.

2 MOLLER, A. R. Auditory Physiology. Academic Press, 1983.

3 FLOWERS, J. B. The True Nature of Speech. *Journal of the Acoustical Society of America*, 1916.

4 DAVIS, K. H., BIDDULPH, P. and BALASHEK, S. Automatic Recognition of Spoken Digits. *Journal of the Acoustical Society of America*, Nov. 1952, **24**.

5 WHEDDON, C. and LINGGARD, R. Speech and Language Processing. Chapman and Hall, 1990.

6 HOLMES, J. N. Speech Synthesis and Recognition. Van Nostrand Reinhold, 1988.

7 RABINER, L. and JUANG, B. Fundamentals of Speech Recognition. Prentice Hall, 1993.

8 RUMMELHART, D. E. and MCCLELLAND, J. L. Parallel Distributed Processing. MIT Press. Cambridge, Mass.

9 STENTIFORD, F. Now You're Talking....Voice Services. Issue 1, Dec. 1992, BT Laboratories.

Biography

Denis Johnston
BT Development and Procurement

Denis Johnston graduated in Electrical Engineering at Queen's University, Belfast, in 1971. After a spell in the aerospace industry, he spent two years teaching in Sri Lanka before returning to complete an M.Sc. in Electronics and Communications. He joined BT (then the Post Office) in 1974 and has worked in the field of speech technology since then. Much of his early career was concerned with the assessment of coding methods, channel-sharing techniques and echo-control devices for deployment in the telephone network. In 1983, he became head of a group concerned with the application of speech recognition to telephony products. He is currently a Senior Advisor in the Speech Applications Division primarily concerned with evaluation methods.

David Haskins

Spatial Information Paradigms for Strategic Analysis

Geographic information systems as a special case of meta-data in business information

Engineering and planning data in the telecommunications environment can be extremely useful to those making commercial decisions. The challenge is to be able to link this data to wider information sources to maximise its effectiveness. This article discusses the issues involved.

Introduction

The construction of information models of aspects of BT's current or potential business to support BT marketing strategists has identified a major fault-line or disjunction between information apparently useful and usable by professional telecommunications engineers, and information useful and usable by those charged with maintaining and increasing market share and developing new products and new markets. The illustrative power of *geographical information systems* (GISs) has acted as the lens which has brought this disjunction into focus and many of the examples cited in this article are therefore shown by maps and diagrams. The main point is that telecommunications engineers and planners can greatly assist BT not only by designing record systems about plant and customers from their own business viewpoint, but also by understanding how other information is in practice held about those same customers...who they are...what they do...where they live and go. Often this is collected outside BT in forms over which we have little control and an understanding of this wider information marketplace could transform engineering and planning data into more useful business information, and vice-versa.

Spacial Concepts

How and why we use spatial concepts is a fascinating problem for information systems designers, and is one which has only relatively recently come to the top of the heap of issues requiring our attention. This is probably because the technical characteristics of the devices which we can use to interact with information systems now allow us to actually see 'as in real life' how things look and where they actually are. This is of course implicit for real-time video links, but is beginning to be the case for stored information too as we store images, sounds, as well as descriptions and structured data for later use and analysis. We are able to begin to use 'where' ideas in our information handling processes in more and more realistic and natural ways.

The short history of geographical information systems may seem fairly confused but might be understood as having involved for two quite different strands of human information processing activity:

- the automation of map production, and
- the automation of navigation.

The resulting systems do not easily come together as they address quite different information problems, that of map-making and that of map-use, and it is becoming clear that, to say that spatial information handling has been successfully automated, these two activities must be understood as being sides of one coin. In considering this issue the reader should ask what is a map made for.

Other questions arise. Why map things at all? Why is one person's map simply some marks or symbols to another?

What is a Map?

The best answer is to give examples:

- a schematic diagram,
- a public transport grid,
- an aerial or satellite image,
- a routing algorithm,
- an A-to-Z city guide.

We could ask whether these share any common features: they all share the feature of being smaller or more manageable than the things which they purport to map. They are stylised simplifications used to aid cognitive explorative activity; they guide one more 'ignorant' through another's subjective interpretation of what exists; they contain or entirely consist of semantic clues which stylise some real-world feature to point it out or make it the obvious interpretation. The clues show the point behind the map, which is its intended use. The public transport grid irons out irrelevant topological complexity and shows where to get on, off, and change trains. The A-to-Z guide shows and cross-indexes named roads. The road atlas uses numbered roads and junctions. The Ordnance Survey 1:25000 Outdoor Leisure series maps show walking routes and passed sights very clearly.

What is clear is that when the map becomes as detailed (the same-size?) as what it is a map of, then it is not a map but a facsimile. Obviously this doesn't normally happen with physical artefacts other than dressmaking patterns or engineering jigs, but the limitations which seem obvious (you couldn't roll up a full-scale map of anywhere) in the physical case aren't so true when it comes to building information systems. The concept of scale doesn't really seem to apply when the data is in digital form—it might give a clue to relative accuracy should the data have been captured off a paper map, or give some sense of the sensitivity of the data-capturing camera, but theoretically a computer could contain a spatial 'model' which 'is as big' (detailed?) as the thing modelled. The obvious example which BT recent experience throws up is the question of when would a map of our network (if one were to zoom in continuously) suddenly transform itself into a schematic diagram? Should the repeated, generic data about the detail at one node be seen as spatial information at all, or should the idea of 'this is a map' stop at some point, and 'this is now a diagram' then start? While it may be an issue about the difficulty of current systems to integrate and handle spatial, qualitative and quantitative data, it is arguable that BT's past weakness in this area is actually about a fundamental conceptual confusion about maps and diagrams rather than anything to do with basic project skills.

So the kinds of questions being addressed are:

- Is a map a type of geographic information system, or a product of one?
- What might digital spatial semantics be like? We talk about 'towns', 'roads', 'nearby', 'my house', 'the local exchange', but can machines handle these ideas?
- Have any semantic systems yet been made—what are good examples of working spatial language which have been successfully computed?
- Are they complex mysterious runes or a rich open language?

Is it sensible or foolish for information scientists to postulate some universal digital spatial information base able to generate maps, rules, tools, views, aids, plans, whatever, for the endless list of human subjective purposes? Or is this some kind of classical early-Wittgenstein-style 'Tractatus Logico Geographicus' which may in time prove a melting vision—some real fundamental error? Indeed what **is** the intellectual function of believing this current dream of the GIS industry?

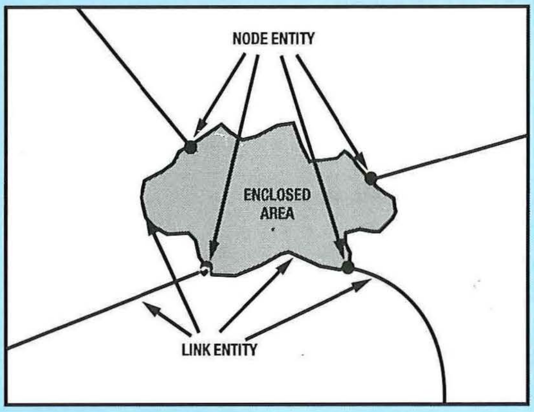
What is a Digital Map?

Building from the earliest work in producing reproducible, scalable, maintainable record systems able to drive large colour plotters and replace human draughtspeople, the information elements usually have included two basic types of information:

- *Vector coordinate records* for points which can be combined to form lines or closed shapes representing earth-surface objects. These follow real-world coordinate spaces like latitude and longitude, and can be mathematically transformed into projections such as the national grid which fulfil localised needs such as best-fit of the total area of interest on a standard paper dimension.
- *Raster bit-images,* which are either the print or screen image of all the many stacked layers of vector data, as well as filling colours and labels, or an actual scanned photograph from photogrammetric or satellite imaging, which can be manipulated as an entire digital object using rubber-sheet geometry and used as a kind of realistic 'wallpaper' against which to project vector layers which highlight features.

Early *vector* digitisation simply captured marks on hand-drawn maps and was essentially unconcerned about what they were marks of as long as the resultant record structure enabled the plotting of masters for the suitable colour printing process. This resulted in a split between

Figure 1—Spacial data entities




```
enclosedArea  ...(sometimes called Faces)
{
        <startingVectorLink>,
}
vectorLink ...(sometimes called Edges)
{
        <startNode>,
        <endNode>,
        <rightEnclosedArea>,
        <leftEnclosedArea>,
        <directionOfDigitisation>,
        <pointerToCoordinateSet>,
        <angleOfVectorSegmentAtStartNode>, ...(can be computed on the fly)
        <angleOfVectorSegmentAtEndNode>,
}
node
{
        <attachedVectorLinksList[]>,
        <pointerToCoordinateSet>
}
coordinateSet
{
        <SegmentListOfXYZtriples[]>
}
enclosedAreaSet
{
        <attributeList[]>,
        <enclosedAreaList[]>
}
networkSet
{
        <attributeList[]>,
        <vectorLinkList[]>
}
pointSet
{
        <attributeList[]>,
        <nodeList[]>
}
```

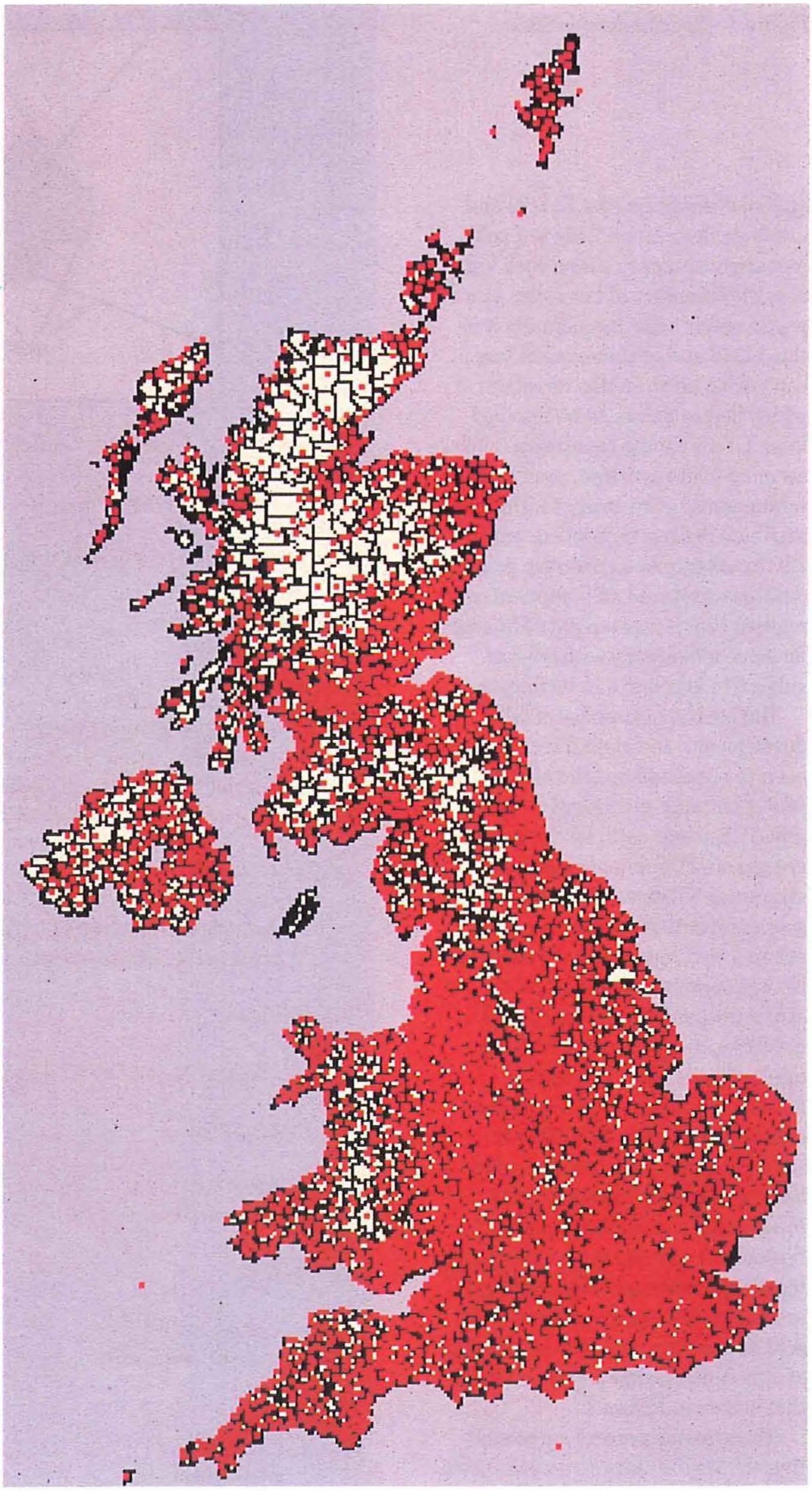
spatial description (the X, Y, Z) and attribute description (this is a point on parish boundary, also a road kerb, also the boundary of Leicester, also a postal sector edge, but actually it is the A4342 and at the moment you can't drive north in this direction...) which had somehow to be resolved later. Of course the coordinate triples seem inviolate, pristine, somehow the fundamental pure entity to which the attribute list can be applied, but it isn't true because as the very point data was captured off a paper map it implied that it was a point of interest for someone already with a set of subjective attributes in its baggage.

During the next phase of GIS development, the main direction has been to impose some kind of order to this essentially unordered or 'spaghetti' digitised data, by the general acceptance of link-and-node data structures which enable fundamental geometric entities of both point entities, and linked sets of points to be organised into coherent record structures, and thus if needed into networks, and polygons enclosing areas of ownership, interest, or fundamental type. This allows the actual spatial data (X, Y, Z) to be stored in some huge less-classified heap, while meta-data about the coordinates is described separately, typically in structures which support topological tracing of both enclosed area polygons, or network traversal. A plausible *data structure* (with loads of data redundancy) might resemble that shown in Figure 1.

There seems general agreement that the kind of data structure shown in Figure 1 supports most of the information processing we may want to do with vector data, and various implementations of this kind of approach are common although not uniform. Some issues about actually making the unordered mass of sets of coordinate strings (which were typically digitised manually), actually start and end at nodes, and for link records to be accurate, do exist, and many GIS implementation issues arise from data integrity.

The task of converting BT digital map data (see Figure 2) for exchange locations and boundaries from the network planning Intergraph system for use in the BT marketing spatial analysis toolkit required the creation of an entire link-and-node description layer for wholly unordered sets of coordinates of 6000 exchange boundaries for the UK, and the manual repair of 'overshoots and undershoots' in digitisation. Only then could any enclosed area identifiers be applied (exchange 1141 codes) and that achieved by point-in-polygon analysis of exchange XY locations derived from an altogether different network planning data set, and only after

Figure 2—UK map of BT exchanges



verification and cleaning of duplicates could attribute sets be applied to permit attachment of business information to these spatial entities. Then we could start using all this in a commercial context to analyse attributes (numbers of customers with this, or that...) in a spatial way.

Of course *raster data* is essentially just a bitmap with certain 'handles' or digitised surface points for rubber-sheet transformation and much work now focuses on deriving vector data including Z coordinates, links, and nodes, from raster images using imaging techniques. Interestingly the printing process to generate paper maps from vector data actually requires a reverse vector-to-raster transformation with considerable manual tweaking to get a good print image and a coherent representational style for intersecting roads and features that physically are co-located but need to be visually distinct for the map to work. This demonstrates the conceptual gulf between a GIS and a map.

Some Current Problems with Digital Maps

The most common problems being encountered in geographic systems at present are about merging data created for one human activity with data plausibly about the same thing but coming from a quite different direction. From the description above of how spatial data is actually digitally organised, it probably is clear that the kinds of things that go wrong are as follows:

- What you are interested in just hasn't been digitally captured yet.
- The objects you are interested in are partially, differently, or incoherently described in another data set because the meta-data has not been written even though the base data might be digitised. You might be able to see the feature visible but it is not referenced in a way that makes sense to your information base.
- The attributes or geocoded descriptors are 'too firmly attached' to discrete polygon, link, node or even coordinate records rather than residing in a meta-data layer as information *about* the links, nodes, and areas (which are themselves meta-data *about* the coordinate records). Unnormalised repeated sub-fields of identical coordinate strings in possibly a quite different node and link structure may coexist in the same data set, so that, say, a co-locating road, boundary, edge of a building may not be structured in the manner described above but be explicitly repeated in the data.

The commercial attractiveness of dynamically merging spatial information at one point of need, for say tourist information kiosks, or legal conveyancing services, into customised user packages about:

- water, electricity, gas, CATV, telephone network plant;
- road and rail links, nodes and timetables;
- traffic information and routing data;
- local authority land use and title deeds;
- what's on where and when;
- accommodation and booking services;
- demographic information

may drive toward a common data standard for spatial reference, but it has not happened anywhere yet and will require a great deal of work before it will be feasible.

The level of success at *automatic feature extraction* from semi-structured spatial data is not yet advanced, and uses attempts to create topological descriptions from what is there both by inference, and by decomposition.

Inference works by saying that a road mid-line digitised for some navigation application should always plausibly be inside a pair of road edge or kerb lines even though only 58% of the 'nearest edges' by that criteria are actually described as kerbs, so an additional 'kerb' descriptor might be applied to the remaining 42% candidate edges which might up to now be called building edges, or parish boundaries. This is fuzzy logic in operation.

For *decomposition* an 'imaging' technique can transform the entire vector data set into a three-dimensional array in which each vector layer creates a horizontal element set in a plane and every single vertical stack of cells in the array contains data about that point on the ground. Each cell has a single identifying tag according to the scale of decomposition—so this bit of world is a '1 metre by 1 metre chunk of road type X' *and* 'has water pipe owned by Thames Water at 3 feet deep' *and* 'is in Kingston-upon-Thames' *and* 'is in KT2 5TH postcode' *and* 'is at 120 feet above sea level'. Analysis by this decomposition is referred to as *map algebra* and is pretty heavy on machine resources and again fairly fuzzy.

What all these techniques are doing is trying to bridge the gap between what information has actually explicitly been put into a system and what you need to get out. You may think it is in there, and a map produced by the data may even plausibly mark it, but it is not really in there at all, or not in a way that relates to how all your other ways of talking, thinking about or processing information might refer to it.

Detail and Gestalt

From the discussion about extracting information from highly granular detail comes the issue of exactly how detailed do you need to be to support strategic decision analysis with maps for data visualisation? Marketing data is generally incomplete and has a confidence factor built in. A decision with half the data but in half the time might be best for business. How exactly can locations of a whole range of types of things be actually found in order for at least some kind of aggregation to be possible, and how accurate need they be? Will the nearest town do? Even in a road navigation system, if your engineer-driver is sighted do you actually need to pinpoint the manhole to 10 centimetres, or just tell him which road it is in?

This is precisely the domain of conflict between traditional telecommunications engineering approaches to spatial information, and that actually needed by those who will market the company's products, and decide what those products should be. The picture of central London, taken from a BT digitised map, in Figure 3, shows an excessive attention to detail but with absolutely no machine extractable business information; in it there are labels but these are not logically attached to the roads or buildings, which have no existence as data entities to which information (which does exist) could be attached. The telephone number points are placed by approximation from a

Figure 3—Detailed London map with no information that is processable

Table 1 'Location' Entity Decomposition

1 Fundamental Entity/Super Entity Definition

LOCATION is how BT refers to where its plant and services exist or should exist, where its competitors' plant and services exist or may exist, and where and how parties can or could be reached by its services or competitors' services, its sales and engineering people, and its postal mail messages and bills

2 Decomposition of Fundamental Entity/Super Entity

Fundamental Entity	Breakdown =>subtype	Data Type	Example	Description/Definition
LOCATION	Spatially accurate or exact geoposition or set of positions (in latitude/longitude (or local grid) derived from surveying, map digitisation, satellite global positioning, aerial imaging, or scanning from paper maps)	**Vector Data:** Point entity (node)	Exchange location in national grid (xy) Customer private circuit end point (xy) Cellular transmitter (xyz) Cable junction, road junction (xy)	Node-and-link vector data structure enables network traversal, point, link and polygon feature extraction, point in polygon analysis, with suitable attribute assignment to node, link, and enclosed area (polygon) entities
		Link entity	Network segment or road segment digitised from map in topological form	Follows topological rules in its format as well as geography
		Polygon entity	Exchange boundary, postal sector boundary, census ward, buffer coverage zone around cable link or cellular node	Allows aggregation, territory management, data visualisation, and is (usually) constructed of topological links
		Trajectory	Microwave, tropospheric scatter, satellite uplink	Typical transmission or reception as in ballistics
		Spaghetti data	Typical BT WN digital map base in Intergraph format	Unstructured into node-and-link format and without attributes
		Raster Data	Aerial or satellite images scanned maps	Manipulable background for realistic superimposition of vector data features and for vector feature extraction by image analysis
			Attributed raster cell arrays	Decomposition of multi-layered vector data and image-derived point data into a notional 3D array of single-attribute cells suitable for map algebraic analysis
			Screen or print image of vector data	Presentation output of vector data suitably processed for visual use
	Pragmatic or workable expression, rule or artefact for location	Postal address	Full or partial address Postcode, zip code etc Country, state, county, city, suburb name within address data	Allows delivery of mail and people, and approximation of real world location by fuzzy matching within a hierarchical framework of rules, algorithms, and 'sounds-like~' (for foreign languages)—is very dependent on the country of interest
		Transport network reference (named road, rail point or link) expressed as start, end and direction (with speed and time)	'Autoroute', vehicle tracking drive-time analysis mobile market, routing schedules	Relies on node and link topological data but permits time and distance and 'buffering around' analysis, tracking of moving points within a spatial frame of reference and drive-time analysis to create reachable catchment areas
		Map	Photograph, sketch, detailed vector drawing, schematic	Leverages human perception (gestalt) and semantics to allow often minimal, selective or stylised world detail to allow effective communication

postcode-to-grid lookup table and are fuzzy. But even though fuzzy, if we are viewing thousands of such points we see them cluster and can make planning decisions. Meanwhile, the meticulously digitised map acts as some backdrop; a scanned image would do just as well.

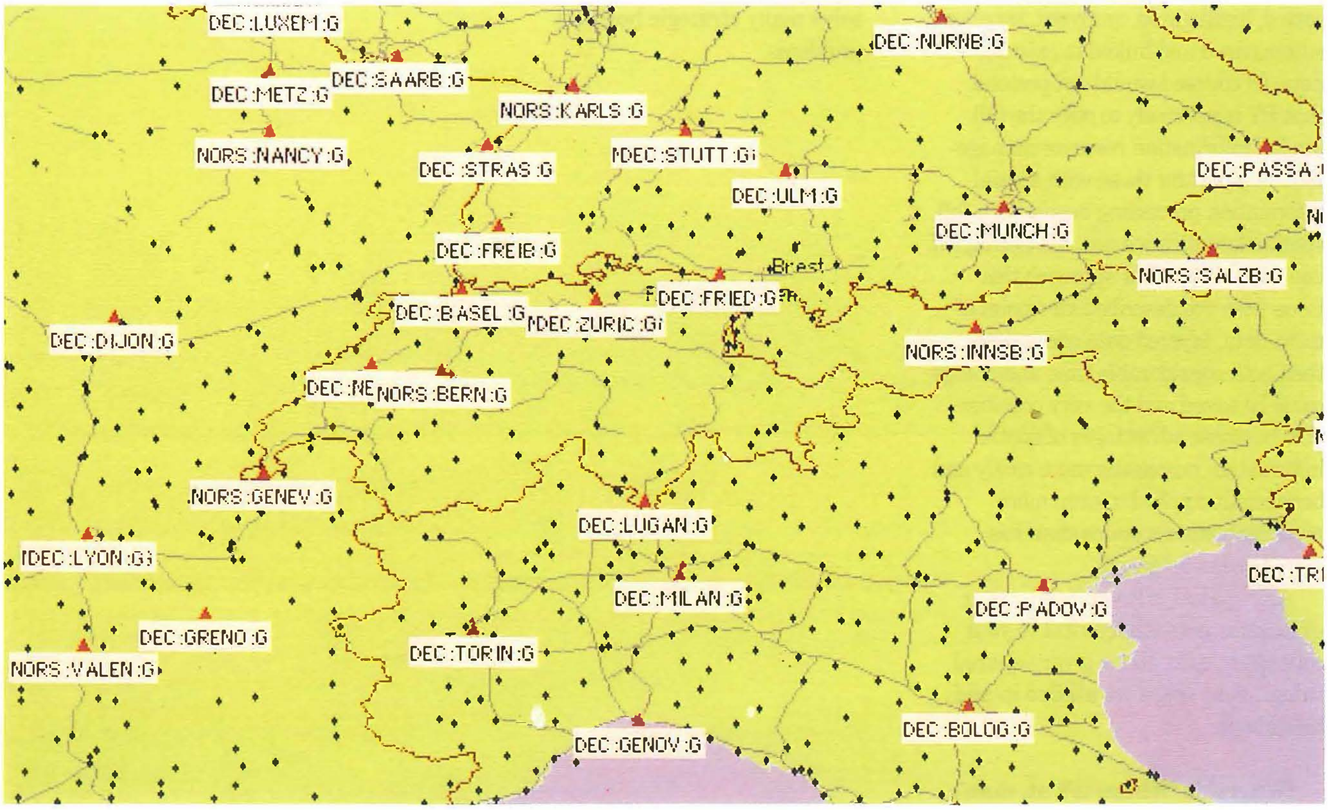
To get around all this confusion, it might be desirable to construct a pragmatic, fuzzy, location fixing tool which uses a variety of approaches to get a best-fit for a strategic problem. Table 1 suggests how such a tool might work.

Conclusion—Where is BT in all this?

It seems to be a core concept rooted deep in telecommunications engineering history that some circuits are charged by actual point-to-point real-world distances; hence we have meticulously recorded actual coordinate points for each exchange, and for some private wire end points but often we never actually named these points as data entities which we could link to commercial business records. Having in the last few years brought this digitised vector data to bear on marketing analysis issues, we have now an increasingly rich picture of BT's spatial information. In adding commercially acquired vector layers for the full AA road map set, postal Area, Sector, and full postcode points, census ward areas and the like we have a workable reference model for the UK for many kinds of strategic business analysis. For the rest of the world, a very complex situation exists in that the quality of digitised data available is mixed; mechanisms (ZIP codes, postcodes etc.) for location-by-address vary country by country if existing at all. We have available a world dataset but it lacks much in the way of meta-data and can only act as essentially wallpaper for incremental enhancement. City detail in the UK is probably to be addressed by the External Plant Records project as a raster layer to which node points linked to plant record data will be applied, but even this focuses more on what is in the ground than what should be in the ground. For this latter analysis it is strategic information about customers and competitors which will need to be integrated with such engineering data; the challenge to the company is to unify network and marketing data in a single information resource so each discipline may share a single approach (see Figure 4).

The tasks ahead could include restructuring all our existing vector data into proper meta-data layers in common data structures and eventual use in whatever user analysis or network design tool is needed. In approaching new applications the company should manage its spatial information assets more coherently than it has in the past and, recognising

Figure 4—World map detail showing potential business customers in North Italy near potential Tymnet nodes. Generated from DCW map data and marketing strategy address data



the peculiar feature of digitised spatial data—its high cost to license and inflexibility when in rigid data structures—avoid as far as possible implementing custom solutions which may prove expensive and difficult to develop and maintain. The peculiar role of multi-layered meta-data in GIS which can make a system you may find quite useful actually not any use to the rest of the company **even though it seems to be 'about' the same things and places** is a quite difficult concept to grasp, and is not one which has yet led to off-the-shelf GIS tools 'for all seasons'. GIS is an immature market, is expensive to enter, and can bankrupt those who jump without thinking twice.

The golden rule for anyone wanting to add to the set of spatial information resources might be this:

> **'If something is worth putting on a digital map it is worth finding out how to name it as an entity, preferably as a type of entity on which someone already keeps information....'**

It can then be cross-referenced, sorted, highlighted, analysed, targeted, summarised and linked to existing data. Of course I would not pretend that BT is yet ready to offer the full kind of information resource management support for those with spatial information processing needs, with full spatial data dictionaries, but perhaps if enough people could approach this issue from the described viewpoint of meta-data, layered data about data, then both considerable time and energy could be saved and the very considerable business advantages of spatial information processing more easily and beneficially applied to even more pressing business issues than has already been achieved.

Any spatial information in your systems is potentially vital to your colleagues as it about a very shared thing...that space we all live in and talk about...

'Where' is 'Where it's at, man'.

Biography

David Haskins
BT National Business Communications

David Haskins joined BT in 1981 after some years as a hardware engineer with Sperry Univac (now Unisys) and then having studied Philosophy in London University as a mature student. In his BT career he has enjoyed a focus on how people use (or cause catastrophe by using carelessly) information systems of all kinds, and how such systems change the basic set of conditions we all have to deal with. Most recently he has designed an accessible, cheap, and easily used set of geographic analysis tools within BT's Marketing Information Centre. The toolset and approach is now formally supported by D&P, has a help desk, and was successfully presented in D&P Innovation '94. It is being widely applied across BT to solve many strategic business problems.

Bernie Rowlands

ITU Network Management Development Group Meeting, Brazil, 1993

The annual meeting of the ITU Network Management Development Group was held in November 1993 in Brazil. This article summarises the function of the group and the topics covered at the meeting, including a paper presented by AT&T on the car bomb attack on the World Trade Centre in New York and its effects on telephone service.

Network Management Development Group

The Network Management Development Group (NMDG) exists to encourage wider international participation in the identification, development and implementation of network management activities and to share the findings of member countries. The group reports to the International Telecommunication Union (ITU) Study Group 2, which deals with service operations issues concerning the 170 member countries of the ITU.

The NMDG has a 3 year work programme centred on the principles of freely interchanging advice and data on network management and allied activities. While the group concerns itself broadly with network traffic management issues, the scope of its remit is sufficiently wide to include issues such as support system developments, signalling management and disaster management. These can be broadly grouped as shown in Figure 1.

NMDG Annual Meeting, 1993

The NMDG's annual meeting is hosted by a member country—in 1993, Telebras, the Brazilian state telephone company, hosted 51 delegates from 18 countries representing over 30 operators and suppliers.

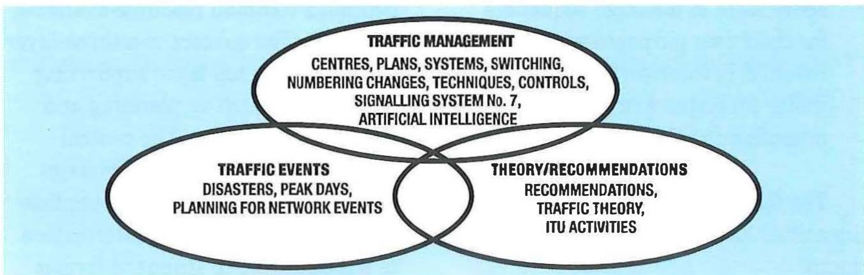
A total of 27 papers were presented covering the following functional areas:

- review of major events and peak-day traffic events impacting the network;
- planning to avert network failures and national disasters,
- network management controls and techniques,
- automatic network management,
- common-channel signalling network management.

Papers of particular interest

The range and scope of all the papers were valuable and gave an up-to-date view of network management developments in European, American, Pacific Rim and third-world countries. A necessarily brief summary of some of

Figure 1—NMDG work areas



the challenge to the company is to unify network and marketing data in a single information resource so each discipline may share a single approach

the more significant papers is included here together with a more detailed review of an AT&T paper covering the World Trade Centre terrorist explosion.

Signalling System No. 7 Management

A paper by Telstra Australia gave an interesting insight into the PTT's experiences with Signalling System No. 7 (SS7) management. The author was frank about the problems experienced and advocated very tight management to avoid problems. Five major areas were identified in need of careful management:

- software quality/integrity,
- network design topology,
- network management and disaster planning,
- operational and implementation procedures,
- planning and design rules.

Many practical examples were given where conventional thinking based on traffic management fell down and more rigorous procedures had proved essential. These included:

- avoidance of circular routings which can be difficult to identify in complex networks;
- design rules; for instance to avoid putting all signalling links in the same cable/duct/fuse supply etc;
- improved specification standards; examples were quoted where manufacturers were allowed to use spare slots in message sequences for their own purposes which resulted in interworking compatibility problems with different manufacturers' systems.

The following parameters were identified for sound SS7 management:

- manufacturers must be given very stringent specifications;
- telecommunications authority design and roll-out rules must be followed without deviation;
- disaster recovery plans must exist;
- a management control system must exist, but kept simple;
- managing SS7 is a company-wide issue;
- the network is a hands-off one and must be treated as such.

Traffic Management Developments in Europe

Papers were presented by Telia Sweden, Deutsche Bundespost and Telecom France. These gave a useful progress report on current traffic management developments in the three countries.

The Swedish paper gave a good picture of Telia's (formerly Swedish Telecom) plans to roll-out an Ericsson Telecom Management Operations Support System (TMOS) technical platform. Swedish Telecom had started out like most with a series of standalone systems covering traffic management and planning/utilisation. They went through the phase of attempting interconnection which proved to be expensive and inefficient. The administration has now taken the major step of investing in an Ericsson TMOS. One of the reasons for choosing TMOS was seen by Telia as the advantages it gave in creating well-defined interfaces between the applications and the data used, thus allowing data to become a common resource available to others. The concept is a three-layer system with a top layer supporting applications such as planning and traffic management. The central layer is an integrated TMOS database. Below this sit element application handlers supplying information to the database in standard format and drawing data from the network elements in standard Q format.

The system will operate on Sun and Hewlett Packard workstations with physically distributed databases interconnected by a 2 Mbit/s local area network (LAN). Each geographic region is planned to be supported by its own database. The system will open in 1994 managing 200 network elements and switches and it is planned to grow to 300 switches and 5000 routes by 1997. The final network will have an international centre, a national centre and eight regional centres.

The German approach was characteristically thorough but with a relatively slow roll-out.

The German PTT developed their traffic management concept in 1990 and expect to have a comprehensive operational system in place by 1997. To date they have defined a trial, invited 30 manufacturers to tender, subsequently reducing this to three and later to two successful tenders. Two pilots have been opened—one for each manufacturer and each with four switches connected. By mid-1995, the trial will have been evaluated and the geographic coverage extended with some 46 switches connected giving a traffic overview of their transit layer (top layer) of switching. Experience with this trial will give the necessary confidence to implement national standards and roll-out a comprehensive system with a national centre in Bamberg. The goals are economic usage of the network with quality of service safeguarded.

By contrast with the Swedish and German approaches, the French gave a presentation of their second generation of traffic management development which had been in operation for two years. This had been preceded by a first-generation development which was operational in 1980.

The current French system has three significant features:

- *Transmission system management* Data from both analogue and digital transmission systems is collected via interfaces developed by Telecom France and using their Sparte system. This permanently monitors performance and availability and in the event of failure can automatically switch at the 140 Mbit/s level. The service protection network is centrally controlled and managed.

- *Signalling system management* An application called PSG, which went into service in 1991, monitors quasi-associated signalling routes and STP working. Monitoring includes circuit, circuit group and STP status.

- *Traffic Management* The application supervisor 2G is used to give real-time traffic information on routes and switches. A further application known as *Violette* gives information in real time on call attempts.

France Telecom has connected both digital and analogue transmission systems into their management centres and developed the necessary interfaces. The Transpac packet switch network is used to provide a communications infrastructure.

Two levels of management centre are in existence, a national level with a single centre located in Blagnac near Toulouse in southern France and 35 regional centres responsible for the lower parts of the network hierarchy. A separate centre exists for Paris.

Use of Artificial Intelligence in International Network Management

A paper by KDD/Japan gave an interesting view of possible future developments in traffic management employing artificial intelligence. The KDD representative was a section manager in KDD's international network management centre. A brief history of the development ran as follows.

KDD established its international network management centre in 1986 and started developing a prototype support system (NMEX) in 1987. A subsequent system NEMS commenced in 1992 and was brought into service in 1993. NEMS was developed because of the need to restrain the likely need for growth in staffing the management centre. KDD's previous experience in the mid-1980s indicated that artificial intelligence looked a promising development route, but at the time this was constrained by a lack of computing power. The computing power used by NEMS is between 5 and 15 times faster than that used by NMEX and the artificial intelligence rule base has been increased from 260 to 1000. Processing power now used is 32 bit 29 MIPS compared with that previously of 16 bit 0·8 MIPS. Language is OPS 83L and previously OPS 5C. NEMS is used to detect abnormal traffic conditions and uses its rule base to reason the likely causes. It does not apply network controls directly; this is currently left in human hands, but further studies are taking place in this area. The system reviews the following parameters:

- routing data,
- existing exceptions,
- destination parameters,
- route parameters, and
- route status.

Network status is examined by looking at exception reports in the above areas and assigning these to one of the following categories:

- sharply increased,
- increased,
- constant,
- decreased, and
- sharply decreased.

NEMS holds a copy of the routing database which is updated daily. NEMS suggests to the operator the most likely cause of the problem, the reasons for this and the appropriate controls to apply. Data used in the reasoning process is made available via a screen-based drill-down process.

It is too early to form firm judgements in this interesting development but progress to date indicates the following:

- The existing design gives high priority to speed of analysis and more accurate results will be available by trading speed for accuracy. It is expected that the measuring speed will be reduced to 2–3 minutes to give optimum balance between speed and accuracy. To allow for this, the system is being upgraded (completion due 3/94) to optimise reasoning speed versus accuracy of result.

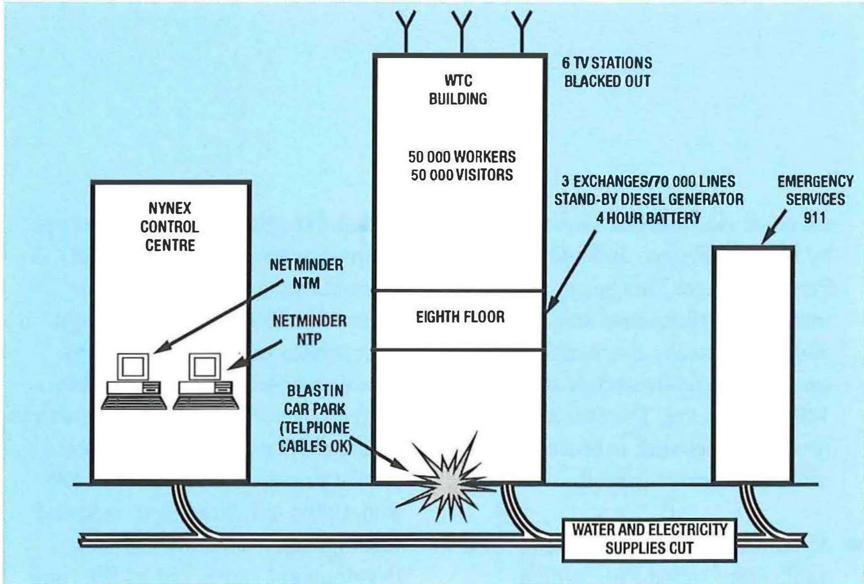
- KDD is optimistic that this will prove a useful development using an area of technology not previously fully exploited.

Future Standards for Switch Data

AT&T presented an interesting paper entitled 'Proposal of Discrete Data as Standards'. They identified the need to establish comprehensive international standards in the area of switch data. Work undertaken by AT&T over the past few years had shown that the signalling system management was subject to the effects of very short transient conditions which need to be carefully managed. The transients can occur too quickly to be seen by conventional traffic management systems sampling at 5 minute intervals.

The switch data in question is binary, that is, a conditions exists or does not exist, and is reported by switches at approximately 30 second intervals. An example of such data might be a congestion status monitor

Figure 2—Explosion at the World Trade Centre, New York



indicating congested or non-congested status. Other examples are:

- status of circuit groups carrying traffic or signalling data and either able or unable to terminate their traffic;
- changes to signalling system status taking place at typically 30 second periods such as congestion control messages from another signal transfer point;
- switch multifrequency receiver capacity overflow status; and
- status changes to common control resources such as increased usage or decreased usage.

It was identified that the volume of such data could be considerable and extreme care had to be taken in presenting it to an operator. The key principle used was to display data in a matrix format and use simple pattern-recognition techniques developed after extensive human-factors research.

AT&T had taken the principle into practice and they used discrete data to trigger network controls automatically. To avoid the unwelcome effects that momentary transients could have, the controls are applied progressively and major changes are based on receipt of continuous evidence of network status changes.

Although the principles have evolved out of AT&T's work on signalling system management, they contended that the principles held good to supplement information derived from conventional traffic management techniques. Without agreed international data standards, AT&T could only extract data from their own manufactured switches such as 5ESS and they were keen to see international standards set to allow this area to be opened up further. The paper was received with interest and the work area seen as appropriate for further study.

Terrorist Attack on World Trade Centre, New York

The explosion at the World Trade Centre (WTC), the subject of an AT&T paper, remains the most significant terrorist event ever in New York State. The WTC building normally houses 50 000 people with thousands of additional visitors on any one day. It is a major international site known throughout the world.

The explosion was caused by a terrorist car bomb placed in the car park in the basement of the WTC. It caused a major disruption to the New York telephone network and attracted worldwide media attention. The explosion was very extensive and caused all buildings in the locality to shake; smoke and fire occurred in the area of the explosion and, as a result, commercial power to the area was cut. Although the WTC building was seriously damaged in the basement, the remainder of the building remained sound.

The power cut resulted in the three Nynex public network switches (see Figure 2) located on the eighth floor of the WTC switching to emergency stand-by power. However, a state of emergency was declared by the authorities and the water supplies were cut off as part of the emergency procedures. This resulted in the failure of the three stand-by water-cooled generators. It is interesting to note that because of fire restrictions, fuel and cooling water for the stand-by engines are pumped externally up the building to the eighth floor. As a consequence of the power failure, the three switches fell back to 4 hour stand-by batteries.

The cables feeding the WTC exchange survived the explosion, the force of which ripped the cables off their supports where they passed through the car park area, but the cables flexed and did not sever.

An orderly evacuation of thousands of people from the whole area commenced as emergency services implemented pre-planned evacuation procedures.

Nynex staff responsible for telephone service in the area were located in the Nynex network control centre close to the WTC building. The control centre was hit by the power cut and many workstations not connected to uninterrupted power supplies went down. The use of uninterrupted power supplies was limited for economic reasons, which is perhaps surprising given the scale and sensitivity of the geography served by Nynex.

The centre had, however, just commissioned an AT&T Netminder Network Trouble Pattern Surveillance System (Netminder NTP), which is capable of surveilling both public switch voice services and advanced network services, and was connected to an uninterrupted power supply. The system was originally developed with the objective of managing common-channel signalling and the availability of this support system was instrumental in enabling Nynex to manage the crisis effectively. The Netminder system draws on both call and switch data from many exchanges in the New York area including the three in the WTC building, which alone totalled over 70 000 lines. Typical failure information made available by Netminder includes:

- time of event,
- reporting switch,
- digits received/transmitted,
- failure symptoms,
- circuit group affected, and
- circuits affected.

The system also correlates switch failure data and common-channel signalling failure data. Netminder is capable of exhibiting data within the normal 5-minute traffic measurement period used by many conventional traffic-monitoring systems. This enables the system to identify very rapid changes in the state of the network, which is critical for effective management of signalling systems.

The bomb explosion affected the WTC and immediate area, but the crisis grew quickly when six TV channels serving the area went off the air because their transmitters were located on top of the WTC building and were thus affected by the power cuts. When this happened, the New York public switched to another TV channel transmitting from the Empire State Building. This channel proceeded to transmit a news flash entitled 'Terror in the Tower'. This created great alarm with the general public, many of whom had relatives in the WTC building.

The effect was a huge and immediate focused telephone traffic overload to the three WTC switches in the now unattended WTC building. At the Nynex Centre, the Netminder NTP kit picked up multifrequency receiver attachment delay reports caused by the incoming calls to the three WTC switches being unable to locate sufficient incoming MF receivers. Netminder NTP identified that no specific numbers in the WTC building were being called.

The WTC building and the surrounding area were evacuated and telephone traffic controls had to be applied to deal with the call build-up. This was done by blocking calls to congested switches at source and reducing traffic sufficiently to allow outgoing calls to materialise. During this stage of the crisis, Netminder was used effectively to identify a shortage of emergency operators dealing with the 911 (our 999) service, as large numbers of calls remained unanswered at the emergency centre but no evidence existed of switch blocking.

Heavy telephone traffic congestion occurred in the whole area surrounding the WTC throughout the day. TV stations broadcast emergency numbers for special classes of emergency service and Netminder NTP was used to monitor the state of the emergency services throughout the emergency. Rerouting commands were applied where necessary by the control centre to overcome local congestion problems. The depth of detail available from the Netminder system complemented conventional traffic monitoring and gave greater detail of specific calling patterns and call failure detail, making possible more informed management decisions.

The three switches in the WTC building remained operational, supplied by the 4 hour stand-by battery, and the problem presented to the Nynex engineers as battery capacity reduced was solved in a novel way. The three switches were remotely reconfigured in data from duplex processor working to simplex working; the resulting reduced load extended the battery life by a further three hours, which was just long enough for commercial power to be restored.

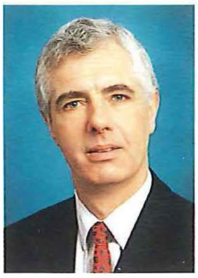
After the immediate crisis was over, Nynex staff used Netminder to detect and count unanswered calls to PBXs in the area and from this to create a prioritised list of customers from the 10 000 companies affected by the disaster. The list was used to help direct and support sales of restoration and call-forwarding services to high-usage customers. Apparently Nynex's customers were very satisfied with the telco's speed of response and were prepared to pay a premium for prioritised restoration of service which they saw as a valuable offering.

Discussion with the presenter of the paper elicited the following information:

- detailed recovery plans existed for major customer buildings in New York including the World Trade Centre,
- the cellular service was used heavily as an expedient and prearranged plans existed whereby both cellphone and base-station capacity can be augmented quickly to meet unforeseen situations.

In summary, the AT&T paper indicated that Nynex had handled a major network event effectively. In doing this, they had made good use of support systems, which played a key part in the early stages of the emergency and later when sales staff used the information to support a sales drive aimed at companies in most urgent need of restoration of telephone service.

Biography



Bernie Rowlands
BT Worldwide Networks

Bernie Rowlands joined BT in 1959 in Liverpool. He has spent much of his career in core network planning and operations roles and is currently Operations Manager for Worldwide Networks North West Zone.

Michael Goddard

The New International Telecommunication Union

The International Telecommunication Union (ITU) is the pre-eminent international body in the world of telecommunications. It has broad responsibilities for the development of telecommunications, including harmonisation and standardisation, and for coordinating the use of the radio-frequency spectrum. Following an ITU conference in 1992, the organisation has undergone many changes in its structure and working methods, resulting in the abolition of many familiar terms. This article describes the background to the changes and the new arrangements.

Introduction

The International Telecommunication Union (ITU) is the oldest specialised agency of the United Nations and plays a key role in the global development of telecommunications. The history of the ITU goes back to the International Telegraph Union, which was founded in 1865, before the invention of the telephone and well before the exploitation of radio transmissions.

The purposes of the Union are laid down formally in the ITU's Constitution. They refer in general to the maintenance of international cooperation between all members for the improvement and rational use of telecommunications of all kinds. (It should be noted that in the context of the ITU the definition of telecommunication is extremely wide. It reads 'any transmission, emission or reception of signs, signals, writing, images and sounds or intelligence of any nature by wire, radio, optical or other electromagnetic systems.' Thus, in addition to traditional telecommunications networks, the ITU covers broadcasting, satellite communications, radio navigation, amateur radio, radio astronomy etc.) One of the key defined purposes of the Union is 'to promote the development of technical facilities and their most efficient operation with a view to improving the efficiency of telecommunication services, increasing their usefulness and making them, so far as possible, generally available to the public.'

The specific functions of the Union include the following:

- the allocation of frequency bands, the allotment of radio frequencies and registration of frequency assignments and any associated orbital positions in order to avoid harmful interference between radio stations in different countries;
- to coordinate efforts to eliminate harmful interference between radio systems of different countries and to improve the use made of the radio-frequency spectrum and the geostationary satellite orbit;
- to facilitate the worldwide standardization of telecommunications, with a satisfactory quality of service;
- to foster international cooperation in the delivery of technical assistance to the developing countries;
- to coordinate efforts to harmonise the development of telecommunications, notably those using space techniques, with a view to full advantage being taken to their possibilities;
- to foster collaboration among members with a view to the establishment of rates (that is, accounting rates for telecommunications traffic) with levels as low as possible consistent with an efficient service;
- to promote the adoption of measures for ensuring the safety of life through telecommunication services;
- to undertake studies, make regulations, adopt resolutions,

the ITU recognised that its working methods would have to be changed to respond to the rapidly changing international telecommunications environment.

formulate recommendations and opinions, and collect and publish information concerning telecommunication matters;

- to promote the establishment of preferential and favourable lines of credit to be used for the development of social projects aimed at extending telecommunication services to the most isolated areas in countries.

From the above, it can be seen that the ITU has a substantial and fundamental role in the development of telecommunications of all kinds.

The membership of the ITU comprises the governments of the member countries, which now total over 180. Traditionally, the governments have been represented in most cases by the Ministry of Posts and Telecommunications (or its equivalent) of the member states. Following the separation of regulator and operator function in many countries, the government's regulatory body for telecommunications and radiocommunications maintains the formal ITU membership and is known as the *administration*. In the UK, this responsibility is carried by the Department of Trade and Industry, with the Telecommunications and Posts Division taking the lead and covering telecommunications issues in general. The Radiocommunications Agency (an Executive Agency of the DTI) leads in the radio activities of the ITU as well as playing an increasing part in general ITU matters. The major UK telecommunications operators, including specifically BT, play a major role in the preparation of UK policy towards the ITU, by participating in many conferences and meetings of the ITU, and by contributing financially to the Union.

The structure of the ITU has remained basically the same from 1947 until very recently. However, the ITU has now undergone a fundamental restructuring which, among other things, has swept away many of the traditional and familiar terms such as the CCITT (The International Telegraph and Telephone Consultative Committee) the CCIR (The International Radio Consultative Committee) and the IFRB (The International Frequency Registration Board). The background to these changes is described below together with a detailed account of the new ITU structure.

Background to the Changes

The mechanism for bringing about major changes of policy in the ITU is the ITU's Plenipotentiary Conference. Such conferences have been held at intervals of 5 to 7 years in the past. One of their duties had been to review the ITU's structure and working methods. These were defined in the ITU's Convention and it was one of the functions of the Plenipotentiary Conference at each meeting to revise and update the Convention as necessary. At the Plenipotentiary Conference held in Nice in 1989, the ITU recognised that its working methods would have to be changed to respond to the rapidly changing international telecommunications environment. This new environment was characterised by the increasing role of non-government players in telecommunications, following the separation of the regulator and operator function in many countries, the introduction of competition between operators, the convergence of information technology and telecommunications and the increasing integration of different forms of telecommunications into common networks. There was also a major concern about the efficiency and effectiveness of the ITU to cope with an ever-increasing complexity of workload coupled with very severe budgetary constraints on the finances of the Union.

To tackle this problem, the Nice Plenipotentiary Conference established a High Level Committee (HLC) to review the organisation and structure of the ITU. The United Kingdom's member of the HLC was Mr Michael Morris, previously of BT. The HLC produced a report entitled 'Tomorrow's ITU: The Challenge of Change'. This contained many recommendations for the restructuring of the Union and for improving the effectiveness of the organisation. A drafting committee of the HLC subsequently produced proposals for revised texts for the ITU's Constitution and Convention incorporating the recommendations made by the HLC.

In order to consider these developments, the ITU held an Additional Plenipotentiary Conference in 1992 (APP-92). This conference generally accepted the recommendations of the HLC and to a large extent adopted the draft text of the Constitution and Convention prepared by the drafting group. The new Constitution and Convention enter into force on 1 July 1994, but the new structure of the ITU and the new working methods were implemented on a provisional basis from 1 March 1993.

New Structure of the ITU

The activities of the ITU have been reorganised into a federal structure consisting of three sectors: The Telecommunication Standardization Sector, The Radiocommunication Sector and The Telecommunication Development Sector. The Telecommunication Standardization Sector takes over the work of the CCITT, together with some further standardization work transferred from the CCIR. The Radiocommunication Sector combines the activities of what used to be known as World Administrative Radio Conferences, with the CCIR, and the International Frequency Registration Board including its Secretariat. The Telecommunication Development Sector combines the activities of the previous Bureau of Telecommunications Development (BTD) and Centre for Telecommunications Development (CTD). These changes are shown in

Figure 1—ITU changes

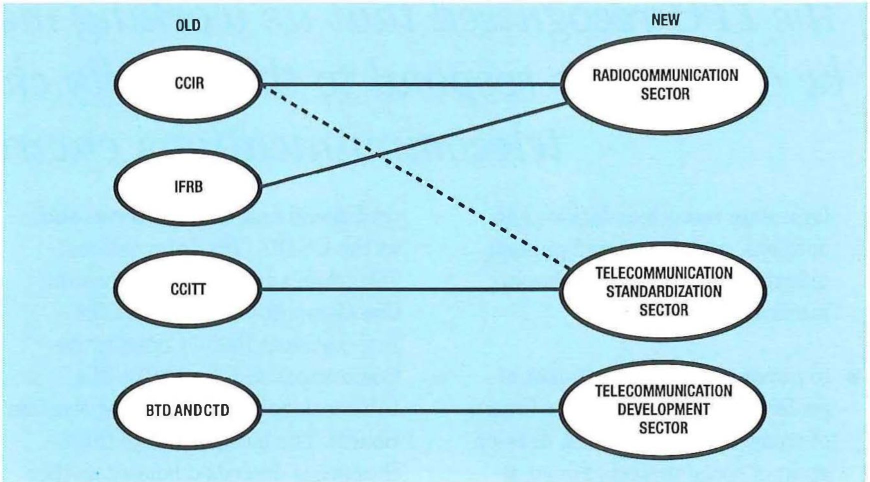


Figure 1 and the new overall structure of the ITU shown in Figure 2.

There will continue to be Plenipotentiary Conferences but in future these will be held at regular 4 yearly intervals. The next Plenipotentiary Conference will be held in Kyoto, Japan, in September/October 1994. The pattern of holding conferences at regular intervals is extended throughout the new ITU structure.

In order to provide day-to-day management oversight of the ITU, the Union has in the past had an Administrative Council the membership of which comprises about one quarter of the members of the Union. This arrangement will continue under the new structure but with the title now simply being the *ITU Council*. One feature of the new ITU will be a greater emphasis on longer-term strategic planning and there are already signs of the Council taking a greater role in this activity than previously carried out by the Administrative Council. The UK was for many years a member of the Council until it lost its seat in 1989. The UK will be seeking election at the Plenipotentiary Conference in 1994.

The ITU has a substantial secretariat in its Geneva Headquarters (Figure 3) headed by a Secretary General who is assisted by a Deputy Secretary General. Under the new structure, which is shown in Figure 4, each of the three sectors has a corresponding Bureau in the ITU Headquarters in Geneva. Each Bureau is headed by a Director. At the APP-92, arrangements were made for the then Director of the CCITT (Mr Irmer of Germany) to act as the Director of the Telecommunication

Figure 2—The new ITU structure

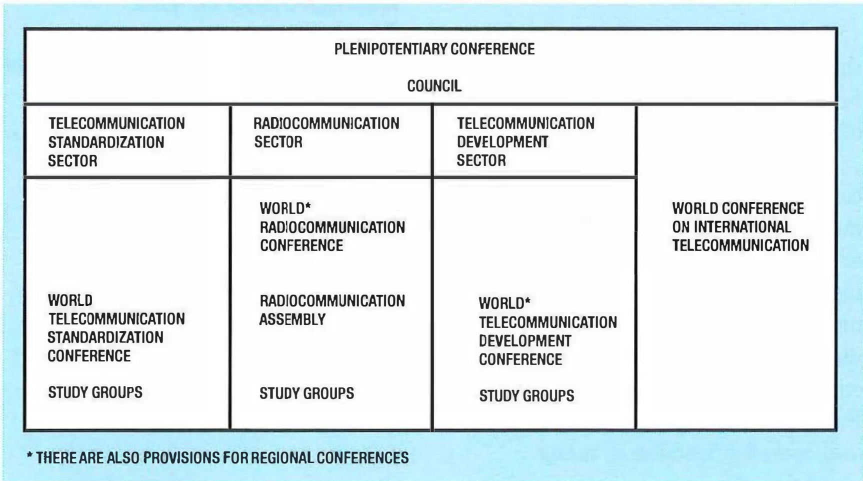


Figure 3—The ITU Tower in Geneva

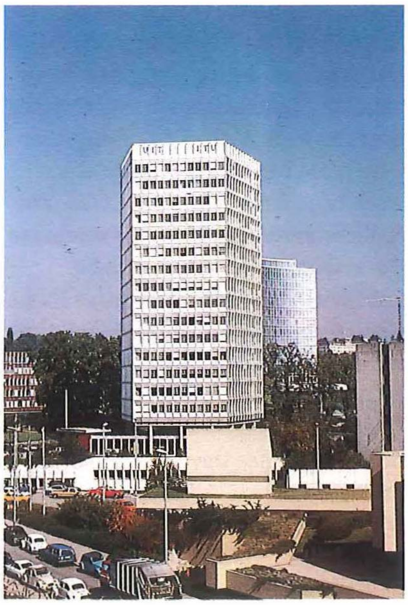
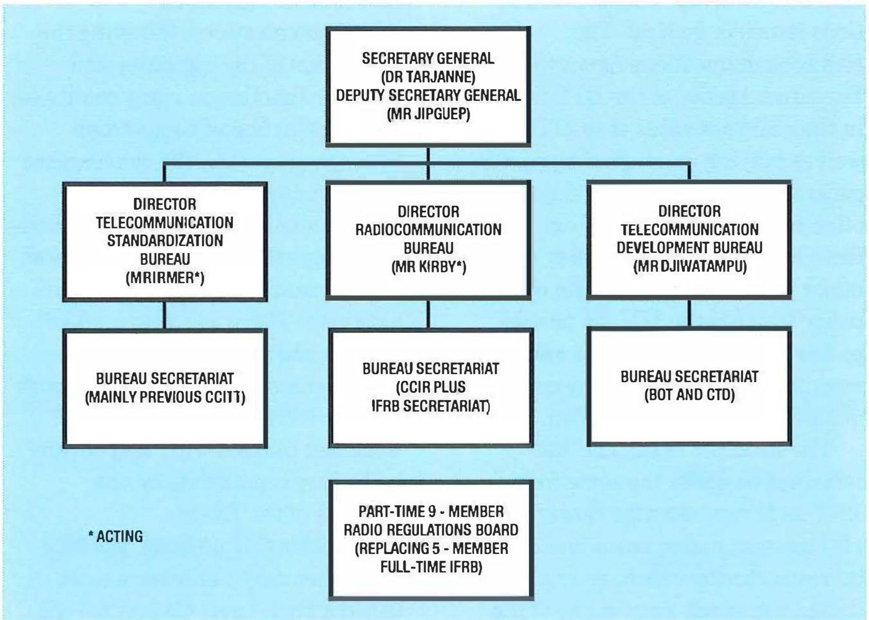


Figure 4—The ITU headquarters organisation



Standardization Bureau. Similarly Mr Kirby (United States), who was the Director of the CCIR, has taken on the role of Director of the Radiocommunication Bureau. A new Director was elected for the Telecommunication Development Bureau—Mr Djiwatampu of Indonesia. All three posts, together with those of Secretary General and Deputy Secretary General, will be subject to re-election in the Kyoto Plenipotentiary Conference of 1994.

The following sections describe each of the three sectors in turn.

The Telecommunication Standardization Sector

The functions of the Telecommunication Standardization Sector are to fulfil the purposes of the Union 'by studying technical, operating and tariff questions and adopting recommendations on them with a view to standardizing telecommunications on a worldwide basis'. The Sector works through the World Telecommunication Standardization Conference (WTSC) and Telecommunication Standardization Study Groups. These relate directly to the previous CCITT Plenary Assembly and the CCITT study groups respectively. The abbreviation generally used for the Telecommunication Standardization Sector is ITU-T.

Following the decisions of the APP-92, WTSCs are to be held every 4 years. To show the commitment to the new ITU structure, even though the new arrangements would not come into effect even on an interim basis until March 1993, the APP decided that the 10th CCITT Plenary Assembly, which had been due in 1992, should become the first WTSC. This took place in Helsinki, Finland, in early 1993.

In addition to a review of working methods and the work programme for the study groups, the WTSC reviewed the study group structure and numbering. The current list of ITU-T study groups is shown in Table 1. In addition, the WTSC established five joint coordination groups (JCGs) to coordinate the activities of the relevant study groups in particular areas of work. The JCGs cover:

Table 1 ITU-T Study Groups

1	Service Definition
2	Network Operation
3	Tariff and Accounting Principles
4	Network Maintenance
5	Protection against Electromagnetic Environment Effects
6	Outside Plant
7	Data Networks and Open System Communications
8	Terminals for Telematic Services
9	Television and Sound Transmission (former CMTT)
10	Languages for Telecommunication Applications
11	Switching and Signalling
12	End-to-end Transmission Performance of Networks and Terminals
13	General Network Aspects
14	Modems and Transmission Techniques for Data, Telegraph and Telematic Services
15	Transmission Systems and Equipment

- Universal Personal Telecommunications,
- Audio-Visual/Multi-Media Services,
- Telecommunications Management Network,
- Quality of Service and Network Performance, and
- Broadband ISDN.

At the APP-92, considerable time was spent debating the distribution of work between the ITU-T and the Radiocommunication Sector (ITU-R). Increasingly, radio is being used for telecommunication services to provide a part, in some cases a major part, of the network. One school of thought was to place all standardization issues within ITU-T including, for example, broadcasting studio standards, leaving the ITU-R to deal only with matters relating directly to spectrum management and utilisation. Others wanted the minimum change to the previous distribution of work between the CCITT and CCIR. A pragmatic solution was adopted which agreed in principle to the transfer of a modest amount of work to ITU-T from the outset, and ensured that the situation will be kept under review. The complex task of implementing this decision in respect of the vast range of subjects under consideration in the two sectors was pursued after the APP.

The UK played a leading role in this process. The WTSC in 1993 endorsed the initial allocation of work and adopted the concept of inter-sector coordination groups (ICGs) to ensure that the related work of the two sectors would continue to be coordinated with the minimum of duplication. Subsequent agreement led to the creation of two ICGs on:

- ISDN and Satellites, and
- FPLMTS (Future Public Land Mobile Telecommunication Systems).

The WTSC also established a Telecommunication Standardization Advisory Group (TSAG) which advises the Director of the Telecommunication Standardization Bureau on working methods, work programme and strategic planning for the sector. Joint meetings of the TSAG with the corresponding Radiocommunication Advisory Group (RAG) provide a further mechanism for ongoing monitoring and review of the distribution of work between the two sectors.

The output of ITU-T is in the form of Recommendations and Reports, as was the case with the CCITT. The WTSC does not have treaty-making authority. There is provision in the ITU Constitution for convening World Conferences on International Telecommunications (WCIT), which have the power to revise the International Telecommunication Regulations, and these conferences will replace the World Administrative Telephone and Telegraph Conferences, the last of which was held in Melbourne, Australia, in 1988. However, these conferences fall outside of the Telecommunication Standardization Sector and no provision is made for such conferences to be held on a regular basis. There are no plans at present to hold a WCIT.

The Radiocommunication Sector

The functions of the Radiocommunication Sector are to fulfil the purposes of the Union relating to radiocommunications by:

- ensuring the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including those using the geostationary satellite orbit, and
- carrying out studies without limit of frequency range and adopting recommendations on radiocommunication matters.

The structure of the sector is more complex than that for ITU-T. Beginning with the similarities, the ITU-R has Radiocommunication Assemblies which have very similar responsibilities to World Telecommunication Standardization Conferences. The first Radiocommunication Assembly was held in 1993 and it undertook a comprehensive review of working methods, work programme and study-group structure. The resulting new ITU-R study-group structure is shown in Table 2. The Radiocommunication Assembly established the Radiocommunication Advisory Group (RAG), which is chaired by the author. The ITU-R does not have the need for joint coordination groups. The first meeting of the RAG was held in April 1994 and at an associated joint meeting with TSAG, further refinements to the distribution of work between the two sectors were agreed. As a result, a block of outstanding work relating to the interconnection of very small aperture terminals (VSATs) in the fixed satellite service to the public switched network will be transferred from ITU-R to ITU-T.

Table 2 ITU-R Study Groups

1	Spectrum Management
2	Inter-Service Sharing and Compatibility (previously Study Group 12)
3	Radio Wave Propagation
4	Fixed-Satellite Service
7	Science Services
8	Mobile, Radiodetermination, Amateur and Related Satellite Services
9	Fixed Service
10	Broadcasting Service—Sound
11	Broadcasting Service—Television

Note: There is no Study Group 5 or 6 in the new structure. The previous Study Groups 5 and 6 have been merged into the new Study Group 3 and the old numbers left unused to avoid confusion.

The Radiocommunication Assembly differs from the WTSC in respect of its support for World Radiocommunication Conferences (see below). Because such conferences are to be held every two years, Radiocommunication Assemblies will be held at the same interval; that is, twice as frequently as World Telecommunication Standardization Conferences.

A major difference between the Radiocommunication and Telecommunication Standardization Sectors is that the former holds regular treaty-making conferences known as *World Radiocommunication Conferences*. These replace the previous World Administrative Radio Conferences (WARCs) but have the same authority to revise the international Radio Regulations. Previously WARCs were held at irregular intervals, often with little indication as to when the next would be held. This made it difficult to plan changes in the Radio Regulations, for example the introduction of new frequency allocations, over a number of years. It also encouraged each conference to deal with a very large number of items simply because no-one knew when the next opportunity would be to address the issues. Under the new arrangements, WRCs will be held on a regular basis every two years. As mentioned above, there is a linkage between WRCs and the Radiocommunication Assemblies and so the APP-92 decided to hold the two events 'associated in place and time'. In 1993, both events were held in Geneva, with the Radiocommunication Assembly taking place first and slightly overlapping the WRC. The next Radiocommunication Assembly and WRC will be held in 1995, probably in Geneva and with the Radiocommunication Assembly preceding the conference but with no overlap.

So far, only one WRC has been held, that of November 1993. Under the new arrangements, each WRC will make proposals for the agenda for the following WRC and proposals for the provisional agenda for the following conference. These proposed agendas will be subject to ratification by the ITU Council. The 1993 WRC was limited in scope in that it had no authority to change the Radio Regulations but was merely asked to prepare the agenda for WRC 95 and the provisional agenda for WRC 97. The agenda for WRC 95 will concen-

the early signs are that the ITU is well placed to respond to the new challenges

trate on consideration of a report from a Voluntary Group of Experts which has been developing proposals for a major simplification of the Radio Regulations. Other items for 1995 include a review of the technical criteria for accommodating mobile satellite services in the spectrum and initial preparations on a number of issues with a view to substantive action at WRC 97. The latter subjects include a possible revision of the frequency plan for satellite broadcasting in the 12 GHz band.

A further unique element in the Radiocommunication Sector is the responsibility for the international notification, coordination and registration of radio-frequency assignments in accordance with the procedures laid down in the Radio Regulations. Previously this function was carried out by the International Frequency Registration Board (IFRB), which consisted of five elected full-time members, supported by a secretariat. In the new ITU structure, the routine responsibilities of the IFRB have been transferred to the Radiocommunication Bureau and, in particular, the Director of that Bureau. However, to maintain a degree of independent oversight of the process, to approve the Rules of Procedure used in the application of the Radio Regulations and to deal with any disputes regarding the application of those Rules of Procedure, the APP decided to establish a part-time Radio Regulations Board (RRB). This board will have nine part-time members. At present the duties of the RRB are being carried out by the IFRB members; the new part-time members will be elected at the Plenipotentiary Conference in the Autumn of 1994.

The Telecommunication Development Sector

The functions of the Telecommunication Development Sector are to fulfil the purposes of the Union and 'to discharge within its specific field of competence, the Union's dual responsibility as a United Nations specialised agency and executing agency for implementing projects under the United Nations development system or other funding arrangements so as to facilitate and enhance telecommunications development by offering, organising and coordinating technical cooperation and assistance activities'. The structure of the Telecommunication Development Sector is very similar to that of the Telecommunication Standardization Sector. The sector works through World Telecommunication Development Conferences, Telecommunication Development Study Groups and the Telecommunication Development Bureau headed by the elected director. World Telecommunication Development Conferences (WTDCs) are to be held every four years. However, unlike ITU-T, the Development Sector (ITU-D) has provision for regional telecommunication development conferences.

A number of regional telecommunication development conferences have been held already and the first World Telecommunication Development Conference was held in Buenos Aires in March 1994. There had been no previous event of this type. WTDCs act as a forum for the discussion and consideration of topics, projects and programmes relevant to telecommunication development and for the provision of direction and guidance to the Telecommunication Development Bureau. The Buenos Aires WTDC made an important start to this process by setting out a programme of future action encapsulated in the 'Buenos Aires Declaration on Global Telecommunication Development for 21st Century'. The conference produced 11 resolutions and two recommendations and agreed to establish two study groups, the first specifically on the subject of telecommunication development.

Again, it should be noted that although the WTDC made significant progress in determining future policies for telecommunication development, the conference itself had no treaty-making powers. Any decisions requiring the formal commitment of governments, for example on the distribution of funding between the sectors, will need to be taken to the Plenipotentiary Conference.

The advisory body for ITU-D corresponding to TSAG and RAG is the Telecommunication Development Advisory Board (TDAB). Although TDAB has similar duties to TSAG and RAG, as its name suggests it has a slightly different status and composition. Firstly, the requirement to establish the TDAB is built into the ITU Convention (as opposed to being covered by a Resolution of the APP as in the case in TSAG and RAG), and secondly the members of TDAB are appointed by the Director of the Bureau (although the membership is likely to be reviewed at the 1994 Plenipotentiary Conference). In contrast, participation in TSAG and RAG is open to all administrations and other participants in the work of the relevant sector. At the WTDC in Buenos Aires earlier this year, emphasis was placed on the need for cooperation between all three sectors of the ITU and it was recommended that joint meetings be held between the three sector advisory groups and board, together with the World Telecommunication Advisory Council (the senior advisory group covering the whole ITU).

Conclusions

The ITU has faced up to the challenge of the new telecommunications environment by reorganising its working methods and structure. The new arrangements have only recently been put in place and it is perhaps too early to say whether they will have a dramatic impact on the ITU. However, the early signs are that the ITU is well placed to respond to the new challenges; there is greater concentration on the need for cost-effective coordination, harmonisation and standardization, and there are very encouraging signs that the ITU is, perhaps for the first time in its long history, taking a serious look at long-term strategic planning. The structure and mechanisms are in place; it is now up to the membership and other participants, whether from governments, operators, manufacturers or international organisations, to build on that foundation and maintain the pre-eminence of the ITU in global telecommunications.

Biography

Michael Goddard
Radiocommunications Agency

Michael Goddard is Director of Spectrum Policy at the Radiocommunications Agency, the UK's radio regulatory authority. He has particular responsibility for frequency policy and international activities and has been involved in frequency management for most of his 25 year career since undertaking a Post Office Student Apprenticeship. Michael has led the UK delegation to several ITU World Administrative Radio Conferences, participated in many other ITU events, and was elected the first chairman of the ITU's Radiocommunication Advisory Group in November 1993. He was also the first chairman of the European Radiocommunications Committee of the CEPT (the European Conference of Postal and Telecommunications Administrations).

Glossary

APP Additional Plenipotentiary Conference (one-off event in 1992)
CCIR International Radio Consultative Committee
CCITT International Telegraph and Telephone Consultative Committee
DTI Department of Trade and Industry (UK)
FPLMTS Future public land mobile telecommunication systems
HLC High Level Committee (of the ITU) (no longer exists)
ICG Inter-Sector Coordination Group
IFRB International Frequency Registration Board
ITU International Telecommunication Union
ITU-D Telecommunication Development Sector of the ITU
ITU-R Radiocommunication Sector of the ITU
ITU-T Telecommunication Standardization Sector of the ITU
JCG Joint Coordination Group
RAG Radiocommunication Advisory Group
RRB Radio Regulations Board
TDAB Telecommunication Development Advisory Board
TSAG Telecommunication Standardization Advisory Group
WARC World Administrative Radio Conference
WCIT World Conference on International Telecommunications
WRC World Radiocommunication Conference
WTDC World Telecommunication Development Conference
WTSC World Telecommunication Standardization Conference

Concert Comes to the Global Stage

In June, BT and MCI unveiled Concert, their $1 billion joint venture company offering multinational companies global communications services from a single source, following clearance from the US Department of Justice.

Concert is the first company to provide such a service. It will put in place the world's most far-reaching advanced network, linking more than 5000 business customer access points in 55 countries by the spring of 1995. Syncordia, the premier global network outsourcing capability, will be merged into the new company.

Concert has established five global customer support centres. Located in Paris, London, Sydney, Tokyo and Cary (North Carolina), these centres currently manage the traffic of a number of BT's and MCI's largest multinational accounts. The centres provide uniform service, regardless of location, as well as a single point of contact for all enquiries and end-to-end problem ownership. They also offer multilingual customer support available 24 hours a day, seven days a week and multilingual, multi-currency support for billing. The Paris centre, for example, will accommodate customer queries in 15 European languages and dialects.

Concert brings together the complementary technical and marketing skills of two recognised powerhouses in telecommunications and information technology. With Concert, MCI and BT become a leading force in tearing down international barriers and establishing new benchmarks for excellence in worldwide telecommunications. Until now, multinational customers have waded through multiple correspondent networks, with varying levels of performance and service. Concert will develop products, services and applications delivered across a single network platform to meet the needs of multinational customers around the world.

The company, with the combined local-market presence and global reach of its parents, will enjoy the economies of scale necessary to provide its distributors, BT and MCI, with the ability to set a new standard in the international communications marketplace. Concert's 'best-of-breed' product portfolio of advanced value-added networking products and services initially includes:

- *Concert Global Virtual Network Services*—uniform enhanced features for virtual private voice, switched data and conferencing;
- *Concert Global Managed Data Services*—low- and high-speed packet switched and frame relay services; pre-provisioned, managed and flexible bandwidth services;
- *Concert Global Application Services*—advanced messaging, electronic data interchange (EDI) and audio/videoconferencing;
- *Concert Global Customer Management Services*—event and configuration management, troubleshooting and reporting, network performance analysis tools; and management, design, planning and security. Concert outsourcing services include facilities management and full enterprise network management under the Syncordia outsourcing services brand.

Both distributors are responsible for customer interface, front-line customer service and invoicing. They also continue to handle correspondent products and relationships. In turn, Concert provides its distributors with:

- a state-of-the-art global network;
- definition and development of end-to-end global products and management services;
- market research;
- technical design and implementation support;
- common architecture and systems for customer service; and
- an integrated billing system.

BT to Offer Internet Services

BT has announced that it is to offer the widest possible range of direct connections to Internet — the world's largest and fastest-growing data network. Starting this autumn, BT will offer everything from dial-up to high-speed broadband access to Internet, a cooperative of more than 10 000 autonomous organisations.

The connections to Internet will be introduced in phases, starting with a service aimed at UK-based customers who have existing corporate networks. Access methods that can be used are the public network, X.25, ISDN, private circuits, switched multi-megabit data services (SMDS) and frame relay. BT will offer corporate customers in mainland Europe access via frame relay and X.25. This will be followed by a service aimed at more occasional users, delivering dial-up services via a modem. BT's software will be suitable for the PC and Apple operating environments. BT will also be introducing a transit service for large academic networks and commercial Internet service providers.

To allow users to take maximum advantage of Internet, BT will be providing enhanced services such as E-mail and a navigation product to enable easy access to Internet services. Future products from BT will include a range of on-line services. In addition, BT will provide domain name system (DNS); file transfer protocol (FTP), which allows customers to connect to another computer and transfer, copy and list files on a directory; and Usenet news, a network of conference and interest groups.

For all access methods, BT will offer 24 hour, 365 day maintenance and support, and can provide all the necessary equipment for an end-to-end service, including the provision of on-site routers from Cisco, modems and software.

BT and Microsoft Alliance Links Workgroups

A computer package which creates a local area network (LAN), allowing 'global' communications, has been launched by BT and Microsoft. The package, ISDN for Workgroups, is the result of collaboration between BT and Microsoft, as well as German company Alcotec, which developed and supplied the software interface.

In place of costly traditional LANs to link personal computers (PCs), the new system uses ISDN lines to allow PCs to be located anywhere in the world where there is access to ISDN. BT and Microsoft believe that ISDN for Workgroups will revolutionise the concept of teleworking and working from home by allowing users to access the office network in just the same way as their colleagues on site. Businesses operating across more than one site can use the system to link separate LANs and, for a fraction of the price, create the effect of an expensive wide area network.

Microsoft's Windows for Workgroups software creates a 'workgroup' by interconnecting any number of PCs and allows them to share resources. Through BT's ISDN, a PC can be located anywhere and remain connected to the workgroup. For example, a contractor working at a remote site linked to ISDN could take along a PC and have direct access to the office network.

BT Launches Network Directory Disc

Any company that has a local area network (LAN) can now have access, from a single CD, to over 17 million telephone listings found in BT Phone Books, following the launch of Network Phone Disc™.

By loading Network Phone Disc onto a LAN, the user is provided with a readily-accessible means of finding telephone numbers from any PC throughout an organisation. Unlike the standard Phone Disc, the new system operates over a LAN, enabling users to increase the functionality of their existing network by making available key business resources company-wide.

When using a conventional phone book the customer's location details are required before searching for an entry, whereas with Phone Disc, a user can search the entire UK armed with just the name or first three characters of a name. For example, if a user enters a company name, all of that company's locations throughout the UK will be shown, normally within 10 seconds.

As the system can run over the majority of DOS computer networks including Novell, LanManager and Workgroups for Windows, many companies with CD-ROM facilities will already have the LAN necessary to run Network Phone Disc. Customers need not incur the cost of investing in new hardware or software.

One of BT's first customers for Network Phone Disc is BBC Radio which is using the networked service to research radio projects throughout the UK.

BT Announces Availability of Billing via EDI

BT has announced the availability of electronic data interchange (EDI) billing for its BT OneBill customers. Called *EDI*Bill*, the service offers all customers who receive consolidated telephone bill information (OneBill) the option of being billed via EDI.

EDI*Bill is presented in an easy-to-understand format which can be incorporated into financial planning systems. It gives customers a clear overview of network charges, without them having to consolidate or key-in copious volumes of data, and provides an invaluable insight into a company's network usage.

EDI*Bill is a strategic tool of value not only to the IT and communications team, but also finance, accounts, bought ledger and other administrative functions. When the OneBill is due, it is sent automatically from BT's computers to the customer's electronic mailbox.

For example, IBM, which helped in the trials of EDI*Bill, used to receive a total of 34 000 bills a year, which took 45 days to process. IBM now receives OneBill via EDI, which facilitates processing and enables an entire month's data to be fed directly into the system in just 30 seconds.

Several market-leading organisations took part in trials of the EDI*Bill service during its development. Most of these pilot customers realised savings ranging from £20 000 to £200 000 per year on bill processing alone.

New Satellite Service for the Financial Community

BT has launched a new shared satellite network service for the financial community. The new network, called FINESS (Financial European Satellite Service) is the first shared satellite communications service to enable the simultaneous broadcast of financial information in the form of both data and voice commentary to financial institutions throughout Europe.

FINESS is fully owned and managed by BT. As a leased service it offers users peace of mind and avoids the capital investment required when using some other networks. Network management is provided remotely from BT's service management centre (SMC) in London, 24 hours a day, 365 days a year.

To use the new service, brokers and information providers transmit data and voice via terrestrial circuits to BT's London Teleport. Signals from various sources are then multiplexed together and uplinked via an 11 m dish to the EUTELSAT II flight 2 satellite located at 10 degrees east.

BT is initially offering the service at two data speeds of 9·6 kbit/s and 19·2 kbit/s. Voice commentary is digitally compressed and will be transmitted at 9·6 kbit/s.

The satellite transmits signals across a footprint extending from Iceland to North Africa and Ireland to Israel. Within this satellite footprint, dishes as small as 90 cm are typically installed at customers' premises to receive the FINESS service (although local restrictions may apply).

At the receive sites, each installation will include a decoder, which is activated to receive designated services, remotely, by the SMC.

Each site's installation is programmed to decode only the data that it is allowed to receive. This data is presented at the decoder interface as a standard RS232 signal. Voice circuits are passed through a digital-to-analogue converter and then delivered to customers via their public address system.

Satellite-based services are independent of local infrastructure and can therefore be installed in remote and developing regions.

FINESS is more reliable than many other services, offering 99·9% availability and an error rate of 1 in 10^7. The satellite capacity utilised is non-pre-emptible and restorable, enabling full recovery in the unlikely event of a transponder or satellite breakdown.

New sites can easily be added to the network and existing sites can readily be provided with further services via FINESS. These installations can be quickly authorised to receive new services, as BT's SMC can remotely programme the decoder to accept new signals. New financial service providers can also easily take advantage of the FINESS service by presenting their financial information to BT's London Teleport for onward transmission to the established FINESS community.

Woolwich Chooses BT

In July, the Woolwich Building Society awarded BT the contract to develop and manage a new national communications network. Six hundred sites will be connected to the network, making the building society BT's largest Global Network Services (GNS) frame relay customer at the time of the announcement.

BT's solution provides the Woolwich with the benefits of a fully managed network which is both technologically flexible and cost-effective. In a 5 year contract, the first task is to interconnect the local area networks at each Woolwich branch, and all the subsidiary companies, to the society's national data centre.

A key factor in BT's successful tender was its 'future-proofing' capacity—the flexibility and adaptability of its managed network to deliver new technologies and services as they become required by the customer. The 1986 Building Societies Act has transformed the Woolwich's business practices and broadened greatly the range of services offered. A traditional data network is unable to support all the new applications required, and with further expansion of services predicted, a costly replacement network could quickly become obsolete. By using BT's technology, the Woolwich can develop its national communications network to meet any future changes in its business practice with minimal capital investment.

BT's service for The Woolwich consists of a shared platform frame relay service and routers with managed ISDN 2 lines for added resilience. This provides full UK coverage with management, resilience and redundancy inherent in the network service package.

The token ring LANs in each Woolwich branch will be connected to a Cisco router which in turn will be connected to the nearest GNS node for frame relay switch access.

The frame relay path was selected for its immediate appropriateness in interconnecting the branch LANs and its capacity to adapt to emerging technologies such as asynchronous transfer mode (ATM), expected to break through in the retail finance sector over the next few years.

Incorporating ISDN 2 lines represents the ultimate in network resilience to the Woolwich. The ISDN 2 lines provide a 64 kbit/s dial-up digital path direct to the hub routers if required. In addition, the ISDN lines offer the branches the bonus of a built-in traffic overflow facility and access to other services such as videoconferencing and voice.

notes and comments

Erratum

In Keith William's and Bryan Law's article 'ISDN Access Signalling' published in the January 1994 issue of the *Journal*, additional reference should have been made to another article: 'ISDN Signalling Standards' by D. R. Davies, K. Baughan and T. Kent published in *British Telecom Technology Journal*, April 1986, **4**(2), p. 26.

industry news

Calling for Agreement

OFTEL—The Office of Telecommunications—has urged the UK telecommunications industry to cooperate to ensure that customers get the maximum benefit from technological developments.

After publication of a statement on customer interface standards for delivering calling line identification (CLI), Director General Don Cruickshank made it clear that he did not currently have the power to enforce a single standard for this service.

Said Mr. Cruickshank: 'the interests of consumers must override the narrow interests of producers in areas where cooperation is required to provide services that customers want.

'I am actively investigating what action OFTEL can take, either on its own or with the DTI, to reduce potential customer confusion and the anti-competitive effects created by the production of a multiplicity of incompatible customer equipment.'

European IS Outsourcing Market to Double

The European market for information systems (IS) outsourcing is expected to nearly double in four years, from an estimated \$4·3 billion in 1993 to \$8·3 billion in 1998.

This forecast comes in a new report from international market research publishers Frost and Sullivan which has based its figures on growth expectations of suppliers and on the outsourcing activity planned by respondents to a major survey of customer attitudes and intentions.

Fifteen national markets were covered in the study, which identifies the UK, France and Germany as the largest markets, each worth around \$1 billion. Spain and Italy are identified as the fastest-growing markets, with annual growth rates of 22 per cent and 18 per cent respectively.

In terms of services, the report reveals that the bulk of present activity lies in systems management, but faster growth is expected in applications management (20 per cent per annum) and network management (25 per cent per annum).

The largest current platform is found to be the data centre, but only modest annual growth of around 7 per cent is expected in this sector, in contrast to forecast growth of 27 per cent for desktop systems and 21 per cent growth for networks.

The Use and Provision of Satellite Communications Systems

The Commission of the European Communities has adopted a Communication concerning future regulatory and policy actions required in the context of the provision of satellite communications capacity for broadcasting, telecommunications and mobile/personal communications through the launching and operation of satellite systems, as well as in the context of the access to the capacity of these satellites systems by the Union's licensed providers of satellite-based business, video and mobile services.

The provision of satellite capacity is currently dominated by the national telecommunications operators through their involvement in the international satellite organisations INTELSAT, INMARSAT and EUTELSAT, and through their ownership in national satellite systems. Access to this capacity in Member States is largely controlled by the national telecommunications operators. This control is becoming a bottleneck which is critical to private satellite service operators which have to obtain satellite capacity as a essential facility in their business.

Also, the increasingly competitive telecommunications environment is now becoming subject of detailed discussions in the international satellite organisations, set up under intergovernmental treaty. These organisations are seriously considering fundamental changes and possibly privatisation. It is in the Community's interest to come forward with a common position regarding further progress.

The Communication thus sets out the Commission's intended policies on what the future organisation of the space segment supply market and its access arrangements should be in the Union, how this can be realised, what arrangements are necessary, and which regulatory and policy actions are required from Member States and the Union. The aim is to ensure the competitiveness of European undertakings on the global satellite communications market.

In the Communication, the Commission has outlined the main objectives and policy lines as follows:

- ensuring throughout the European Union direct access to the space segment, including in particular space segment provided by the international satellite organisations;
- joint action by the Member States in the reform of the international satellite organisations and in particular EUTELSAT;
- joint management in the future of the space segment as a common resource of the Union, in particular concerning future applications to the International Telecommunication Union for orbital positions and related coordination procedures and availability of radio frequencies;
- the establishment of measures in order to ensure comparable and effective access to third countries, in parallel with the Union's market liberalisation; and
- inclusion of satellite-based services in programmes for transeuropean networks as a major priority, in particular, with regard to the emerging technologies.

New Alliance

France Telecom, Deutsche Telekom and Sprint Corporation have announced plans to form a global partnership designed to offer seamless global telecommunications services to business, consumer and carrier markets worldwide. Services will include global international data, voice and video business services for multinational and large business customers, as well as small companies with international communications needs; international services for consumers, initially based on card services for travellers; and 'carrier's carrier' services, providing international transport services for other carriers.

Over time, the services will expand to address the changing dynamics of global, regional and national markets. The partnership plans to develop national networks to provide long-distance services in other countries.

As part of the agreement, France Telecom and Deutsche Telekom will invest a total of approximately $4·2 billion in Sprint over two years, become shareholders of the company and hold a 20 per cent equity investment in the company.

The intention of the three partners is to combine planning for global network facilities allowed by regulation. They believe there are opportunities to work together in other advanced markets and expect to sign a definitive agreement by the end of the year.

ITU Agrees 'V.fast' Standard

The International Telecommunication Union has finally adopted the standard for future high-speed modems.

The adoption of the new standard will give the industry the go-ahead to offer new products using high-performance data transfer technology. Work on the drafting of the standard started some three years ago by Study Group 14 of the Telecommunication Standardization Sector, ITU-T (formerly known as CCITT) and was known as 'V.fast' by industry experts in its development stage. The standard, called V.34, will surpass the current technology used in data transfer via traditional telephone lines.

V.34 future modems will transfer data at twice the current technology. By increasing transmission speed, they will drastically cut down the time needed by computers and faxes and in turn lower user telephone bills. These new modems will have variable data transmission capacity ranging from 2·4 kbit/s up to 28·8 kbit/s.

At the same meeting of Study Group 14, a new standard—Recommendation V.18—which, for the first time, recognises the communication needs of the deaf and hard of hearing, was also approved. This Recommendation, with its capability to interwork with all existing devices, provides the platform on which a universal standard communication device can be built.

OFTEL Calls for Improved Consumer Codes

Don Cruickshank, Director General of Telecommunications has called for major improvements in the Consumer Codes of Practice published by the network operators. The Director General said he expected to see new Codes from BT, Mercury, Kingston Communications and other operators to ensure that the new Codes contain the information customers need.

Don Cruickshank commented: 'With the increasing number of operators to choose from, it is essential that every operator provides a clear guide to what customers can expect in terms of standards of service and getting problems resolved. The Consumer Codes of Practice—required under each operator's licence—are designed to do this job. I am looking for the new Codes to be easier to read and understand and to include more information on important issues such as disconnections and fault repair.

'As part of our drive to improve the Codes, OFTEL, has produced a 'good practice guide' to help operators focus on customers' interests. We see these Codes as public statements of the operators' commitment to their customers.'

Global Telephone Trial Success

Field trials conducted by INMARSAT in Ipswich using two specially-equipped aircraft have demonstrated that the INMARSAT-P global hand-held telephone can automatically select the best quality link among the two or more satellites visible in the sky.

The design of the INMARSAT-P system, which uses an intermediate circular orbit of 10–12 satellites, makes it possible for at least two of the satellites to be accessible at any one time. If a link is disrupted by shadowing, from trees or buildings for example, electronics within the telephone handset which combine the signals from all the available satellites automatically select the best satellite link.

The trials were conducted in outdoor urban and rural environments as well as inside buildings, to test as many conditions which might obstruct a telephone call as possible. BBC newscasts were transmitted from the two aircraft to simulate the transmission of telephone calls.

INMARSAT-P is one of a family of personal mobile satellite communications services being developed between now and the turn of the 21st century under an initiative known as *Project 21*.

Call for IT Standards in EU

The Bangemann group of IT industrialists have said that EU-wide standards were imperative for the development of Europe's information highways. Europe did not lack an IT infrastructure, but national networks were often incompatible.

The 19-member committee was set up by the Commission earlier this year to advise before Corfu on the development of the IT highways outlined in the EC's White Paper on growth. The group also attacked national telecommunications monopolies saying that they should be the exception and not the rule.

The industrialists recommended: the removal of political and budgetary restraints on telecommunications operators expanding commercially; the creation of an EU-regulatory body; a review of EU standardisation procedures; the reduction of tariffs; and the creation of a regulatory framework for satellite communications.

Proposed Rule Change

The following Rule change has been formally advised to Members via their Local Centre Secretary, in accordance with Rule 45, which states that the Rule shall apply from the end of the month following the month in which the proposed change is advertised, unless not less than 2% of the membership demand that the proposal be put to a ballot of all Members.

The proposed new Rule redefines the conditions for retired membership, and is intended to encourage more Members to take up the opportunity of retaining their links with IBTE, while enabling IBTE to control costs more effectively.

The new Rule will read as follows:

16 Any Member retiring or retired from BT or its subsidiaries, who wishes to become a Life Member of IBTE, and will not be undertaking responsible work in the field of telecommunications, will be able to do so at no cost. This Member will continue to enjoy the privileges of membership, including attendance at open meetings of a Centre of his/her choice, but with the exception that he/she will not be eligible to receive the *Journal* except by payment at the BT employee rate or such sum as shall be determined by Council as appropriate having regard to the costs of production of said *Journal*.

(*a*) All existing commuted Life Members as at 1 September 1994 shall be entitled to receive the same benefits, including regular mailings of *Newsview*, and, in line with previous entitlements, will be invited to take up one of three options:

—to continue as a Life Member, but elect not to receive the *Journal*;
—to continue as a Life Member, and to receive the *Journal* at no cost; or
—to recognise that they have already received full value for their one-off commuted payment, and to elect to make a voluntary payment for the *Journal* at the BT employee rate or such lower sum as Council deems appropriate.

BRITISH TELECOMMUNICATIONS ENGINEERING ISSN 0262-401X

Published in April, July, October and January by *British Telecommunications Engineering Journal*, 2–12 Gresham Street, London, EC2V 7AG. (Formerly *The Post Office Electrical Engineers' Journal* Vols. 1–74: April 1908–January 1982.)

The Board of Editors is not responsible for any statements made nor the opinions expressed in any of the articles or correspondence in this *Journal*, unless any such statement is made specifically by the Board.

Printed in Great Britain by Unwin Brothers Limited, The Gresham Press, Old Woking, Surrey, GU22 9LH.

Subscriptions and Back Numbers
Price: Single copy—companies, universities, libraries and other bodies £9·25 (£10 including postage UK, £11·25 including postage overseas); private individuals (paid by personal cheque) £5·00 (£5·75 including postage UK, £7·00 including postage overseas). Annual subscriptions (including postage)—companies etc. £40·00 UK, £45·00 overseas; private individuals (paid by personal cheque) £23·00 UK, £28·00 overseas.

Overseas customers can pay by sterling drafts drawn on London banks or by sterling Eurocheques.

Price to BT staff: £2·25 per copy.

Back numbers can be supplied if available at the single copy rates given above.

Orders, by post only, should be addressed to *British Telecommunications Engineering Journal* (Sales), PP: G012, 2–12 Gresham Street, London EC2V 7AG.

Remittances for all items (except binding) should be made payable to '*BTE Journal*' and should be crossed '& Co'.

Advertisements
All enquiries relating to advertisement space reservations should be addressed to The Advertisement Manager, *British Telecommunications Engineering Journal*, PP: G012, 2–12 Gresham Street, London EC2V 7AG. (Telephone: 071–356 8050. Fax: 071–356 7942.)

Communications
All communications should be addressed to the Editorial Office, *British Telecommunications Engineering Journal*, PP: G012, 2–12 Gresham Street, London EC2V 7AG. (Telephone: 071–356 8050. Fax: 071–356 7942.)

THE INSTITUTION OF BRITISH TELECOMMUNICATIONS ENGINEERS

IBTE Local-Centre Contacts

The following is a list of Local-Centre contacts to whom enquiries about the Institution may be directed. Where no Membership Secretary details are shown, membership enquiries should be directed to the Local Secretary.

Centre	*Local Secretary*	*Membership Secretary*
Aberdeen	Jim Martin (0224 752688)	George Cameron (0224 752857)
East Anglia	Terry Birdseye (0702 373330)	Keith Phillips (0206 717244)
East Midlands	Dave Bostrom (0691 666501)	Ian Bethell (0602 478587)
Lancs & Cumbria	Trevor Hughes (0254 666261)	John Nicholson (0254 666261)
Liverpool	John Johnson (0925 415705)	Clive Maude (051–229 4074)
London	Mike Burgess (0908 693141)	Peter Martin (071–492 6138)
Manchester	David White (061–200 9231)	Mike Ku (061–200 9343)
Martlesham Heath	Peter Shaw (0473 644378)	Ian Gilbert (0473 645500)
North Downs	Nick Smith (0227 474823)	
North East	Richard Jackson (091–261 3560)	
North Wales & Marches	Peter Clay (0743 274205)	
Northern Ireland	Allan MacDougall (0232 215301)	Brenda Sharvin (0232 320092)
Scotland East	Graham Neilson (031–668 5878)	Lewis Shand (031-668 5593)
Severnside	Ian Davis (0272 206701)	Dean Shephard (0272 206457)
South Downs	Graham Lindsay (0273 222633)	
South Midlands	Tony Atkinson (0234 277874)	Jim Coley (0234 274043)
South Wales	Peter Coleman (0222 514524)	Keith Voyce (0222 555075)
Staffordshire	Ray Piper (0785 813483)	Paul Field (0785 762419)
Thames Valley	Stephen Graham (0252 337104)	Dave Prout (0734 501244)
West Midlands	Richard Harwood (0527 550252)	Paul Jervis (021–233 8162)
West Scotland	Graham Wellby (041–220 3997)	Jill Maxwell (041–220 5837)
Westward	Rob Rand (0392 267114)	Chris Gould (0392 267113)
Yorks & Lincs	Steve King (0532 378124)	Paul Horncastle (0532 760242)

Associate Section Zone Committee Contacts

The following is a list of Associate Section Zone contact points to whom enquiries about the Associate Section should be directed.

Zone Committee	*Contact*	*Telephone No.*
London	Terry McCullough	081–456 6798
Midland	John Sansom	0533 513155
North East	Geoff Jenkinson	0532 378163
Northern Home Counties	Andy Edmonds	0473 645171
Northern Ireland	Brian Walker	0232 232565
North West	Phil Holt	061–600 8111
Scotland	Bill Lindsay	0382 302629
Southern Home Counties	Martin Harvey	0424 754385
South West	John Horton	0272 329303
Wales	Ferg Brown	0952 249886

IBTE National HelpLine — 071-356 8008

5th IEE Conference on
Telecommunications

Brighton, 26–29 March 1995

CALL FOR PAPERS

The conference seeks to reflect and explore new developments, evolutionary trends and current best practice in telecommunications in terms of technology, networks, services and applications, and business issues.

The conference will deal with all aspects of research, development, deployment and applications of telecommunications systems, equipment, networks and services and will embrace also the broader operational, organisational, regional, regulatory and business issues inherent in telecommunications today.

Papers are invited for inclusion in the conference programme addressing any aspect of telecommunications as a technology or as an industry within the scope indicated above. Papers should present the results of original research or development, provide a critical, insightful review of the state of the art in a relevant area or offer a case study of broad relevance on a particular application of telecommunications.

Full papers should be submitted to the IEE Conference Services Department on or before 1 November 1994. Papers should not exceed ***five A4 pages*** and should be prepared in accordance with ***'Instructions for Authors'*** available from the Conference Secretariat. Papers accepted for inclusion in the programme will be published in the Conference Proceedings from the ***'camera-ready'*** copy supplied by an author.

Potential contributors are asked to indicate a provisional title/topic for their papers to the Secretariat, who will provide 'Instructions for Authors' on the preparation of the paper. The details of the Conference Secretariat are given below:

Jane Chopping
ICT '95 Secretariat
Conference Services
Institution of Electrical Engineers
Savoy Place, LONDON WC2R 0BL, UK

Tel: +44 71 344 5477
Fax: +44 71 497 3633
Telex: 261176 IEE LDN G

VOLUME 13 PART 2 JULY 1994

Contents