COMMUNICATIONS FOR TEXT PROCESSING: WITH APPLICATION TO ELECTRONIC INFORMATION SERVICES

A Report to the
Department of Communications
Ottawa, Ontario, Canada

by

Dr. Robert W. Donaldson, Professor
Department of Electrical Engineering
University of British Columbia
Vancouver, British Columbia
Canada

January 1977

COMMUNICATIONS FOR TEXT PROCESSING:
WITH APPLICATION TO
ELECTRONIC INFORMATION SERVICES

A Report to the
Department of Communications
Ottawa, Ontario, Canada

by

Dr. Robert W. Donaldson, Professor
Department of Electrical Engineering
University of British Columbia
Vancouver, British Columbia
Canada

January 1977

ABSTRACT

The economic viability of interconnecting geographically separated computers and user terminals via communication links indicates a rather sudden emergence of a number of related electronic information services (EIS).

Following a brief consideration of three typical services, namely word processing, electronic funds transfer and electronic mail, the problem of analysing, designing and evaluating electronic information systems is examined and subsequently subdivided into four separate but interrelated aspects, including: computer-communication network analysis and design, man-machine interfacing, organization and management of distributed data-bases and socio-economics.

The economic aspects are further subdivided into three separate but related aspects, namely: cost of supplying electronic information services, demand for services and government options and alternatives. Each of these topics is considered in some detail.

Because EIS costs depend in large measure on data-traffic volume, the potential EIS data sources are carefully examined. Coding and reconstruction of speech, images (including color), text and facsimilie are discussed, as well as machine recognition of speech and printed characters, and machine verification of speakers. These discussions are directly relevant to man-machine interfacing which is also considered.

Within the framework proposed for analysis, design and evaluation of electronic information systems, further study of the following matters is deemed useful:

1. Socio-economic and technological aspects of implementing various specific electronic information services, including electronic mail.
2. Analysis and design of computer-communication networks (this is to be the subject of a separate report by the author).
3. Organization and management of distributed data bases.
4. Government's role in encouraging and using electronic information services.
5. The relationship and potential integration of existing services, particularly cable television, with new electronic services.
6. Effects of electronic information services on other economic sectors, including the energy sector.
7. Selection of $p \simeq 10^{-4}$ as the required random bit-error probability for data communication channels, with channel encoding being used to achieve lower error probabilities when required.
8. Design of user terminals, with a view towards integration of current relevant knowledge in electronics, computer technology, software engineering, human information processing, speech processing, image processing and various EIS applications.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	
I-1	Electronic Information Services	1
I-2	Approaches to Electronic Information System Design and Evaluation	1
I-3	Objective of the Present Report	2
I-4	Outline and Summary of the Present Report	2
I-5	Recommendations for Further Study	5
I-6	Review of Others' Works	6
I-7	References	7
II.	ELECTRONIC INFORMATION SERVICES	
II-1	Introduction	8
II-2	Word Processing	8
II-3	Banking and Electronic Funds Transfer (EFT)	11
II-4	Electronic Mail	13
II-5	Observations and Comments	18
II-6	References	19
III.	FACTORS AFFECTING THE ANALYSIS, DESIGN AND PERFORMANCE OF ELECTRONIC INFORMATION SYSTEMS	
III-1	Isolation of the Major System Factors	23
III-2	Computer-Communication Network Design	23
III-3	Man-Machine Interfacing	26
III-4	Data-Base Organization and Management	26
III-5	Socio-Economic Issues	27
III-6	Performance Evaluation	28
III-7	References	28
IV.	ECONOMIC ASPECTS OF ELECTRONIC INFORMATION SERVICES	
IV-1	Introduction	31
IV-2	Costs of Supplying Electronic Information Services	33

Page

IV-3 Demand for Electronic Information Services. . . 38

IV-4 Economic Options for Governments. 40

IV-5 References. 45

V. ANALYSIS, CODING, RECOGNITION, SYNTHESIS AND RECONSTRUCTION OF SPEECH; SPEAKER VERIFICATION

V-1 Speech and Electronic Information Services. . . 48

V-2 Conventional Digital Coding and Reconstruction of Speech 49

V-3 Adaptive Differential Encoding of Speech. . . . 53

V-4 Speech Analysis and Synthesis 59

V-5 Machine Recognition of Speech 63

V-6 Speaker Verification. 65

V-7 Summary and Conclusions 66

V-8 References. 67

VI. IMAGE CODING, STORAGE, TRANSMISSION AND RECONSTRUCTION

VI-1 Images and Electronic Information Services. . . 72

VI-2 Image Properties, Characterizations and Distortion Measures. 72

VI-3 Evaluating Image Quality. 75

VI-4 Coding and Reconstructing Monochrome Images . . 78

VI-5 Coding and Reconstructing Color Images. 81

VI-6 Concluding Remarks. 83

VI-7 References. 85

VII. CODING, STORAGE, TRANSMISSION AND RECONSTRUCTION OF FACSIMILIE, TEXT AND TEXT SEGMENTS

VII-1 Facsimilie, Text and Electronic Information Services 89

VII-2 Properties of FAX Information Sources 89

VII-3 Run-Length Coding 90

VII-4 Continuous Tone Images on Bilevel Displays. . . 92

Page

VII-5 Direct Coding of Text. 93

VII-6 Coding Text Segments 94

VII-7 Optical Character Recognition. 96

VII-8 Conclusions and Summary. 98

VII-9 References 98

VIII. DESIGN AND EVALUATION OF USER TERMINALS

VIII-1 Problem Definition 102

VIII-2 User Considerations. 102

VIII-3 Hardware 107

VIII-4 Software 110

VIII-5 Conclusions and Suggestions for Further Study. 115

VIII-6 References 118

LIST OF FIGURES

		Page
Fig. 2-1:	Electronic mail system.	14
Fig. 2-2:	Pattern of mail flow (U.S.A. 1974).	16
Fig. 4-1:	Typical supply and demand curves.	32
Fig. 5-1:	Conventional PCM system	50
Fig. 5-2:	Quantizer characteristic.	51
Fig. 5-3:	Adaptive differential PCM system (ADPCM). . . .	54
Fig. 5-4:	SNR vs bit rate for various speech coders . . .	58
Fig. 5-5:	Schematic of the human vocal tract	60
Fig. 5-6:	Vocal tract analog.	61
Fig. 6-1:	Image coding and transmission/storage system. .	76
Fig. 8-1:	Request for file.	112
Fig. 8-2:	Displayed file.	112
Fig. 8-3:	Menu-selection example.	113

I INTRODUCTION

I-1 Electronic Information Services

It is now economically viable to interconnect geographically separated computers and user terminals via communication lines. Thus, users at various locations are potentially able to communicate with one another and with various computers.

It is argued in Chapter 4 that the cost of a specific electronic information service (EIS) will be reduced if a variety of related services such as electronic mail, word processing, electronic funds transfer and community information services are available to share various common costs. Thus, the emergence of a number of different services over a relatively short time frame is a potential reality. Varying estimates as to the eventual size of the EIS industry are available. Most agree that the industry will contribute significantly to the gross national product of our post-industrial society.

Information is unlike any other commodity. It motivates people's actions and its manipulation by special interest groups can produce a variety of effects, not all of which are desirable.

How does one design electronic information services and systems? How are the systems to be evaluated? What are the salient problems and issues?

I-2 Approaches to Electronic Information System Design and Evaluation

The answers to the questions in the previous paragraph are not immediately obvious; nor is the approach to be taken to obtain the answers obvious.

One approach is to list all the relevant questions, and then attempt to answer them. The difficulty is that there is no way to know whether or not all of the important questions have been asked.

An alternative approach is to examine potential applications and thereby extract the important issues. Again, one can't be certain of asking all the right questions, but the chance of omitting important ones seems less if enough applications of sufficiently broad scope are adequately considered. We adopt the second approach.

I-3 Objective of the Present Report

Our objective in the present report is to examine in an integrated way the emerging EIS industry, to develop a coherent, logical and pragmatic approach for isolating and dealing with the significant issues and problems, and to examine in some detail those aspects which relate directly to text processing. Our interpretation of text is the broad one which includes speech and visual images.

I-4 Outline and Summary of the Present Report

Chapter 2 includes a brief examination of word processing, electronic funds transfer and electronic mail. Although further study of each of these as well as other listed applications is warranted, we are able to isolate the following four separate but related areas of importance:

1. Computer-communication network and analysis and design
2. Man-machine interfacing

3. Organization and management of distributed data bases
4. Socio-economics

In Chapter 3, each of these areas is further subdivided, and problems to be solved are identified. The network design problem, which is of particular relevance at this time, is to be considered in detail in a separate report by the author. System software, whose relative cost contribution is large and growing, also warrants further study; the data-base technology aspects are of particular urgency.

Economics will ultimately determine the time, nature and extent of development of various electronic information services. Chapter 4 considers economic matters under three separate but related headings:

1. Cost of supplying electronic information services
2. User demand for services
3. Economic options of governments

Calculation of supply costs is shown to be uncomplicated in principle, although practical problems exist, including the forecasting of future costs and technological developments. Readily available data detailing user demand for various services is virtually non-existent. Because many of the services are relatively unfamiliar to potential users, obtaining reliable demand data presents difficulties. Issues facing governments include consideration of how and to what extent to encourage development of various services, potential effects of various services on other economic sectors including the energy sector, and potential relationships of existing information services, particularly cable television, with proposed services. These and other issues reinforce the need for accurate estimates of EIS supply costs and user demand.

Chapter 5 deals with the analysis, coding, recognition, synthesis and reconstruction of speech, as well as speaker verification. If the actual acoustic signal is to be coded into binary digits for subsequent reconstruction, then adaptive differential 10 Kbs encoders of acceptable cost and complexity will yield speech which is subjectively preferable to 4-bit PCM speech. Vocoders which encode speech phonemes are potentially more efficient, but currently too expensive. Speech synthesis is applicable now for some situations including voice response to terminal enquiries. Recognition of words spoken from limited vocabularies is feasible, as is speaker verification at approximately 95% accuracy. This accuracy is acceptable for most day-to-day applications. Considerable improvements are likely in word recognition, speech synthesis and speaker verification.

Chapter 6 reviews coding and reconstruction of monochrome and color images. We consider mainly the faithful reproduction of images, and not their synthesis from stored text. Marginally cost effective schemes now exist for coding both monochrome and color images using 2 bits per picture element, which implies 1.3×10^5 bits per image sampled at a 256 x 256 spatial sampling rate. Significant bit rate reductions seem unlikely until we understand how image building blocks are specified and assembled to generate images. This same lack of understanding is a serious impediment to computer synthesis of all but the simplest images.

Chapter 7 considers images all of whose picture elements are either black or white. Run-length coding of such images, called facsimilie (FAX), is relatively efficient, resulting in bits per picture element codes of 0.1 for a typical business letter and 0.4 for weather maps. Direct character-by-character transmission of text, however, is more efficient than run-length coding by a factor of 16 for a page of double-spaced text and 28 for a typical business letter. However the data channel error requirements for directly transmitted text are much more severe; bit error probabilities of 10^{-10} or 10^{-12} are required for text, while 10^{-4} is adequate for digitally coded FAX, speech and images.

Chapter 8 deals in an introductory way with the design of user terminals. The topic is a vast one which clearly warrants further detailed study to provide meaningful integration of knowledge from a number of diverse disciplines.

The various chapters can be read independently of each other. However material in different chapters is interrelated. Thus, the material in Chapters 5, 6, and 7 is relevant to the economic considerations in Chapter 4, since the cost of transmitting or storing a message depends in part on the number of bits needed to represent a message. This same material is relevant to the design of user terminals, since communication across the interface can be via speech, images, facsimilie and text.

The broad array of subject matter inherent in the EIS industry motivates subdivision of the material into relatively disjoint, manageable pieces. We suggest that the subdivisions proposed throughout this report constitute a useful contribution to the present work.

I-5 Recommendations for Further Study

Listed below are those topics which have been identified as deserving further study.

1. Socio-economic and technological aspects of implementing various specific electronic information services, including electronic mail.
2. Analysis and design of computer-communication networks (to be the subject of a separate report by the author).
3. Data-base organization and management, particularly distributed data-base technology.
4. The role of government in encouraging the development and use of various electronic information services.
5. The relationship and potential integration of existing services, particularly cable television, with new electronic services.
6. Effects of electronic information services on other economic sectors, including the energy sector.
7. Selection of $p \simeq 10^{-4}$ as the required random bit-error probability for data communications channels, with channel encoding being used to achieve lower error probabilities when required.
8. Design of user terminals, with a view towards integration of current relevant knowledge in electronics, computer technology, software engineering, speech processing, image processing, human information processing and various EIS applications.

I-6 Review of Others' Works

Because of the diversity of the subject matter of the present report, specific references to others' works appear at the end of each chapter. References cited have been selected with care, and are intended as sources of useful additional information. An examination of these reveals great diversity of subject matter. The following specific references [1-13] are offered as a means of gleaning a minimal general exposure to much of the subject matter of interest.

I-7 References

1. N. Abramson and F.F. Kuo, Ed., Computer-Communication Networks. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

2. L. Kleinrock, Queueing Systems, Vol. 2: Computer Applications. New York: Wiley, 1976.

3. J. Martin, Systems Analysis for Data Transmission. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

4. J. Martin, Design of Man-Computer Dialogues. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

5. J. Martin and A. Norman. The Computerized Society. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

6. J. Martin, Principles of Data-Base Management. Englewood Cliffs, N.J.: Prentice-Hall, 1976.

7. J.J. Verma, Ed., Proceedings of the Third International Conference on Computer Communication, Toronto, Canada, August 1976.

8. Proceedings of the Second International Conference on Computer Communications. Stockholm, Sweden, August 1974.

9. S. Winkler, Ed., Computer Communication, Impacts and Implications: Proceedings of the First International Conference on Computer Communication. Washington, D.C., October 1972.

10. W.M. Newman and R.F. Sproull, Principles of Interactive Computer Graphs. New York: McGraw Hill, 1973.

11. R.L. Pickholtz, "Computer communications: an emerging discipline," IEEE Trans. Commun., Vol. COM-25, p. 1, Jan. 1977. (This issue is devoted entirely to computer communications and contains 15 excellent papers.)

12. W.W. Chu, Ed., Advances in Computer Communications. Dedham, Mass.: Artech House, 1975.

13. P.E. Green, Jr. and R.W. Lucky, Computer Communications. New York: IEEE Press, 1975.

II ELECTRONIC INFORMATION SERVICES

II-1 Introduction

Our intent here is not to conduct a detailed study of the various electronic information service applications. Such a study is potentially useful but beyond the scope of the present work. Rather, we examine three applications which illustrate the salient aspects of an emerging industry. The three, cited below, were selected because they are either in use or their use is iminent, their potential economic and sociological impact is considerable, and, taken together they expose important aspects of the electronic information services industry. The applications considered include:

1. word processing
2. banking and electronic funds transfer
3. electronic mail

Listed below are other EIS applications together with a few of the many available references which provide additional details.

4. Reservation systems for airlines, hotels, and other transportation and hospitality facilities [1,2].
5. Teleconferencing [3-5].
6. Computer-aided instruction, including via home terminals [1,6-9].
7. Health care and hospital information systems [9-14].
8. Remote accessing of files detailing library books available, transportation schedules, weather forecasts, and coming events [1,6,9,15].
9. Automated reading of utility meters [1,6].

II-2 Word Processing

Word processing implies different activities to different users. Under its umbrella we include assembly of documents from stored text segments, text editing, spelling correction and final document preparation. Although the specific configuration of word processing systems varies widely, a system would normally include a keyboard, a display such as a CRT, digital storage, logic for searching, retrieving, updating, and editing stored information, and a hardcopier.

A typical system user would be a lawyer who must prepare affidavits, agreements and court forms which consist largely of standard paragraphs or minor variations thereof. For example, a mortgage would be prepared by manually typing the name, address, and occupation of the parties. The appropriate paragraphs would then be retrieved from storage, edited, and displayed on a screen or on sheets of paper. The resulting draft would be proofread, edited further, and then accepted.

Several individuals could work simultaneously on preparation of a single document provided each worked on a different segment. The various individuals could be in different geographical locations, so long as each had a terminal for communication with a central processing unit. Thus, document preparation from one's own home is potentially feasible.

A readable description of a computer program (called WYLBUR) for assembling and editing text appears in [16]. Following limited training users are able to use WYLBUR to correct text as it is typed, to insert, modify and delete parts or all of any line or line sequence, to store text segments, to retrieve stored segments for examination or insertion into a larger body of text, and to align text for final document preparation. Text segments are referred to by line number or by association. For example, a user can specify that all references to J.E. SMITH in lines 21 to 356, inclusive, or in paragraphs 2 to 10, inclusive, be replaced by A.B. JONES.

Somewhat surprising is the fact that few quantitative measures of performance of various word processing configurations are available. Oren [17] developed a mathematical model of the document assembly process. He then used this model to devise an optimum strategy whereby an operator could decide whether to retrieve a stored text segment or to type it manually. The resulting processing time was calculated. The decision parameters and expected processing time were expressed in terms of the document's length, the expected length of the stored file segments (exponentially distributed lengths were assumed in the examples considered) and the various man-machine parameters including machine display rate (including retrieval), operator's typing rate, proofreading rate, mean editing rate and final copy production rate. Oren [18] conducted a similar type of quantitative analysis for text editing.

Use of language redundancy for spelling correction is implied by research originally motivated by the desire to improve the recognition accuracy of optical character readers [19-21]. The approach involves modelling the English language as an N^{th}-order Markov source, with the result that neighbouring characters can be used to help establish whether or not a specific character is misspelled. Simulation results for $N \leq 2$ show an error rate reduction of between 0.25 and 0.67 the original rate. Further improvement would result from use of a higher-order Markov model. Implementation of the correction algorithm is economically viable for small values of N by means of the Viterbi algorithm [22,23] originally developed for decoding sequences of binary digits transmitted over a noisy communication channel. Further improvement would also result from use of a dictionary; any word not in the dictionary would be considered as misspelled, and would be replaced by the dictionary word "most similar" to the misspelled word. The cost of implementing spelling correction increases with N and with the dictionary size.

Additional work is needed to make word processing systems "convivial" [24] in the sense that they are tools that are easily used by ordinary people. In particular, better terminals are needed for the man-machine interface, and improved software is required to facilitate automatic isolation of text segments, arrangement of these in storage, and updating and retrieving of stored segments. Significant progress in either of these areas requires a better understanding of the human learning, memory, language, cognition, and problem solving processes.

II-3 Banking and Electronic Funds Transfer (EFT)

Money is information. It is of minimal value in itself; it serves only as a medium of exchange and is, in the final analysis a convenience to obviate the need to barter goods and services. Being information, money can be stored electronically.

To secure cash from a bank or credit union it is first necessary to identify oneself and to then establish that sufficient funds are on deposit.

Financial institutions are currently in various stages of automation [25-28]. In most of the industrialized world's banks, the clerk or teller can visually examine a daily computer printout for an account's current status. In some banks the information is available within a few seconds via a keyboard-display connected to a CPU holding the account's current balance as well as other information. Many banks which have not already done so are in the process of installing electronically controlled cash dispensers which provide withdrawals to a fixed amount on a 24-hour basis.

Various retail outlets have installed point-of-sale (POS) systems [29] whereby a customer's account is automatically debited by the amount of his purchase; the retailers inventory is usually adjusted at the same time, and in those instances where a sales commission is appropriate, this is credited to the appropriate account.

It is not difficult to visualize an extension by stages whereby electronic banking and POS systems are merged and developed to the point where cheques become obsolete, credit cards are replaced by one EFT card which identifies the bearer and fits into terminals for automatic recording of the bearer's EFT account number, and cash is reserved for payment of paper boys, parking meters and the small corner grocer.

EFT systems imply lower costs associated with accounting and cash dispensing, elimination of bad cheques and account overdrafts, elimination of the inconvenience of having to remember to pay fixed charges at regular intervals, and execution of financial transactions via remote terminals.

EFT systems have potential disadvantages including the consumer's inability to stop payment for defective goods and services, remote theft from and manipulation of accounts by those with the requisite knowledge, and the monitoring of one's movements and behavioral patterns by examination of his financial transactions. Remedies are available, but these add to the overall system cost.

II-4 Electronic Mail

The essential components of an electronic mail [30-34] (EM) system appear in Fig. 2-1. The sender's input to the electronic mail centre (EMC) would consist of appropriately formatted magnetic tapes on which "mail" has been recorded, or paper mail which would be converted via facsimilie (FAX) scanners or optical character readers (OCR) to electronic format. The actual tapes could be taken physically to the electronic mail centre; alternatively, their contents could be transmitted via data links from the users' premises. The recipient EMC would store the received "mail" electronically for delivery. For recipient users having appropriate facilities, distribution would involve delivery of either the actual magnetic tapes to the recipient's premises or electronic signals to the recipient user's electronic storage facilities. For other recipient users, electronic "mail" would be converted at the recipient EMC to hard copy format for conventional hand delivery.

Electronic transmission of mail is restricted to those documents for which the textual content alone is of interest. In some instances the physical document would have to be delivered; for example when an original signature is required, or when the document is a scented love-letter or a drawing from a favorite nephew or niece.

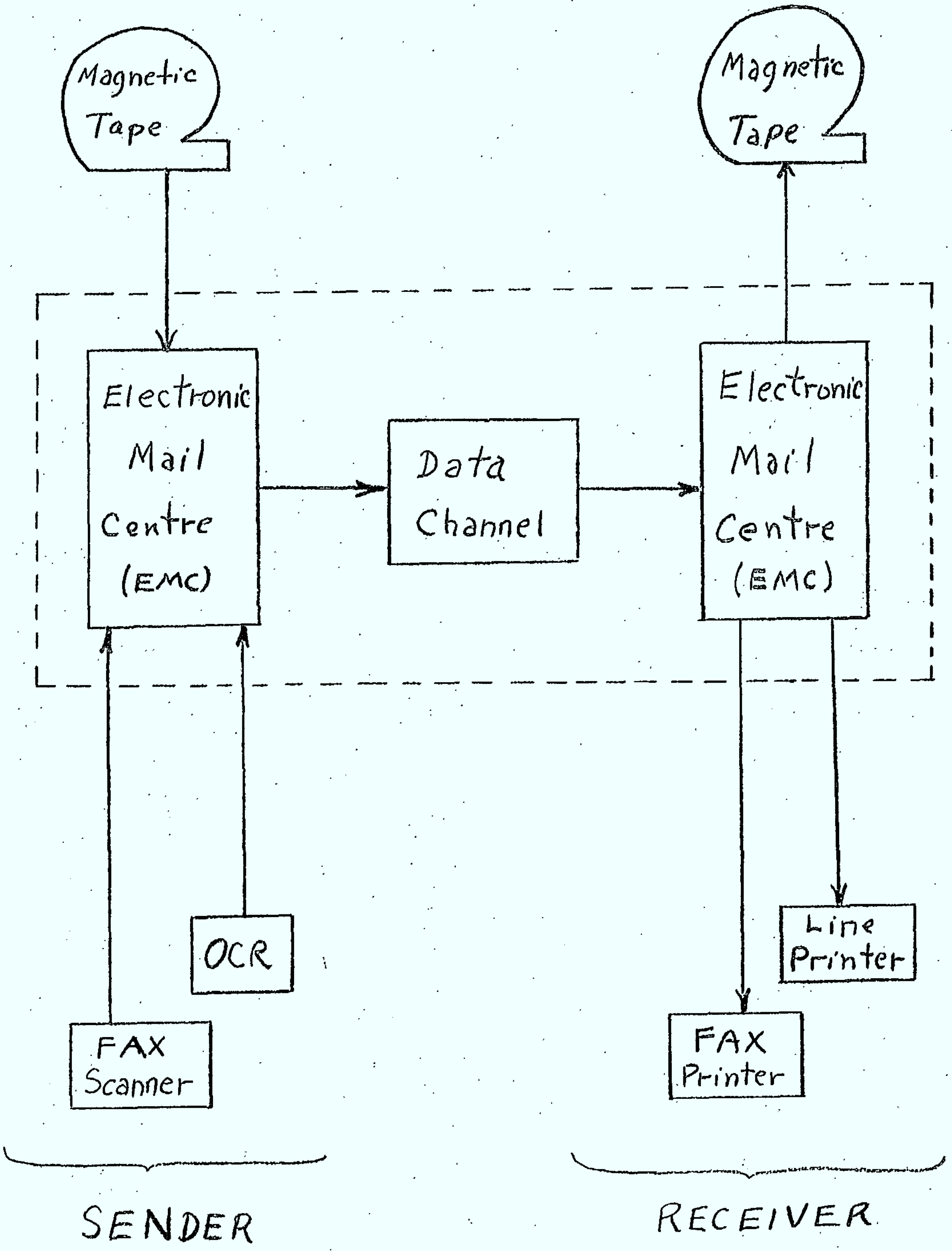

Fig. 2-1: Electronic mail system

Electronic delivery of recorded voices has been proposed [35]. The sender's voice would be converted to digital format, transmitted as "mail", and reconstructed. Acoustic transducers would be required at the sending and receiving end, preferably at the user's premises. A "voice-gram" service would obviate the need for preparing and sending a letter if a called party could not be reached, a situation which occurs, allegedly, with a probability of at least 0.8 on any given attempt [35]. A widely available EM service would tend to make voice-gram service unnecessary.

Fig. 2-2 shows mail probabilities associated with the various sources and recipients, as well as the probabilities of mail from each source to each destination for the USA in 1974 [33]. The largest flow is from businesses to individuals; approximately 46% of the total mail flow follows this route. Current projections show that an EM service in the USA could be required to handle 100 million mail pieces per day, which implies a transmission volume of 6.5×10^{12} bits/day, much of which would require high quality FAX [30,33].

We now give brief consideration to some of the EM issues. Regarding technology, adequate communication facilities will likely be in place by 1985 in the USA [30] as well as in Canada. Communication links envisioned include voice grade lines leased from the carriers, satellite channels, or packet switched data networks. However, processing of paper documents may involve problems. Current scanners handle individual sheets at the rate of one per minute; an EM system would require a capability of between four and ten

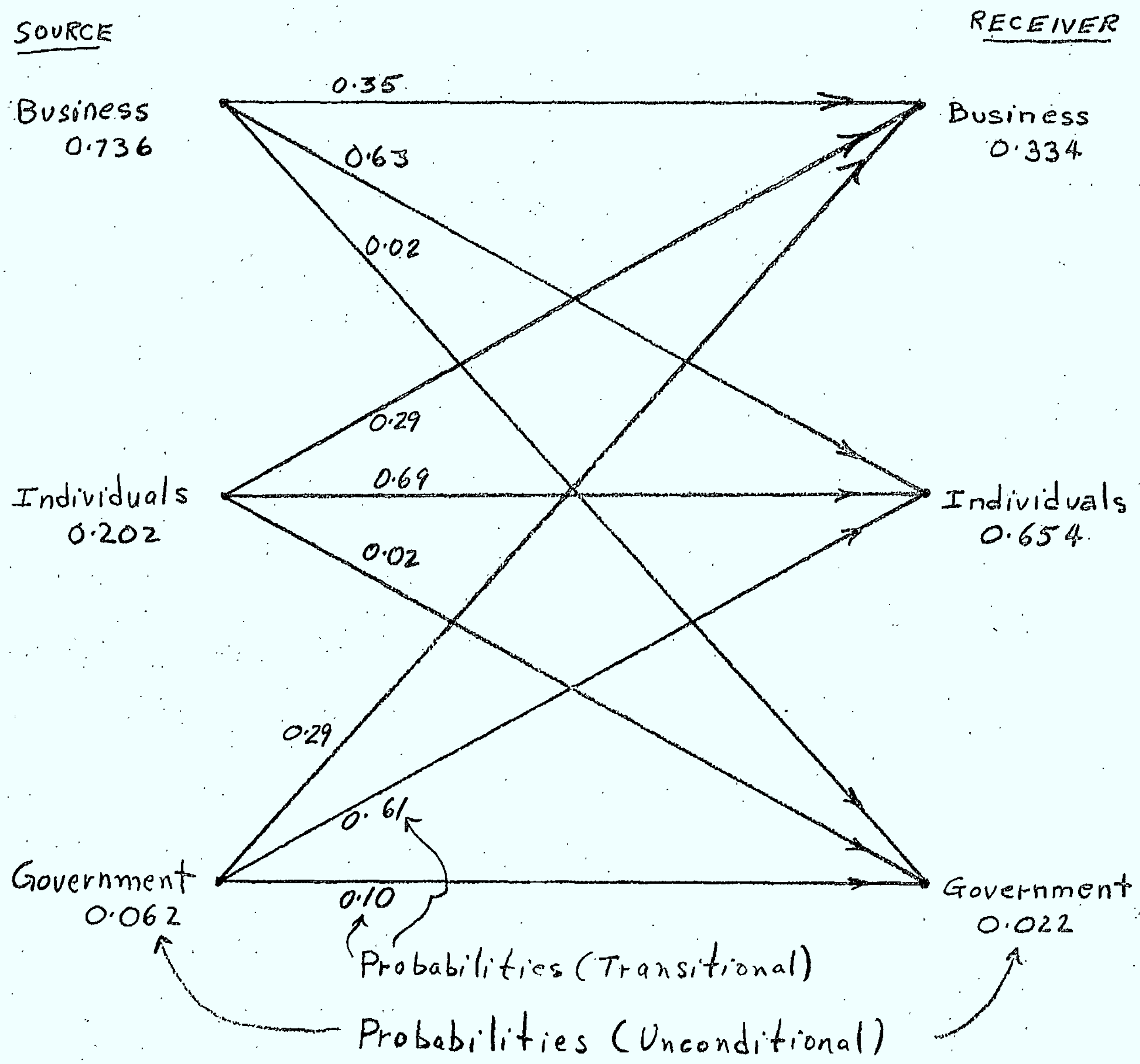

Fig. 2-2: Pattern of mail flow (U.S.A. 1974)

sheets per second. The most serious problem, however, is the reliability of high speed printers [30].

Accurate estimates of the economics of EM are difficult to obtain. Potential economic advantages depend in part on developments in office automation, electronic funds transfer, and other electronic message service applications. The economic viability of these other applications depends, in part, on the costs and availability of EM facilities. The greater the trend toward the use of other electronic information services, the less the need for OCR and FAX facilities, and the less the need for conventional mail services. Reductions in conventional service would encourage the use of electronic services. Determination of actual costs is further compounded by various subsidies, both direct and indirect. If actual costs were used to assess postal tariffs, electronic mail tarrifs might be somewhat different than tarrifs on conventional mail; for example, letters sent from one source to many recipients might all be sent at rates below the single letter rate, since little additional electronic storage is needed for the name and address of the various recipients.

The broader economic and sociological effects are even more difficult to predict. For example, airlines are currently subsidized by the PO for transporting mail. If a large percentage of conventional mail were replaced by electronic mail, airline revenues might drop, fares might then rise, alternative means of transportation which consume less energy would then become more attractive, and fuel costs and pollution levels might drop in response to reduced air traffic demand. Fuel usage in collection and delivery of conventional mail would also drop with a decrease in conventional mail volume.

An electronic mail service would likely be phased into use, in order that a reduction in or retraining of postal workers could proceed smoothly, with little or no increase in unemployment. In any case, post-office automation will not necessarily be met with enthusiastic responses from labour unions.

Finally, we note that there may be some question as to whether or not the post-office has the legal right to be involved in handling electronic "mail". One approach is to permit its involvement only when physical documents require handling at either an electronic mail centre or a conventional postal station. In any event modification of existing statues may be required to guard against electronic mail theft, and "opening" of mail by unauthorized persons.

II-5 Observations and Comments

From the preceeding discussion it is seen that while each application has its own peculiarities and problems, the three are strongly interrelated and involve many common issues. For example, both word processing and EFT involves use of terminals. It is natural to enquire as to how such terminals and their software should be designed and whether or not one type of terminal will serve both applications.

Word processing, EFT and electronic mail would all be more widely accessible if terminals were available in the home. Again, many questions arise. Can standard television receivers be modified for use in these applications? Should the CATV networks be incorporated as part of an information service network? How much, on a monthly basis, would consumers be prepared to pay for such a service?

All three applications require networks linking data storage and processing depots to other such depots and/or terminals. Where should these depots be located? What data should be stored at the various depots? How should the data be routed over these links?

Other questions follow. How does one assess the economic costs and benefits of electronic information systems? How is development to be financed? What are the potential sociological impacts of electronic information systems?

The diversity of disciplines involved in the EIS industry implies a partitioning of the totality of subject matter into relatively disjoint packages for detailed consideration by different groups of experts. How should the subject matter be partitioned, and later integrated?

Perhaps the most important question is the following one: can an EIS system be implemented in such a way that it is inherently impossible for special interest groups to use the system for undesirable activities?

To these larger questions we now turn, realizing that much detailed study and analysis remains to be done on the many specific applications.

II-6 References

1. J. Martin and A. Norman, The Computerized Society. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

2. J. Martin, Design of Man-Computer Dialogues, Englewood Cliffs, N.J.: Prentice-Hall, 1973.

3. J. Vallee, R. Johansen, H. Lipinski, and T. Wilson, "Pragmatics and dynamics of computer conferencing: A summary of findings from the FORUM project," in Conf. Rec., Third Inter. Conf. Comp. Communications, Toronto, Canada, Aug. 1976, pp. 208-213.

4. H.M. Lipinski, and R.H. Miller, "FORUM, a computer-assisted communications, medium," in Conf. Rec., Second Int. Conf. Comp. Communications, Stockholm, Sweden, 1974, pp. 143-147.

5. G.W. Jull, R.W. McCaughern, N.M. Mendenhall, J.R. Storey, A.W. Tassie, and A. Zalatan, "Research Report on Teleconferencing," vol. 1 and vol. 2, CRC Rept. No. 1281-1 and 1281-2, Ottawa, Canada, Jan. 1976.

6. P. Baran, "Broad-band interactive communication services to the home; Part I - potential market demand," IEEE Trans. Commun., vol. COM-23, pp. 5-15.

7. H.A. Peele and E.M. Riseman, "The four faces of HAL: a framework for using artificial intelligence techniques in computer-assisted instruction," IEEE Trans. on Systems, Man and Cybern., vol. SMC-5, pp. 375-380, May 1975.

8. W.L. Haney, W.C. Brown and J. Brahan, "The computer-aided learning program at the National Research Council of Canada," Int. J. Man-Mach. Studies, vol. 5, pp. 271-288, July 1973. (The entire issue of 13 papers is devoted to computer-aided instruction.)

9. K. Chen, "Cable communications policy issues: an overview," IEEE Trans. on Systems, Man and Cybernetics, vol. SMC-6, pp. 727-734, Nov. 1976.

10. T.C.S. Kennedy, "Experience with a mini-computer based hospital administration system," Int. J. Man-Mach. Studies, vol. 5, pp. 237-250, March 1973.

11. T.C.S. Kennedy, "The design of interactive procedures for man-machine communications," Int. J. Man-Mach. Studies, vol. 6, pp. 309-334, May 1974.

12. M.E. Silverstein, S. Rosenberg, and M.A. Cremer, "Medical diagnostic terminals: their nature and description of an archetypal labour saving system," in Conf. Rec., Int. Conf. Comp. Commun., Stockholm, Sweden, August 1974, pp. 53-58.

13. M.L. Rockhoff, "An overview of some technological/health-care system implications of seven exploratory broad-band communication experiments," IEEE Trans. on Commun., vol. COM-23, pp. 20-30, Jan. 1975.

14. M.E. Silverstein, "The wired medical community: an emerging reality," in Conf. Rec., Int. Conf. Commun., Stockholm, Sweden, August 1974, pp. 41-45.

15. R.K. Summit and O. Firschein, "On-line reference searching," IEEE Spectrum, vol. 12, pp. 68-71, Oct. 1975.

16. R. Fajman and J. Borgett, "WYLBUR: An interactive text-editing and remote job entry system," Commun. ACM, vol. 16, pp. 314-322, May 1973.

17. S.S. Oren, "A mathematical theory of man-machine document assembly," IEEE Trans. on Systems, Man, Cybern., vol. SMC-5, pp. 520-527, Sept. 1975.

18. S.S. Oren, "A mathematical theory of man-machine text editing," IEEE Trans. on Systems, Man, Cybern., vol. SMC-4, pp. 258-267, May 1974.

19. D.L. Neuhoff, "The Viterbi algorithm as an aid in text recognition," IEEE Trans. Inform. Theory, vol. IT-21, pp. 222-226, March, 1975.

20. J. Raviv, "Decision making in Markov chains applied to the problem of pattern recognition," IEEE Trans. Inform. Theory, vol. IT-13, pp. 536-551, October 1967.

21. G.T. Toussaint and R.W. Donaldson, "Some simple contextual decoding algorithms applied to the recognition of handprinted text," in Proc. Annual Can. Comp. Conf., Session 72, Montreal, Canada, June 1972, pp. 422101-422116.

22. G.D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," IEEE Trans. Inform. Theory, vol. IT-18, pp. 363-378, May 1972.

23. G.D. Forney, "The Viterbi algorithm," Proc. IEEE, vol. 61, pp. 268-278, March 1973.

24. I. Illich, Tools for Conviviality. New York: Harper and Row, 1973.

25. F.J. Jensen, "Centralization or decentralization in banking," Datamation, vol. 22, p. 60, July 1976.

26. N. Foy and W. Helgason, "Europe claims the lead in banking," Datamation, vol. 22, pp. 57-59, July 1976.

27. B. Sköldborg, "Real time banking system as an application," in Conf. Rec., Int. Conf. Comp. Commun., Stockholm, Sweden, August 1974, pp. 71-82.

28. G.E. Passant, "Operational impact of real time computing in trustee savings banks," in Conf. Rec., Int. Conf. Comp. Commun., Stockholm, Sweden, August 1974, pp. 83-85.

29. P.U. McEnroe, H.T. Huth, E.A. Moon and W.W. Morris III, "Overview of the supermarket system and the retail store system," IBM Systems Journal, vol. 14, pp. 3-15, Jan. 1975.

30. W.J. Miller, "Technology assessment for electronic message handling", in Conf. Rec., IEEE Int. Conf. Commun., Philadelphia, Penn., 1976, pp. (15-10)-(15-12).

31. M.A. Robbins, "Error objective in the electronic mail system," in Conf. Rec., IEEE Int. Conf. Commun., Philadelphia, Penn., 1976, pp. (15-7)-(15-9).

32. C.M. Laucht, "Electronic mail for the Canadian environment," in Conf. Rec., IEEE Int. Conf. Commun., Philadelphia, Penn., 1976, pp. (15-3)-(15-4).

33. G.D. Hodge, "An electronic mail system -- will it happen?" in Conf. Rec., Int. Conf. Comp. Commun., Stockholm, Sweden, August 1974, pp. 351-357.

34. R.L. Williams, "Mailgram -- an electronic mail service," in Conf. Rec., IEEE Int. Conf. Commun., Philadelphia, Penn., 1976, pp. (15-5)-(15-6).

35. C.R. Kraus, "Meeting the publics' communication needs," in Conf. Rec., IEEE Int. Conf. Commun., Seattle, Wash., 1973, pp. (6-1)-(6-6).

III FACTORS AFFECTING THE ANALYSIS, DESIGN AND PERFORMANCE OF ELECTRONIC INFORMATION SYSTEMS

III-1 Isolation of the Major System Factors

Any new and potentially large industry involves a broad spectrum of issues and problems, some of which do not become apparent until the industry matures. Notwithstanding, articulation of the salient aspects at an early date often avoids mistakes which might later prove costly and perhaps irreversible.

We note that the applications in Chapter 2 involve the following topics:

1. The linking of computers and computer peripherals (including terminals and memory modules) via data communication links to form computer-communication networks.
2. Man-machine interfacing, including conversion of messages such as text, speech and visual images into and reconstruction of these from electrical signals.
3. Data-base organization and management, including storage, retrieval, modification and processing of computer data distributed throughout the network.
4. Socio-economic matters, many of which impact on government policy.

Accordingly, we select these as the component aspects affecting systems analysis, design and evaluation, acknowledging that such a subdivision is convenient and somewhat arbitrary and that these four aspects are interrelated.

III-2 Computer-Communication Network Design

Computer-communication implies transmission of information between nodes at which computers and/or terminals are located. Information processing, including storage, retrieval, and modification of data may occur from time to time at various nodes.

Regarding the design of a computer-communication network, the following questions arise:

1. Where geographically should the nodes be located and what computing resources should be allocated to each; ie., how much memory, and if terminals are required what should be their specifications?
2. How should the nodes be interconnected? What should be the capacity of the data lines used? Where should concentrator/multiplexors be located?
3. At which nodes should the various data files be located? (It may be desirable to store some files redundantly at several nodes, and to move files from one node to another in accordance with changing usage patterns.)
4. What network protocols including routing of data, multiplexing, congestion control, and transmission request and acknowledgement schemes should be used to enable data transmission?

The fact that these four questions define the computer communication network design problem was not formally articulated until recently [1]. Consequently, it is not surprising that these questions have not, in general, been answered.

As noted earlier, the network design problem is to be considered in detail in a separate report by the author [2]. At this point, the current status of some aspects of the problem is summarized, as follows:

1. If all user terminals are connected to one node, the network is centralized. Given the terminal traffic and terminal locations, reasonably effective procedures exist for specifying data-link capacities and routes as well as concentrator-multiplexor locations and capacities [3-5]. Network cost is minimized, subject to a constraint on the average time delay due to queueing at the terminals. A recently developed terminal clustering approach appears particularly attractive [3].
2. In distributed computer-communication networks, data files are located at two or more nodes and possibly at one or more terminals. Much remains to be learned regarding the design of such networks. It is not clear how to distribute data files and computing power among the various node locations. Attempts have been made to deal with this problem in specific and tightly constrained situations [6-8], but serious study of the problem has hardly begun.
3. Protocol design is in its infancy [9-11]. Those currently in use are heuristic and only partially understood. Many have not been adequately tested.
4. Point-to-point data transmission itself a well understood and highly developed technology, involving the design of modems, and channel encoders and decoders [12-16]. Further limited gains and economics resulting from improved modems and coders will undoubtedly reward hard work. Recent developments are the subject of two current publications [13,14] describing studies demonstrating the feasibility of transmitting data over good quality voice-

grade lines at rates of 12000 and 14400 bits/sec. A few years ago a rate of 9600 bits/sec was regarded as an upper limit.

III-3 Man-Machine Interfacing

Man-machine interfacing involves two related problems:

1. Conversion of text, speech and images into electronic form either for "understanding" by a machine (the pattern recognition or scene analysis problem) or for reconstruction at some later time (the source coding problem).
2. Design of terminals which enable users to interact with stored information as well as with other users.

In source coding the objective is to represent text, speech and images using as few bits of information as possible. As noted in Chapters 5, 6 and 7, it is reasonably clear what can and what cannot be achieved. Pattern recognition problems of interest and considered in this report include recognition of spoken words and printed characters, and speaker verification.

As noted earlier, Chapter 8 provides an introduction to the design and evaluation of user terminals. The need for further study is clearly indicated.

III-4 Data-Base Organization and Management

The 1970's are becoming the decade of the data base [17-19]. Data-base organization and management, which involves specification of data structures as well as procedures for assessing, retrieving, updating and storing data is now the subject of serious scientific and engineering study. Design goals include flexibility, application independence and ease of maintenance. Flexibility includes the ability of a data base to provide quick answers to unanticipated questions.

The two basic and different data-base organizations are the centralized heirarchical structure, and the newer relational structure proposed by Codd [19]. It is not clear which structure is best. Neither is it clear how to organize and manage distributed data bases. As indicated earlier, this latter problem has received but superficial consideration [6-8].

Now is the time to study seriously the design of distributed data bases linked by actual data-communication channels. These are subject to (variable) queueing and transmission delays as well as transmission errors. To what extent should organization and management of distributed and centralized data bases be similar? We don't really know.

Data-base technology is not specifically considered in this report; however, our discussions in Chapter 8 on user terminals indicate the relationships between data-base technology and man-machine interfacing.

III-5 Socio-Economic Issues

As noted earlier, economic matters are considered in some detail in Chapter 4 under three subheadings: cost of supplying electronic information services, demand for services, and government economic options.

There is a clear relationship between supply costs and network design; specifically the latter involves the arrangement of data links, concentrator-multiplexors, computers, and computer peripherals to provide a given grade of service at minimum cost.

There is also a clear relationship between the cost of transmitting a message and the number of bits needed to encode the message.

The sociological issues are not given detailed consideration in this report, since they have been discussed at length elsewhere [20-25]. These issues raise questions over which reasonable people would disagree. For example, how much of an increase in system cost is warranted to increase the security and privacy of an individual's data?

A recent article [20] contains a summary of many of the important sociological considerations.

III-6 Performance Evaluation

Performance assessments of EIS systems will depend on the values and value judgements of different users, and these differ among individuals.

This dilemna we deal with as follows: If system configuration A provides the same services with equal user satisfaction as system B but at less cost, then system A is better than system B by the amount of the cost difference. Determination of equality of user satisfaction is not always easy but reasonable estimates of equality are often obtainable.

III-7 References

1. S.K. Chang, "A model for distributed computer system design," IEEE Trans. Systems, Man and Cybern., vol. SMC-5, pp. 344-359, May 1976.

2. R.W. Donaldson, "Design of Distributed Computer-Communication Networks", Report to the Department of Communications, Ottawa, Canada, March 1977, in preparation.

3. H. Dirilten and R.W. Donaldson, "Topological design of teleprocessing networks using linear regression clustering," IEEE Trans. Commun., vol. COM-24, pp. 1152-1159, Oct. 1976.

4. H. Frank and W. Chou, "Topological optimization of computer networks," Proc. IEEE, vol. 60, pp. 1385-1397, Nov. 1972.

5. R.R. Boorston and H. Frank, "Large-scale network topological optimization," IEEE Trans. Commun., vol. COM-25, pp. 29-47, Jan. 1977.

6. W.W. Chu, "Optimal file allocation in multiple computer system," in Computer-Communication Networks. N.A. Abramson and F.F. Kuo, Ed., Englewood Cliffs, N.J.: Prentice-Hall, 1973, Ch. 3.

7. D.J. Foley and E.H. Brownlee, "A model of distributed processing in computer networks with application to satellite graphics," in Conf. Rec., Second Int. Conf. Computer Commun., Stockholm, Sweden, Aug. 1974, pp. 331-335.

8. S.P. Ghosh, "Distributing a data base with logical associations on a computer network for parallel sharing," IEEE Trans Software Engr. vol. SE-2, pp. 106-113, June 1976.

9. L. Kleinrock, Queueing Systems, Vol. 2: Computer Applications. New York: Wiley, 1976, Ch. 6.

10. H. Rudin, "Chairman's remarks: An introduction to flow control," in Conf. Rec., Third Annual Conf. Comp. Communication, Toronto, Canada, Aug. 1976, pp. 463-466.

11. R.G. Gallager, "Basic limits on protocol information in data communication networks," IEEE Trans. Inform. Theory, vol. IT-22, pp. 385-398, July 1976.

12. R.W. Lucky, J. Salz and E.J. Weldon, Principles of Data Communication. New York: McGraw Hill, 1968.

13. D. Falconer and F.R. Magee, Jr. "Evaluation of decision feedback equalization and Viterbi algorithm detection for voiceband data transmission," IEEE Trans. Commun., vol. COM-24, pp. 1238-1245, Nov. 1976.

14. F.R. Magee, Jr. "A comparison of compromise Viterbi algorithm and standard equalization techniques over band limited channels," IEEE Trans. Commun., vol. COM-23, pp. 361-367, Mar. 1975.

15. W.C. Lindsay and M.K. Simon, Telecommunication Systems Engineering. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

16. S. Lin, An Introduction to Error-Correcting Codes. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

17. J. Martin, Principles of Data-Base Management. Englewood Cliffs, N.J.: Prentice-Hall, 1976.

18. J. Martin, Computer Data Base Organization. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

19. E.F. Codd, "A relational model of data for large shared data banks," Comm. ACM, vol. 13, pp. 377-387, June 1970.

20. ACM Committee on Computers and Public Policy, D.D. McCracken, Chairman, "A problem-list of issues concerning computers and public policy," Comm. ACM, vol. 17, pp. 495-503, Sept. 1974.

21. R.E. Butler, "International co-operation and regulation; foundations for development," in Conf. Rec., Second Inter. Conf. Computer-Communication, Stockholm, Sweden, Aug. 1974, pp. 11-17.

22. D.F. Parkhill, "Society and communications policy," Conf. Rec., Third Inter. Conf. Comp. Commun., Philadelphia, Penn., Aug. 1976, pp. 11-17.

23. J. Martin and A. Norman, The Computerized Society. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

24. T.D. Sterling, "Guidelines for humanizing computerized information systems: A report from Stanley House," Commun. ACM, vol. 17, pp. 609-613, Nov. 1974.

25. H. Von Baeyer, "The quest for public policies in computer/communications - Canadian approaches," in Conf. Rec., Second Inter. Conf. Comp. Commun., Stockholm, Sweden, Aug. 1974, pp. 19-23.

IV. ECONOMIC ASPECTS OF ELECTRONIC INFORMATION SERVICES

IV-1 Introduction

The costs of goods and services, as well as the amounts consumed, are determined by supply and demand [1]. The above statement is true even though each may be controlled somewhat artificially; for example, supply by cartels and demand by coercive advertising. Fig. 4-1 shows typical supply and demand curves for a commodity or service. The point of their intersection determines commodity price per unit as well as the number of units produced and consumed.

Supply and demand curves for a given commodity depend on many factors. Demand for a particular item is affected in varying degrees by the per-unit costs of other commodities, particularly those which compete with the commodity in question. Supply costs depend on costs of material, labour, land, and capital.

Government activity affects supply and demand. Taxation and regulatory laws affect producer costs, while direct and indirect subsidies to consumers raise the per-unit price that a consumer will pay for a given quantity. Fiscal and monetary policies affect supply and demand through control of interest rates and employment levels.

The foregoing comments motivate us to consider economic factors affecting the EIS industry under three separate but related headings:

1. cost of supplying electronic information services
2. demand for electronic information services
3. economic options for governments

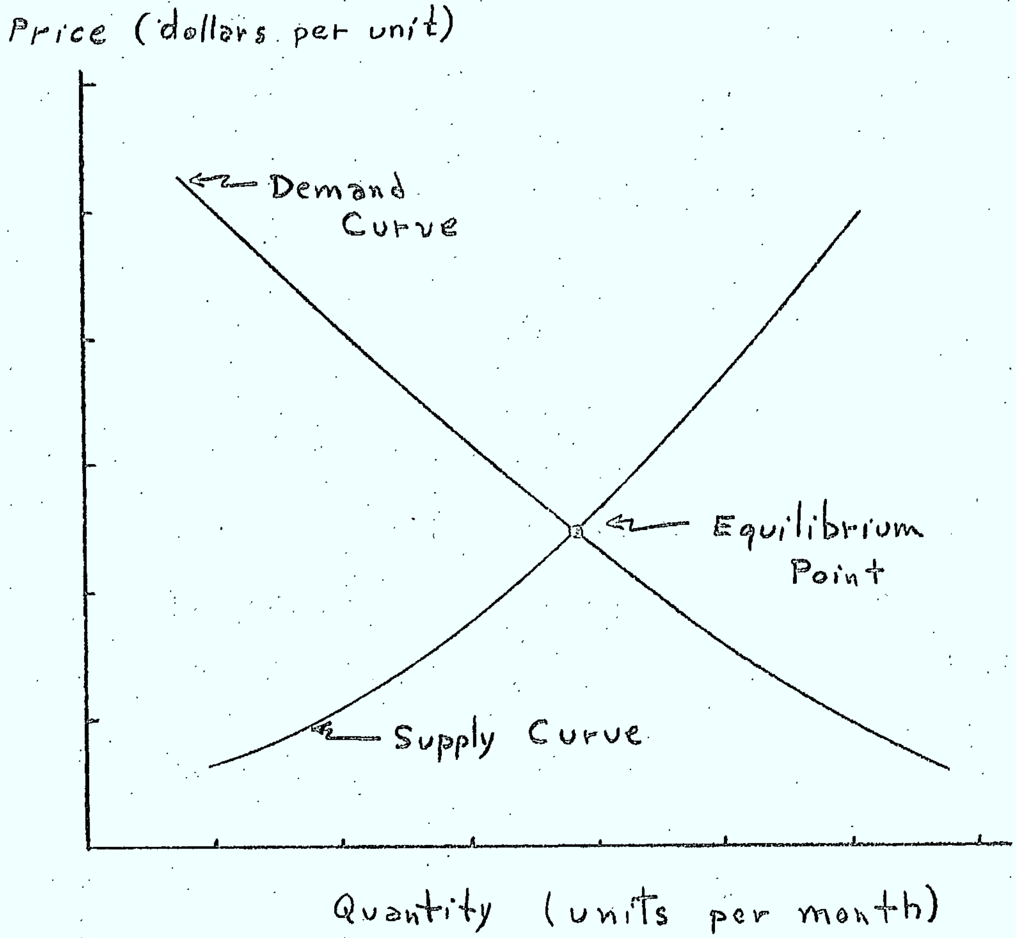

Fig. 4-1: Typical supply and demand curves

The eventual size and relatively long time horizon for the EIS industry motivates assessment of future cost trends as well as current costs.

IV-2 Costs of Supplying Electronic Information Services

The cost of supplying a service is the product of the per-unit cost of the service and the number of units required to meet user need.

Per-unit costs depend, to a considerable extent, on the total number of units of service provided to the user population (see Fig. 4-1). Per-unit costs would normally decrease as the number of units of service increases; the latter increases as the number of different but related services increases. For example, the costs of supplying a single company with a word processing system might be prohibitive. However, if an electronic mail service and an electronic funds transfer service were also available, and if many companies were to share these services, the per-unit costs of each to any one company might well be acceptable.

The number of units of service required to meet a user's need depends, in large measure, on the number of bits required to encode a message. As stated earlier, reasonably complete knowledge of what economics can and cannot be achieved is available as summarized in Chapters 5, 6 and 7.

The component costs for EIS systems are as follows:

1. cost of data transmission hardware
2. cost of computer hardware, including memory
3. cost of user terminals, including software
4. cost of system software

Both construction and rental costs of common carrier data transmission facilities show stepwise increases with capacity [2,3]. A unit increment in capacity yields a less than unit increment in cost. Construction costs increase linearly with distance, whereas rental costs increase more slowly.

Terrestrial communication costs have been decreasing since 1960, mainly because improved modem designs have increased rates at which data can be reliably transmitted [4-6]. Roberts [7] estimates an annual cost decrease in the USA of 11 per cent, which implies a factor of 10 reduction in a 22-year period. The breakthrough which might result in a drastic decrease in terrestrial communications costs is the realization of fibre optic channels [8].

Satellite channel costs have decreased much more rapidly over the past decade. Roberts [7] estimates a US decrease of 40% per annum, which implies a 10-fold cost reduction in 6.7 years. These estimates are based on limited data for a new technology. It seems unlikely that this rate of cost decrease would continue.

In addition to common carrier terrestrial channels and satellite channels, one might include cable TV channels in EIS networks. These exist as broadband channels suited, following some modification, for transmission of messages to and from the home [9-13]. Cable system modifications are envisioned whereby the wideband video and audio signal into the home is complemented by a narrowband enquiry data channel from the home [9-11]. The economic, legal and technological issues involved in including cable TV channels as part of an electronic information system warrant further study.

Regarding computer hardware costs, including random access storage costs, the trend has been a factor of ten decrease in cost over a five-year period [7]. A continuation of such a trend for the next several years seems likely on the basis of improved CPU architecture, including multiprocessing capabilities, and decreased random access storage costs [3].

It is not easy to forecast CPU developments in terms of size and architecture. Prior to 1968, large systems were more efficient than small ones in terms of cost per processing operation [3]. However, the newer small machines may help reverse this trend. The issue of size profile will depend in part on whether studies on network design and data-base organization and management indicate a preference for many smaller computers operating in distributive fashion or a more centralized system with fewer, larger machines.

Memory costs vary inversely with the time needed to retrieve and store data. Thus, magnetic tape systems, which are slowest in terms of access time, are least expensive in terms of storage cost per data bit; discs are next; and core and semiconductor memory which is fastest (access time approximately 0.3 μsec) is most expensive.

Core and disc costs seem unlikely to decrease much [3]. Magnetic tape costs may decrease if signal design and detection techniques used for data transmission over bandlimited channels can be modified to increase the density with which data can be packed on magnetic tape [14]. Increasing the packing density would decrease the access time. Improvements in semiconductor memories may reduce storage costs further. Drastic reductions are not expected; however, forecasting here is difficult [15]. Alternative storage techniques including optical, magnetic bubble and thin film methods offer some likelihood of further reductions in per-bit storage costs at fast access rates [16-18]. Again, accurate cost estimates and dates of market arrival are difficult to predict.

Regarding user terminals, there exist a wide variety of types including CRT-keyboard, OCR, keyboard printer and special purpose keyboards including touchtone units. Some decrease in terminal costs can be expected as a result of an increase in the number and variety available, and because of reduced memory and visual display costs. Terminal software will become more powerful but is not likely to decrease in cost. A more detailed discussion of terminals is included in Chapter 8.

System software has become the major cost factor and potential limitation to anticipated developments in computer-communication systems [3,19]. This trend is not likely to be reversed quickly, although decreasing hardware costs will allow the transfer of some software functions to hardware. Further, as computer-communication networks develop, software used at one computer site will be available to distant users who would otherwise have to develop their own. In fact, sharing of software and other resources is one of the primary motivations for development of computer-communication networks.

Software cost increases occur because programs, whose costs increase faster than their size, continue to grow in size and complexity. A major cost component results from communication among programmers, many of whom work as artisans rather than as disciplined professionals adhering to well defined and well established formats.

The need for a disciplined approach to software design has been recognized. The IEEE now publishes a transactions entitled Software Engineering. Standardization of data base organization and management [20-22] is rapidly gaining acceptance as is standardization of software protocols in computer-communication systems.

The following statements are offered as an aid to estimate software costs [3]:

1. The programming cost per phrase remains constant over time.
2. The cost per phrase increases with program size; if a 1000 phrase program costs $5.00 per phrase a 10,000 phrase program might cost $10.00 per phrase.
3. The cost of writing a program for a specific task decreases at a rate of 25% per annum.

The improvement of software techniques and user languages will undoubedly occur. A better understanding of human language structures, memory processes, learning, cognition and problem solving is needed. In comparing new data-base organization and management schemes [20-22] one observes a growing similarity with models of human memory, learning and cognition [23].

In principle, then, a system's per-unit cost is determined by adding together the per-unit costs of its component parts. The system configuration having the lowest per-unit cost while meeting specified constraints is optimum. Practical difficulties include obtaining cost data, particularly for future costs and estimation of present values of capital assets. Analytical minimization of system cost is usually impossible, with the result that heuristic and simulation approaches are usually employed [24]. As stated earlier, the paramount design objective is to avoid systems which are highly non-optimum.

The actual time of entry of EIS system suppliers into the marketplace is uncertain. There are large development costs, most of which are borne by those first to offer services. Initial revenues which would finance expansions in service offerings may be forced downwards by competitive suppliers who did not participate in the expensive development phase [25].

Reliable cost data for specific services is scarce. One preliminary analysis indicates a cost for electronic mail of approximately $1.00 per "letter" in 1976 decreasing to $0.25 by 1985 [26].

IV-3 Demand for Electronic Information Services

A product's potential demand may be estimated either by asking people what they would be willing to pay or by extrapolations based on past usage patterns. The weakness inherent in the first approach is that people may find demand assessment difficult, particularly when the product is a service with which they have had virtually no experience. The weakness of the second approach is that past purchasing patterns do not fully expose the total demand curve, but only its equilibrium point (See Fig. 4-1), nor do they fully expose the dependence of the demand curve on per-unit costs of other goods and services.

Data reflecting potential demand for electronic information services is scarce. Much of what exists is proprietary.

Baran [9] reports the results of a 1971 assessment of potential demand for broadband interactive services in the home. The services considered are listed below, without the amplifying descriptions given to the test respondents:

1. cashless-society transactions
2. dedicated newspaper
3. computer-aided school instruction

4. shopping transactions (store catalogue)
5. person-to-person (paid work at home)
6. plays and movies from a video library
7. computer tutor
8. message recording
9. secretarial assistance
10. household mail and messages
11. mass mail and direct advertising mail
12. answering services
13. grocery price list, information and ordering
14. access to company files
15. fares and ticket reservation
16. past and forthcoming events
17. correspondence school
18. daily calendar and reminder about appointments
19. computer-assisted meetings
20. newspaper, electronic, general
21. adult evening courses on television
22. banking services
23. legal information
24. special sales information
25. consumers' advisory service
26. weather bureau
27. bus, train and air scheduling
28. restaurants
29. library access
30. index, all services served by the home terminal

Two detailed and "rather painful [9]" questionnaires were completed by 200 experts in the area of future systems. The respondents were asked for various estimates including the most likely year of introduction and the average value of each service. (The services are clearly not mutually exclusive.) Following completion of the first questionniare, the respondents were given the responses of the other participants and then asked to modify their initial responses (the Delphi technique). No mention was made to the respondents regarding the communication medium to be used to provide the service.

The totality of services was deemed worth twenty 1970 U.S. dollars by the respondents. The largest share of the market was given to computer-aided instruction (18%), person-to-person communication (15%) and plays and movies from a library (10%). The smallest market share was for fee and ticket reservations (0.3%), bus, train and air scheduling (0.3%) and restaurants (0.3%). With 90 million households and 10% penetration by 1980, the projected income totalled 2.2 billion dollars per year.

An analysis [3] which extrapolates the growth in data sets in the US Bell system indicates a faster rate of growth for data than for voice. The projected annual growth rate of 38% for data sets would result in 12% of the total sets on switch network, wideband network and leased network plant being data sets by 1980, vs 0.1% in 1970. The same survey projects wideband cable to 35 million US households by 1980.

Reliable estimates of future demand for EIS services probably requires a fairly extensive description of services offered followed by cost/benefit analysis by prospective users prior to their responding to market surveys.

IV-4 Economic Options for Governments

By means of legislation, regulatory actions, and direct and indirect subsidies including tax incentives, governments can influence both the supply of and the demand for electronic information services.

It is not our intention here to seriously consider the extent to which government should be involved in the EIS industry. The matter is one over which respected economists and other knowledgeable persons would disagree. Nobel Laureate (1976) Milton Friedman argues for less government involvement in the marketplace; John Kenneth Galbraith argues for more [27]. Ours is the pragmatic view that governments will continue to be involved to some degree in the EIS industry. Our efforts are therefore confined to a brief consideration of some government options. The subject is too large for thorough consideration here.

Government actions can strongly influence the direction and rate of growth of the industry. For example, an electronic mail system could be encouraged directly by accelerating its development in the post office or indirectly by encouraging the common carriers or others to provide the service, while increasing rates and reducing service for conventional mail. As explained earlier, implementation of an electronic mail service would increase the utility of word processing systems, electronic funds transfer facilities and other services. Contrary government actions would have contrary effects.

Of major importance is the issue of competition vs monopoly. It is generally agreed that large enterprises can benefit from economics of scale, provided the enterprise is sufficiently specialized to avoid diseconomics resulting from its inability to respond to new circumstances, including demand shifts and new technology. The difficulty is to define the boundary between economy and diseconomy [28-30].

In western democracies citizens are protected against monopolistic injustices by government laws, and by regulatory procedures whereby the industry is permitted a fair rate of return in exchange for an agreed upon product or service. The difficulties lie in determining what constitutes a fair rate of return, whether or not an organization's claims to costs are reasonable, and whether or not products and services provided meet previously agreed upon standards.

The ways in which various industry groups might serve the EIS industry is also related to the monopoly vs competition issue. This question has been resolved, at least for now. In Canada, EIS communication facilities are to be provided solely by the common carriers, under the following conditions [31]:

1. total ban on any carrier involvement with content
2. an obligation of the carriers to meet any reasonable demand for service
3. a legal requirement on the part of the carriers to distribute the services of all suppliers on a non-discriminatory basis at authorized tariffs.

In the USA, private organizations are permitted to apply for licences from the Federal Communications Commission for purposes of providing message service communications facilities; one of these organizations (Datran) has already achieved bankruptcy. The non-carrier segments of the industry in both Canada and the US are to be served by private enterprise in the normal way.

Of considerable interest is the potential effects of electronic information services on activity in other economic sectors. One of the few such studies available considered replacing a centralized downtown office by regional offices staffed by local workers [32]. The study was motivated by the fact that many workers face daily the unpleasant and time-consuming task of commuting to a central location from the suburbs even when the job itself involves information processing rather than frequent face-to-face interactions with other people.

In a case study involving a Los Angeles insurance company, annual telecommuting costs using existing equipment and voice grade lines ranged from $300 to $1000 per employee. The reduction in annual commuting costs ranged from $650 to $1600 with additional savings realized by selling the car once used to commute. The energy used to operate a telecommuting system was estimated to equal 4% of that used for actual commuting. In fact, this estimate is probably conservative, since it did not include energy savings to others commuting in a potentially more efficient manner because of reduced traffic volume. Wider ranging effects of telecommuting, including overall changes in land usage due to larger suburbs and reduced pollution were not considered in the study.*

The employees' incentives to telecommute include more free time and reduced commuting costs. Incentives to the company (who must finance the system) include reduced costs of employee parking and hot lunches; elimination of salary differentials to city workers; possibly reduced office rental costs; more and better qualified applicants for available positions due to the reluctance of many people including working mothers, to work in the city; higher productivity and better customer service. Because approximately four percent of the total energy usage in the USA is for commuting, the

---

*See [33-35] for informative recent reports on telecommuting vs transportation. Of particular relevance are the comments [34] regarding potential shortages of captial for new facilities and the effects of diverting capital from other industries, particularly energy.

energy savings potentially attributable to electronic information services is clearly significant, particularly in view of the possibility of extending telecommuting facilities to reduce intercity travel.

An understanding of the likely changes through which our conventional information services will pass as a result of the growth of EIS services is of interest. Conventional services involve broadcasting, where information flows one way from a source to one or more users, and interactive usage, where information flows to and from a user in a conversational mode. Broadcasting includes AM and FM radio, over-the-air television, conventional cable television, newspapers, magazines and books. Interactive services include telephone and telex. Mail and most of the new information services listed in Chapter 2 probably fall into both categories.

Telephony's growth will probably continue unabated. Telex would undoubtedly be replaced by new electronic services, as would a portion of the conventional paper broadcast media. As indicated earlier, future roles for cable television requires clarification.

In summary, the following matters warrant further study:

1. How and to what extent should the government encourage development of electronic information services?
2. What are the potential effects of various EIS applications on other segments of the economy?
3. How will conventional information services change with the growth of EIS services? In particular, what are the technological, economic and legal aspects of including as part of an EIS network broadband cables currently used for cable television?

Problems associated with monopoly and regulation exist but are not peculiar to the EIS industry.

The three items above reinforce the need for reasonably accurate estimates for costs of and demand for specific electronic information services.

IV-5 References

1. P.A. Samuelson and A. Scott, Economics, 3rd Canadian ed., New York: McGraw-Hill, 1971.

2. H. Dirilten and R.W. Donaldson, "Topological design of teleprocessing networks using linear regression clustering," IEEE Trans. on Commun., vol. COM-24, pp. 1152-1159, Oct. 1976.

3. D.A. Dunn and A.J. Lipinski, "Economic considerations in computer-communication systems," in Computer-Communication Networks, N. Abramson and F.F. Kuo, Ed., Englewood Cliffs, N.J.: Prentice-Hall, 1973, Ch. 10.

4. R.W. Lucky, J. Salz, and E.J. Weldon, Principles of Data Communication. New York: McGraw-Hill, 1968.

5. D.D. Falconer and F.R. Magee, Jr., "Evaluation of decision feedback equalization and Viterbi algorithm detection for voice-band data transmission," IEEE Trans. on Commun., vol. COM-24, pp. 1238-1245, Nov. 1976.

6. F.R. Magee, Jr., "A comparison of compromise Viterbi algorithm and standard equalization techniques over band-limited channels," IEEE Trans. Commun., vol. COM-23, pp. 361-367, March 1975.

7. L.G. Roberts, "Data by the packet," IEEE Spectrum, vol. 11, pp. 46-51, Feb. 1974.

8. K. Chen, "Cable communications policy issues: an overview," IEEE Trans. Systems, Man, and Cybern., vol. SMC-6, pp. 727-734, Nov. 1976.

9. P. Baran, "Broad-band interactive communication services to the home: Part I - potential market demand, IEEE Trans. Commun., vol. COM-23, pp. 5-15, Jan. 1975.

10. A.G. Hare and A.H. Ithell, "Multipurpose wide-band local distribution - proposals for an integrated system," IEEE Trans. Commun., vol. COM-23, pp. 42-48, Jan. 1975.

11. K. Rose and R.K. Stevens, "Design of a switched broad-band communications network for interactive services," IEEE Trans. Commun., vol. COM-23, pp. 49-55, Jan. 1975.

12. I.T. Frisch, "Technical problems in nationwide networking and interconnection," IEEE Trans. Commun., vol. COM-23, pp. 78-88, Jan. 1975.

13. R.E. Babe, Cable Television and Telecommunications in Canada, An Economic Analysis, East Lansing, Mich.: MSU International Business and Economic Studies, 1975.

14. J.C. Mallinson, "Tutorial review of magnetic recording," Proc. IEEE, vol. 64, pp. 126-208, Feb. 1976.

15. D. Hodges, "A review and projection of semiconductor components for digital storage," Proc. IEEE, vol. 63, pp. 1136-1159, Aug. 1975.

16. D. Chen and J.D. Zook, "An overview of optical data storage technology," Proc. IEEE, vol. 63, pp. 1207-1230, Aug. 1975.

17. H. Bobeck, P.I. Bonyhard and J.E. Geusic, "Magnetic bubbles - an emerging new memory technology," Proc. IEEE, vol. 63, pp. 1176-1195, Aug. 1975.

18. W.C. Hughes, "A semiconductor nonvolatile electron beam accessed mass memory," Proc. IEEE, vol. 63, pp. 1230-1240, Aug. 1975.

19. R.E. Butler, "International cooperation and regulation foundations for development," in Conf. Rec., Second Conf. Computer Commun., Stockholm, Sweden, Aug. 1974, pp. 11-17.

20. J. Martin, Computer Data-Base Organization. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

21. J. Martin, Principles of Data-Base Management. Englewood Cliffs, N.J.: Prentice-Hall, 1976.

22. E.F. Codd, "A relational model of data for large shared data banks," Commun. ACM, vol. 6, pp. 377-387, June 1970.

23. D.H. Lindsay and D.A. Norman, Human Information Processing. New York: Academic Press, 1972.

24. R.W. Donaldson, "Design of Distributed Computer-Communication Networks," Report to the Department of Communications, Ottawa, Canada, March 1977, in preparation.

25. P. Baran, "Broadband interactive systems to the home: Part II - impasse," IEEE Trans. Comm., vol. COM-23, pp. 166-170, Jan. 1975.

26. R.R. Panko, "The outlook for computer message services: a preliminary assessment," Preliminary Report, Stanford Research Institute, Menlo Park, Calif., March 1976.

27. J.K. Galbraith, Economics and the Public Purpose. Scarborough, Ont.: New American Library of Canada, Ltd. 1975.

28. A.D. Hall, "An overview of economics of scale in existing communication systems," IEEE Trans. on Systems, Man, and Cybern., vol. SMC-5, pp. 2-14, Jan. 1975.

29. W.H. Melody, "Relations between public policy issues and economics of scale," IEEE Trans. on Systems, Man, and Cybern., vol. SMC-5, pp. 14-22, Jan. 1975.

30. J.M. Taplin and M.P. Beere, "The economics of new information networks," IEEE Trans. on Systems, Man, and Cybern., vol. SMC-5, pp. 40-43, Jan. 1975.

31. D.F. Parkhill, "Society and computer communication policy," in Conf. Rec., Third Inter. Conf. Computer Commun., Toronto, Canada, Aug. 1976, pp. 11-17.

32. J.M. Nilles, F.R. Carlson, P. Gray and G. Hanneman, "Telecommuting - an alternative to urban transportation congestion," IEEE Trans. Systems, Man, and Cybern., vol. SMC-6, pp. 77-85, Feb. 1976.

33. R.C. Harkness, "Selected results from a technology assessment of telecommunication-transportation interactions," in Conf. Rec., Int. Conf. Commun., Philadelphia, Penn., June 1976, pp. (26-1)-(26-7).

34. R.R. Panko, R.W. Hough, and R. Pye, "Telecommunications for office decentralization: apparent needs and investment requirements," in Conf. Rec., Int. Conf. Commun., Philadelphia, Penn., June 1976, pp. (26-8)-(26-13).

35. R. Pye and P.I. Weintraub, "Social and organizational implications of decentralization," in Conf. Rec., Int. Conf. Commun., Philadelphia, Penn., June 1976, pp. (26-14)-(26-19).

V ANALYSIS, CODING, RECOGNITION, SYNTHESIS AND RECONSTRUCTION OF SPEECH; SPEAKER VERIFICATION

V-1 Speech and Electronic Information Services

Coding, recognition, synthesis and reconstruction of speech, as well as speaker verification are of importance to the electronic information service industry for two reasons.

First, speech is one of the available system inputs and outputs. For example, responses to enquires from terminals, including mobile terminals [1,2] can be via voice-answerback [3,4]. The various ways to convert acoustic utterances to and from electrical signals and to process these is, therefore, of interest.

Second, electronic information service costs depend to a considerable degree on the number of bits needed to represent a message. The lower the required bit-rate for speech, the lower the cost of its storage and transmission.

In conventional telephony speech is prefiltered to pass only those frequencies between $f_1 \cong 200$ Hz and $f_2 \cong 3.5$ KHz. If digital transmission ensues, the prefiltered signal is sampled at the Nyquist rate, quantized non-uniformly to 128 levels, transmitted digitally, and reconstructed by lowpass filtering at $f_2 \cong$ 3.5 KHz.

Recent advances in speech coding, reconstruction and synthesis include considerable savings in bit rate for acceptable increases in coding complexity and cost. It is reasonable to expect that some EIS users would favour the option of speech input and output whose quality was in accordance with user cost. Potential cost vs quality trade-offs are becoming apparent as explained below.

V-2 Conventional Digital Coding and Reconstruction of Speech

Fig. 5-1 shows a conventional PCM system. Fig. 5-2 defines the quantizer characteristic in Fig. 5-1. The source coder, which consists of a prefilter, sampler and quantizer converts the analog speech signal a(t) into binary digits which are either transmitted over a digital communication channel or stored. The source decoder converts the received digits to analog pulses which are then postfiltered to yield the reconstructed signal m(t).

A commonly used measure of speech quality is the output signal-to-noise ratio SNR, defined as follows; where m(t) and $\hat{m}(t)$ are defined in Fig. 5-1, and E [] denotes expected value:

$$SNR = E\left([m(t)-\hat{m}(t)]^2\right) / E\left(m^2(t)\right) \qquad (5\text{-}1)$$

From analysis [5-10] it follows that SNR for a well designed system can be approximated as follows, when L is the number of quantizer levels, r/2W is the ratio of the sampling rate to the prefilter's Nyquist bandwidth 2W (in PCM r=2W), and p is the bit-error probability of the digital channel which is assumed memoryless; constants A and B, which are close to unity, depend on the statistics of m(t), on the actual quantizer characteristic and on the way in which the quantizer output digits are coded:

$$SNR \approx \left[\frac{A}{L^2} + 4Bp\right]\frac{r}{2W} \qquad (5\text{-}2)$$

The term A/L^2 corresponds to quantization noise, while the term 4Bp results from digital channel transmission errors.

For a waveform channel with a given carrier-to-noise ratio, p decreases as the bit rate $R=r\log_2 L$ increases. For example,

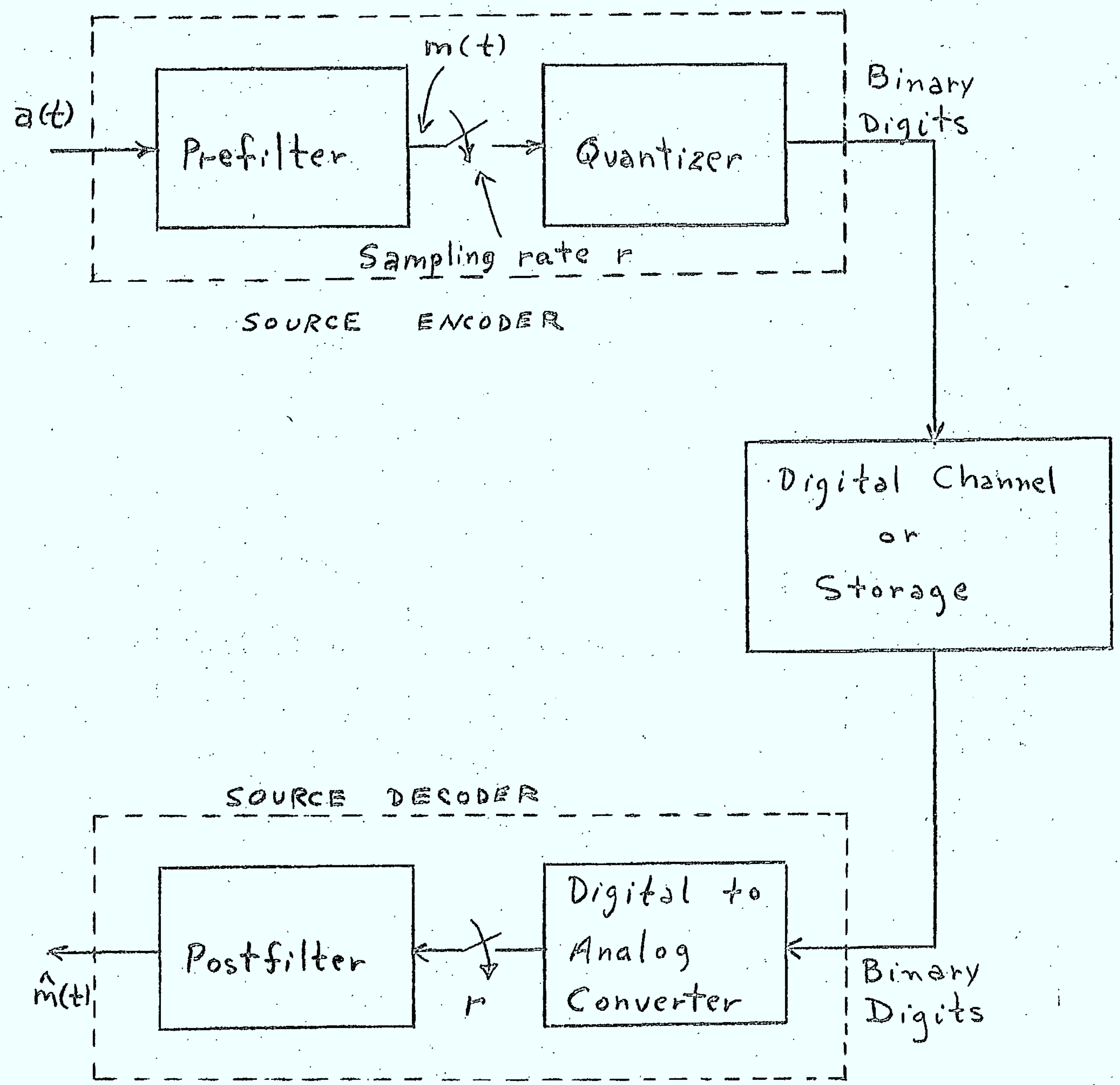

Fig. 5-1: Conventional PCM system

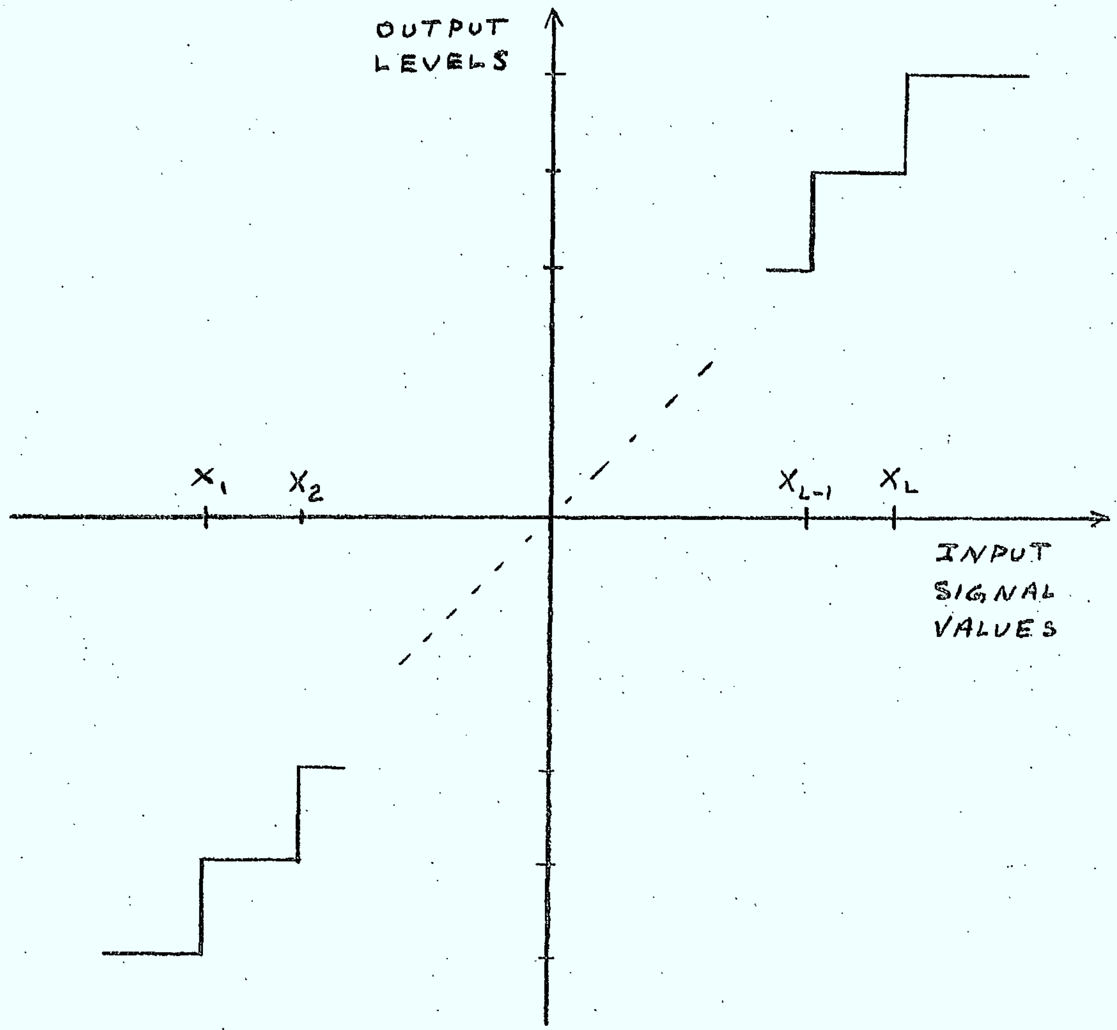

Fig. 5-2: Quantizer characteristic

if the waveform channel's sole source of distortion is white Gaussian Noise, and if optimum incoherent demodulation is used, then [11-13]

$$p=\tfrac{1}{2}\exp(-P/2RNo) \qquad (5\text{-}3)$$

where P is the average power of the received data signal and No/2 is the noise power spectral density.

It is immediately evident that if R is unconstrained then there is an optimum choice for L. Too large a value of L will introduce excessive distortion from digital channel transmission errors. Too small a value of L results in excessive quantization distortion.

Another performance measure is the normalized mean-square difference MSE between the original input signal a(t) and the reconstructed signal $\hat{m}(t)$,

$$MSE = E([a(t)-\hat{m}(t)]^2)\ /E(a^2(t)) \qquad (5\text{-}4)$$

For prefilters with sharp cutoff, analyses [14,15,5] show:

$$MSE \simeq 2\int_{F_c} S_a(f)\ df + \left(\frac{1}{SNR}\right)\frac{r}{2W} \qquad (5\text{-}5)$$

$$\simeq 2\int_{F_c} S_a(f)\ df + \left(\frac{A}{L^2} + 4Bp\right)\frac{r}{2W} \qquad (5\text{-}6)$$

In (5-5) and (5-6), $S_a(f)$ is the power spectral density of a(t), normalized such that $\int_{-\infty}^{\infty} S_a(f)df = 1$, and F_c includes all frequences removed by the prefilter.

The first term in (5-5) is error which results from irreversible removal of some of the frequencies in the input signal a(t).

This term decreases as W increases, while the channel noise term 4Bp increases with W. Thus, (W,L) can be optimized to minimize MSE. In fact, it is not necessary to require r=2W so long as $r \geq 2W$. However, existing evidence [5] indicates that r=2W is the optimum PCM sampling rate.

V-3 Adaptive Differential Encoding of Speech

Fig. 5-3 shows an adaptive differential encoder recently used for encoding speech signals [16]. The dotted lines indicate flow of control information within the source encoder and source decoder. The system in Fig. 5-3 differs from the one depicted in Fig. 5-1 in three respects:

1. The quantizer is not fixed, but adaptive, in the sense that both the horizontal and vertical breakpoints of the quantizer characteristic in Fig. 5-2 change in response to the variance of the signal being quantized, e(kT). The entropy coder in Fig. 5-3 has a buffer which can overflow if quantizer's outer levels, which are normally improbable, occur too often. To prevent this overflow, the levels X_k in Fig. 5-2 are adjusted to prevent buffer overflow. The result is a gradual, rather than sudden and drastic degradation in speech quality.

2. The predictor, which is also adaptive uses a weighted linear sum of previously estimated signal samples $\hat{m}(kT)$ to estimate the current value of the actual input sample m(kT), as follows:

$$z(kT) = \sum_{i=1}^{N} a(i)\ \hat{m}\left[(k-i)T\right] \qquad (5\text{-}7)$$

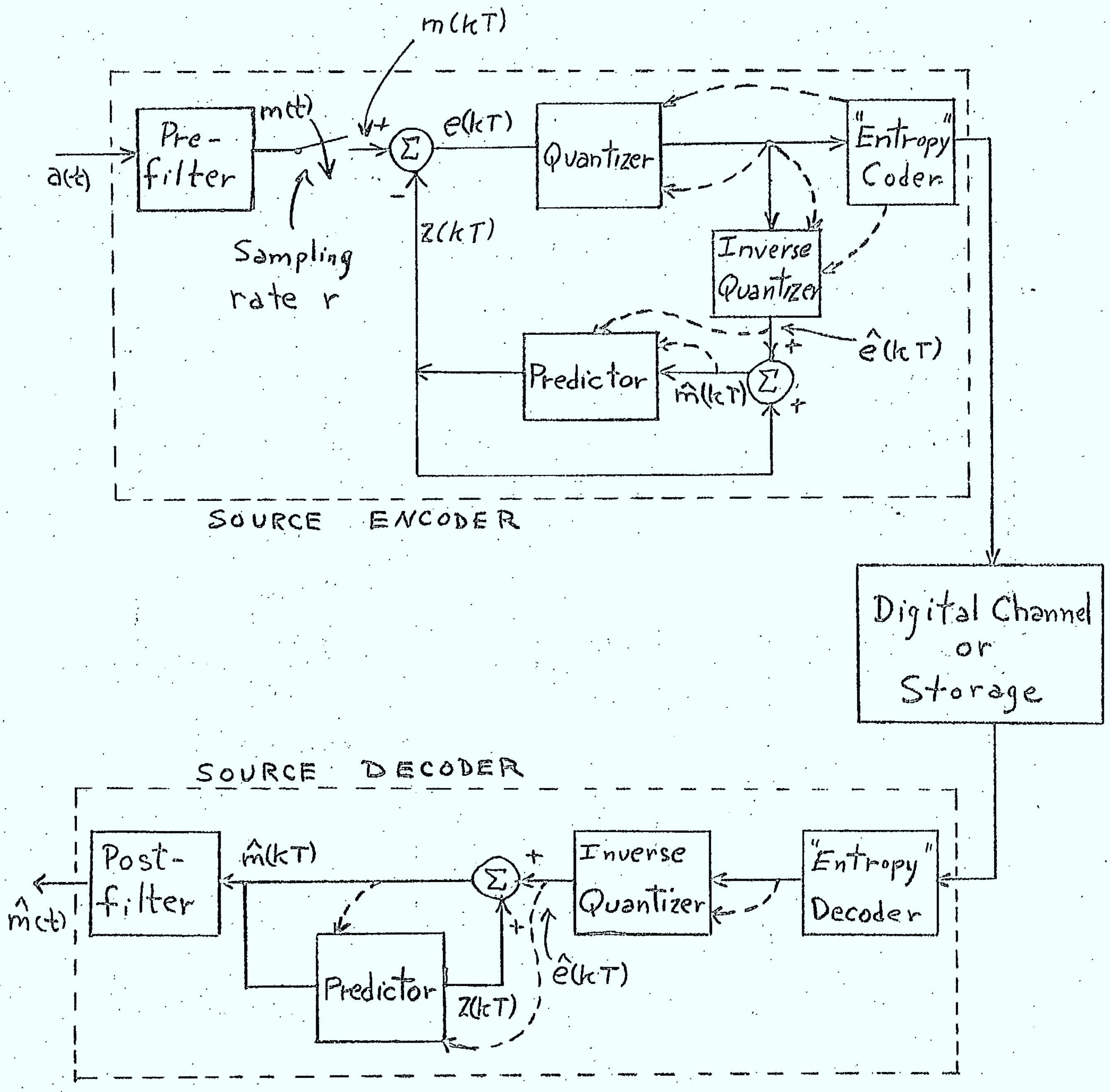

Fig. 5-3: Adaptive differential PCM system (ADPCM)

It is the <u>difference</u> e(kT) between the actual sample m(kT) and its predicted value z(kT) which is quantized. The predictor coefficients a(i) are varied in accordance with estimates of the error signal $\hat{e}(kT)$ and the reconstructed samples $\hat{m}(kT)$.

3. Each quantizer output level is not represented by a fixed number of binary digits. Instead, variable-length sequences of quantizer output levels are coded into fixed-length sequences of binary digits. For example, if the quantizer consists of five output levels, numbered sequentially from lowest to highest, the quantizer output digits are as follows [16]:

Output level sequence	Coder output	Output level sequence	Coder output
333	000	331	111 001
332	001	335	111 010
334	010	31	111 011
32	011	35	111 100
34	100	1	111 011
2	101	5	111 101
4	110		

This conversion from a variable-length input code to a fixed-length output code avoids synchronization problems caused by channel transmission errors; the Huffman coding scheme which maps fixed-length quantizer output sequences into variable-length binary sequences is subject to loss of synchronization if bit-errors occur during transmission [17].

The performance of the above system seems remarkably good [16]. For speech lowpassed at 3.2KHz and sampled at 6.4KHz, the measured output SNR was 13 db and 17 db, respectively [16].

The reconstructed speech was subjectively preferable to PCM speech having higher SNR values; 16 Kbs ADPCM speech fell between 5-bit PCM (32 Kbs) and 6-bit PCM (38.4 Kbs) while 9.6 Kbs ADPCM speech fell between 4-bit PCM (25.6 Kbs) and 5-bit PCM. Little degradation in speech quality was observed with bit-error rates as high as 10^{-3} [16].

As explained below, a differential encoder with faster-than-Nyquist-rate sampling and fewer quantization levels is quite likely to be as good as or better than one restricted to Nyquist-rate sampling. The author is currently involved in a study to optimize the sampling rate using a Cohn-Melsa coder/decoder in which the predictor is fixed. The additional system cost implied by an adaptive predictor may not be justified; Cohn and Melsa [16] estimate that adaptive prediction adds 1.5-2 dB to SNR, but others seem to question this estimate as being too high [18].

The system in Fig. 5-3 has evolved over the past several years, and is intuitively attractive. Correlations between message samples are removed by the predictor, variations in quantizer output level probabilities are used by the coder for further bit rate reductions, and time-varying speech statistics are monitored by the adaptive quantizer and predictor.

If the entropy coder in Fig. 5-3 is removed and if the predictor is non-adaptive, a conventional DPCM system results [5,8,9,14,15]. A conventional delta modulation (DM) system is a conventional DPCM system with two quantizer output levels. Conventional DPCM uses Nyquist-rate sampling, whereas DM samples at 3 or 4 times the Nyquist rate [10,13,14,19-25]. If the predictor is removed from a conventional DPCM system, a PCM system results.

If a DM quantizer is adaptive, an ADM system results [19-21]. As Fig. 5-4 indicates none of these systems yield as high a SNR value for a given bit rate as an ADPCM system, but none are as complex as the ADPCM, either.

For a conventional, well designed DPCM system with a one-tap non-adaptive predictor, analysis yields the following expression for SNR for a quantizer having a number of levels $L \gtrsim 8$ [5,14]:

$$SNR \simeq \left[(A(1 - \phi^2)/L^2) + 4Bp\right] \frac{r}{2W} \qquad (5-8)$$

In (5-8), ϕ denotes the normalized sample-to-sample correlation; $-1 \leq \phi \leq 1$. As sampling rate r increases so does ϕ, which is typically positive for $r \geq 2W$. Thus, as r increases the first term, which results from quantization distortion decreases because $1-\phi^2$ decreases, and r/2W increases. Increasing the sampling rate increases the sample-to-sample correlation and therefore reduces the variance of the quantizer input signal e(kT). It should be noted that the absolute value of the channel distortion need not increase beyond that inherent in PCM systems, contrary to what was initially believed [9]. However the effect of channel distortion relative to quantization noise is larger for DPCM than for PCM [5,7,26].

SNR expressions are available for non-adaptive DPCM systems with an N-order linear predictor [5,8,9] and approximate expressions are available for DM systems [13,20-25]. For adaptive systems SNR is obtainable by measurement on actual or computer simulated systems.

SNR is not always monotonically related to human subjective

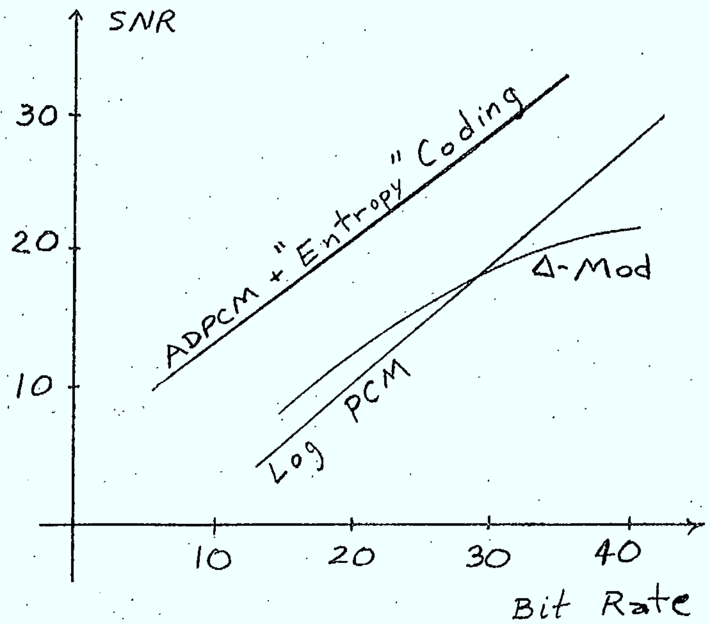

Fig. 5-4: SNR vs bit rate for various speech coders

preference [10,14-16,26-29]. We avoid here detailed discussion of subjective evaluations, other than to say that they are necessary, that they must be done with care, that results of such tests seem to indicate that ADPCM systems having SNR equal to that of PCM systems seem to be subjectively preferable to PCM systems [16,22], and that bit-error probabilities $p \lesssim 10^{-4}$ do not seem noticeable in the reconstructed speech signals of well designed PCM, DPCM, DM or ADPCM systems [18,26,29].

V-4 Speech Analysis and Synthesis

Humans normally speak at a rate of approximately 300 words/minute. An average length of 4.5 characters/word (excluding space) [30,31] and an average entropy of 2.3 bits/character [30-32] gives a speaking rate of less than 60 bits/sec. The actual rate is less because of statistical dependences between words which has been ignored.

If the sole objective of an electronic information system is to convey a replica of the acoustic speech waveform then 9.2 Kbs is more than 100 times larger than necessary. We hasten to point out that it is often desirable to reproduce the speakers emotions as indicated by the subtle variations in pitch, intensity and harmonic content in the acoustic signal.

Drastic reductions in bit-rates below 10 Kbs would seem to require either much more complete, accurate and usable knowledge of speech statistics or knowledge of the human speech production process. Fortunately, the latter is available, and is briefly explained with the help of Figs. 5-5 and 5-6, which represent, respectively, the actual (acoustic) speech production system and a corresponding electronic analog. The analog is obtained by

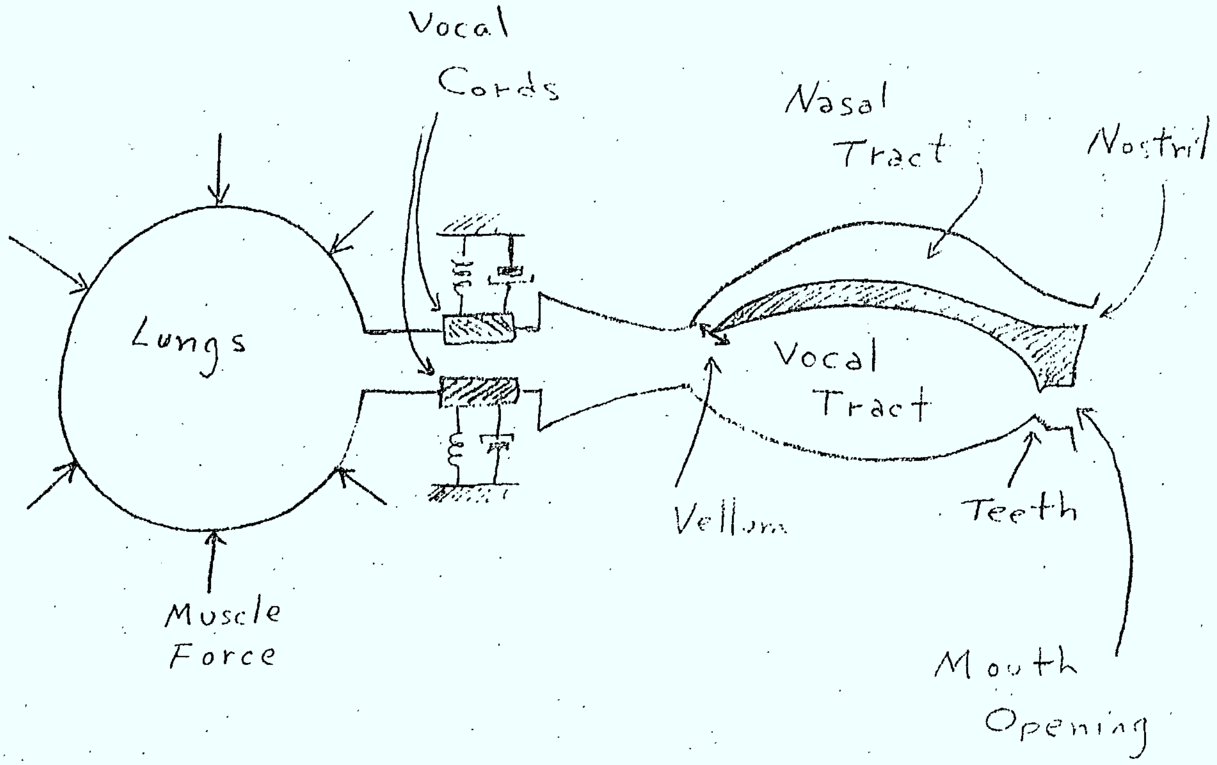

Fig. 5-5: Schematic of the human vocal tract

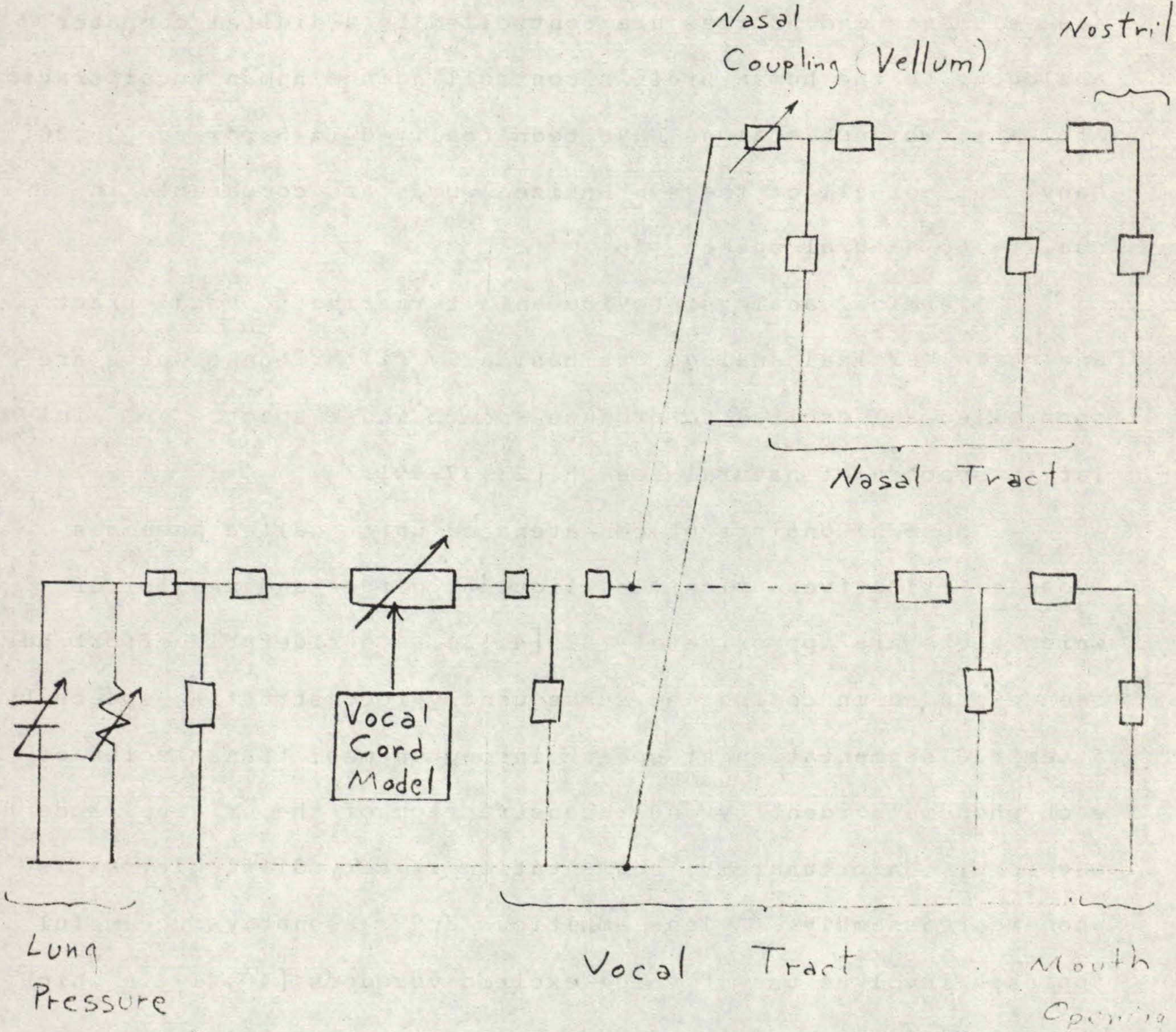

Fig. 5-6: Vocal tract analog

linearizing the fluid flow equations, simulating these by a lumped equivalent of a lossy time-varying transmission line, and driving the line by either a periodic glottal source to simulate voicing or a noise source to simulate turbulence [23,33,34]. The analog's line elements and sources are controlled by a digital computer analogous to the human brain's controlling the human vocal tract. At least two such analogs have been realized in hardware [35,36]. Many, but not all of the synthesized words are comparable in quality to natural speech [36].

Terminal analogs provide an alternative to vocal tract analogs. Terminal analogs are basically filter banks which are controlled and excited to produce speech whose spectre are similar to spectra of natural speech [23,37-39].

Speech consists of concatenated units called phonemes (vowels, fricatives, plosives, liquids, glides and nasals) of which there are approximately 42 [4,33]. Considerable effort has been expended in coding and subsequently reconstructing speech by attempted segmentation of speech into phonemes, transmission of each phoneme's identity and reconstruction of the original speech waveform. Unfortunately, segmentation is very difficult, as is phoneme reassembly. A less ambitious and reasonably successful approach involves use of voice-excited vocoders [40,41] in which the 500-Hz lowpass-filtered speech waveform is sampled and transmitted, along with approximately 12 time-varying linear prediction coefficients (LPC). The excitation function is recovered by nonlinear processing of the 500 Hz signal which is used to drive a terminal analog synthesizer specified by the LPC coefficients. These LPC coefficients are virtually identical to those in the predictor in Fig. 5-3. Speech of good quality has been obtained using 3.6 Kbs. Recent work [42] indicates that if the LPC coefficients are transform coded as explained in Chapter 6 and transmitted differentially, the bit rate required to transmit the LPC parameters would drop considerably. Recent work [43-45] on pitch detection and voice/un-voiced boundary detection may make direct transmission of pitch and voicing amplitude feasible, in which case a bit rate of below 1.5 Kbs for transmission of good quality speech is indicated [42,43].

A further factor of ten reduction in bit rate to approximately 100 units/sec probably requires text synthesis [46-49], whereby stored text segments are converted into signals which control a vocal tract analog or terminal analog to produce natural sounding speech having proper intonation and stress. Although real-time text synthesis of certain sentences is currently feasible, synthesis of arbitrarily selected sentences from stored rules is not and will not soon be feasible. Notwithstanding, text synthesis, as well as formant synthesis, is currently being used to convert computer generated wiring instructions into audio commands acted upon by wire-men who can now accomplish 25% more work because they no longer need to interrupt their work to visually examine wiring diagrams [50,51].

V-5 Machine Recognition of Speech

The general problem of conversing with a machine as if it were a human is of such vast scope and complexity that it is not of immediate technological interest. However, severely constrained versions of the problem, including recognition of words spoken by a co-operative speaker are of interest.

Early algorithms for speech recognition attempted to segment the word into phonemic units, recognize each unit, and finally recognize the word [52-54]. The value of incorporating cotextual dependencies and constraints was recognized [55,56], and attempts were made to use these as an aid in the recognition process. The phoneme-based approach met with only limited success, however, largely because of difficulties inherent in segmentation of the acoustic waveform.

Recent efforts have involved extraction of features whose values vary over the duration of the word, followed by analyses of these features for word recognition. Segmentation is relegated to a limited and manageable role. As in other pattern recognition problems, the decision as to what features to select is the difficult part of the problem [57,58]. Once these are specified and the statistics are available, decisions based on the features are relatively straight forward, at least in principle.

Features employed include speech amplitude (intensity), pitch, zero-crossing rates, or LPC coefficients [60-66]. The LPC coefficients are correlated with spectra, as are zero-crossing rates [65]; zero-crossing rates are also dependent on pitch.

Conclusive evidence of the viability of recent word recognition efforts is available [59,60]. Samber and Rabiner [60] used zero-crossing rate, amplitude and LPC coefficients as features. Each feature is by itself rather unreliable. However the totality of these permits recognition of digits spoken by 25 men and 30 women with 94.4% accuracy [60]. Recent work [61] has demonstrated the feasibility of recognizing a digit sequence. The digits must be segmented from each other; however segmentation of words from a limited vocabulary is less difficult than segmention of individual phonemes within a word.

Word recognition systems provide users with a man-machine interface which permits user mobility, use of both hands for non-communication purposes, vocabulary flexibility, virtually no user training, simple microphonic input, and natural language communication. Some systems are currently in operation in applications involving quality control inspection and verbal command of systems for automated material handling and machine tool numerical control [59].

Developments which will permit use of larger vocabularies, larger speaker populations and improved accuracies for word recognition systems will undoubtedly occur. Attempts to "solve" the larger problem of speech recognition are continuing [62,56]; however, the likelihood of developing a system which recognizes continuous speech with relatively few constraints seems small.

V-6 Speaker Verification

Of growing interest is the possibility of users of electronic information services establishing identity by means of their own voice, thereby gaining access to data and services. Recent evaluations, including evaluations of speech transmitted via telephone lines, indicate verification accuracies of 95% [63,54,66,67], which is adequate for most applications. These systems examine speech intensity and pitch of the acoustic waveform and then align these (slowly) time-varying features to achieve a best match with stored templates of the claimed speaker [63,64,67]. If the specimen's features are sufficiently close to those of the stored templates, the speakers identity claim is accepted, otherwise, the speaker is regarded as an imposter. It should be noted that trained mimics were able to "fool" the system only 4% of the time; human listeners were fooled 22% of the time by mimics [4].

The verification accuracy would likely be improved somewhat by replacing outright rejection of identity claims by instructions to the speaker to speak again, possibly using an alternate sentence. Failure after 3 attempts might then warrent rejection. Use of the LPC coefficients might also yield improvement, since these are characteristic of one's vocal tract, which is not easily modified even by surgical procedures.

V-7 Summary and Conclusions

The progress in coding, synthesis and recognition of speech, as well as speaker verification is apparent. Such progress is due largely to our growing understanding of both the methods of production and the important perceptual attributes of speech. Difficulties in recognizing continuous speech and in synthesizing speech from text results, at least in part, to our lack of understanding of human cognitive processes. The common utility of pitch, intensity and the LPC coefficients as derived from the acoustic speech waveform in ADPCM coding, vocoders, word recognition and speaker verification is apparent.

Regarding speech coding for transmission or storage, cost vs quality trade-offs are potentially available. System users would likely welcome the option of selecting the quality and hence their cost of system usage.

Digital channel error probabilities of 10^{-4} are adequate for transmission of coded speech and probably for those acoustic features used to recognize words and verify speaker identities. Serious consideration should be given to designing channels with 10^{-4} error propabilities, in return for higher data transmission rates. As noted in Chapter 7, error-correcting channel codes are available to provide lower bit-error probabilities for those messages requiring more accurate transmission.

The utility of ADPCM, word recognition and speaker verification is soon to be demonstrated by a commercial airlines reservation system [4]. Designers of user terminals should seriously consider providing voice input/output facilities.

V-8 References

1. W.C. Jakes, Microwave Mobile Communications. New York: Wiley, 1974.

2. N.S. Jayant, "Step-size-transmitting differential coders for mobile telephony," in Conf. Rec., Int. Conf. Commun., San Francisco, Calif., June 1975, pp. (30-6)-(30-10).

3. J. Martin, Design of Man-Computer Dialogues. Englewood Cliffs, N.J.: Prentice-Hall 1973.

4. J.L. Flanagan, "Computers that talk and listen: man-machine communication by voice," Proc. IEEE, vol. 64, pp. 405-415, April 1976.

5. K.Y. Chang and R.W. Donaldson, "Analysis, optimization and sensitivity study of differential PCM systems operating on noisy communication channels," IEEE Trans. Commun., vol. COM-20, pp. 338-350, June 1972.

6. P.F. Panther and W. Dite, "Quantization distortion in pulse count modulation with non-uniform spacing of levels," Proc. IRE, vol. 39, pp. 44-48, Jan. 1951.

7. R.W. Donaldson, "Optimization of PCM systems using natural binary codes," Proc. IEEE, vol. 56, pp. 1252-1253, July 1968.

8. R.A. McDonald, "Signal-to-noise and idle channel performance of differential pulse code modulation systems - particular application to voice signals," Bell Syst. Tech. J., vol. 45, pp. 1123-1151, Sept. 1966.

9. J.B. O'neal, Jr., "Predictive quantizing systems (differential pulse code modulation) for the transmission of television signals," Bell Syst. Tech. J., vol. 45, pp. 689-721, May-June 1966.

10. N.S. Jayant, "Digital coding of speech waveforms: PCM, DPCM and DM quantizers," Proc. IEEE, vol. 62, pp. 611-632, May 1974.

11. J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering. New York: Wiley, 1965.

12. W.C. Lindsay and M.K. Simon, Telecommunications System Engineering. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

13. R.E. Zeimer and W.H. Tranter, Principles of Communications. Boston, Mass.: Houghton Mifflin Co., 1976.

14. R.W. Donaldson and R.J. Douville, "Analysis, subjective evaluation and optimization of the performance capabilities of PCM, DPCM, DM, AM and PM voice communication systems," IEEE Trans. Commun. Technol., vol. COM-17, pp. 421-431, August 1969.

15. D. Chan and R.W. Donaldson, "Optimum pre- and post-filtering of sampled signals with application to pulse modulation and data compression systems," IEEE Trans. Commun. Technol., vol. COM-19, pp. 601-612, Oct. 1971.

16. D.L. Cohn and J.L. Melsa, "The residual encoder - an improved ADPCM system for speech digitization," IEEE Trans. Commun., vol. COM-23, pp. 935-941, Sept. 1975.

17. R.G. Gallager, Information Theory and Reliable Communication. New York: Wiley, 1968.

18. S.U.H. Quireshi and G.D. Forney, Jr., "A 9.6/16 Kb/s speech digitizer," in Conf. Rec., Int. Conf. Commun., San Francisco, Calif., June 1975, pp. (30-31)-(30-36).

19. J.E. Abate, "Linear and adaptive delta modulation," Proc. IEEE, Vol. 55, pp. 298-307, March 1967.

20. P. Cumminsky, "Single integration, adaptive delta modulation," Bell Syst. Tech. J., vol. 54, pp. 1463-1474, Oct. 1975.

21. N.S. Jayant, "Adaptive delta modulation with a one-bit memory," Bell Syst. Tech. J., vol. 49, pp. 321-343, March 1970.

22. P. Cumminsky, N.S. Jayant and J.L. Flanagan, "Adaptive quantization in differential PCM coding of speech," Bell Syst. Tech. J., vol. 52, pp. 1105-1118, Sept. 1973.

23. J.L. Flanagan, Speech Analysis, Synthesis and Perception. New York: Springer-Verlag, 1972.

24. D.J. Goodman, "Delta modulation granular quantization noise," Bell Syst. Tech. J., vol. 48, pp. 1197-1218, May-June 1969.

25. L.J. Greenstein, "Slope overload noise in linear delta modulators with Gaussian inputs," Bell Syst. Tech. J., vol. 52, pp. 387-422, Mar. 1973.

26. J. Yan and R.W. Donaldson, "Subjective effect of channel transmission errors on PCM and DPCM voice communication systems," IEEE Trans. on Commun., vol. COM-20, pp. 281-290.

27. W.A. Munson and J.F. Karlin, "Isopreference method of evaluating speech transmission circuits," J. Acoust. Soc. Amer., vol. 34, pp. 762-774, June 1962.

28. "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust., vol. AU-17, pp. 227-246, Sept. 1969.

29. D.J. Goodman, B.J. McDermott, and L.H. Nakatani, "Subjective evaluation of PCM encoded speech," Bell Syst. Tech. J., vol. 55, pp. 1087-1110, Oct. 1976.

30. H.E. White, "Printed English compression by dictionary encoding," Proc. IEEE, vol. 55, pp. 390-396, Mar. 1967.

31. F. Pratt, Secret and Urgent. New York: Bobbs-Merrill, 1942.

32. C.E. Shannon, "Prediction and entropy of printed English," Bell Syst. Tech. J., vol. 30, pp. 50-64, Jan. 1951.

33. G. Fant, Acoustic Theory of Speech Production. S-Gravenhage, The Netherlands: Houghton and Co., 1960.

34. J.L. Flanagan and K. Ishizaka, "Automatic generation of voiceless excitation in a cord-vocal tract speech synthesizer," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 163-170, Apr. 1976.

35. G. Rosen, "Dynamic Analog Speech Synthesizer," M.I.T. Res. Lab. Electronics, Cambridge, Mass., Tech. Rept. 353, 1962.

36. K.N. Spencer, "Synthesis of Stops, Fricatives, Liquids, and Vowels by a Computer Controlled Electronic Vocal Tract Analog," Dept. Elec. Enggr., Vancouver, Canada, Ph.D. Thesis, Dec. 1971.

37. M.P. Haggard and I. Mattingly, "A simple program for synthesizing British English," IEEE Trans. Audio and Electroacoust., vol. AU-16, pp. 95-99, March 1968.

38. J. Allen, "Speech synthesis from unlimited text," in Speech Synthesis, J.L. Flanagan and L.R. Rabiner, Eds., Philadelphia, Penn.: Dowden, Hutchinson and Ross, Inc., pp. 416-428, 1973.

39. I.G. Mattingly, "Experimental methods for speech synthesis by rule," IEEE Trans. Audio and Electroacoust., vol. AU-16, pp. 198-202, June 1968.

40. B.S. Atal, M.R. Schroeder, and V. Storer, "Voice-excited predictive coding system for low bit-rate transmission of speech," in Conf. Rec., Int. Conf. Commun., San Francisco, Calif., June 1975, pp. (30-37)-(30-40).

41. B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Amer., vol. 50, pp. 637-655, 1971.

42. M.R. Sambur, "An efficient linear prediction vocoder," Bell Syst. Tech. J., vol. 54, pp. 1693-1724, Dec. 1975.

43. L.R. Rabiner, M.J. Cheng, A.E. Rosenberg and C.A. McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-24, pp. 399-418, Oct. 1976.

44. L.R. Rabiner and M.R. Sambur, "Some preliminary experiments in the recognition of connected digits," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 170-182, Apr. 1976.

45. J.J. Dubnowski, R.W. Schaefer and L.R. Rabiner, "Real-time digital hardware pitch detector," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 2-8, Feb. 1976.

46. J. Allen, "Synthesis of speech from unrestricted text," Proc. IEEE., vol. 64, pp. 433-442, Apr. 1976.

47. C.H. Coker, N. Umeda and C.P. Brownman, "Automatic synthesis from ordinary English text," in Speech Synthesis, J.L. Flanagan and L.R. Rabiner, Eds. Stroudsbourg, Penn.: Dowden, Hutchinson and Ross, Inc., 1973.

48. C.H. Coker, N. Umeda and C.P. Brownman, "Automatic synthesis for ordinary English," IEEE Trans. Audio Electroacoust., vol. AU-21, pp. 293-297, June 1973.

49. N. Umeda, "Linquist rules for text-to-speech synthesis," Proc. IEEE, vol. 64, pp. 443-451, Apr. 1976.

50. L.R. Rabiner and R.W. Schaefer, "Digital techniques for computer voice response: implementations and applications," Proc. IEEE, vol. 64, pp. 416-433, Apr. 1976.

51. J.L. Flanagan, L.R. Rabiner, R.W. Schaefer and J.D. Denman, "Wiring telephone apparatus from computer generated speech," Bell Syst. Tech. J., vol. 51, pp. 391-397, Feb. 1972.

52. D.R. Reddy, "Segmentation of speech sounds," J. Acoust. Soc. Amer., vol. 40, pp. 307-312, 1966.

53. G.W. Hughes, The Recognition of Speech by Machine, MIT Res. Lab. of Electronics, Cambridge, Mass., Tech. Rept. 395, 1961.

54. D.R. Reddy, "Computer recognition of connected speech," J. Acoust. Soc. Amer., vol. 42, pp. 329-346, 1967.

55. R. Alter, Utilization of contextual constraints in automatic speech recognition," IEEE Trans. Audio Electroacoust., vol. AU-16, pp. 6-11, Mar. 1968.

56. F. Jelinek, "Continuous speech recognition by statistical methods," Proc. IEEE, vol. 64, pp. 532-556, Apr. 1976.

57. L. Kanal, "Patterns in pattern recognition: 1968-1974," IEEE Trans. Inform. Theory, vol. IT-20, Nov. 1974.

58. R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis. New York: Wiley, 1973.

59. T.B. Martin, "Practical applications of voice input to machines," Proc. IEEE, vol. 64, pp. 487-501, Apr. 1976.

60. M.R. Sambur and L.R. Rabiner, "A speaker-independent digit recognition system," Bell Syst. Tech. J., vol. 54, pp. 81-102, 1975.

61. L.R. Rabiner and M.R. Sambur, "Some preliminary experiments in the recognition of connected digits," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 170-182, Apr. 1976.

62. D.R. Reddy, "Speech recognition of speech by machine: a review," Proc. IEEE, vol. 64, pp. 501-531, Apr. 1976.

63. A.E. Rosenberg and M.R. Sambur, "New techniques for automatic speaker verification," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-23, pp. 169-176, April 1975.

64. M.R. Sambur, "Selection of acoustic features for speaker identification," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-23, pp. 176-182, April 1975.

65. M.R. Ito and R.W. Donaldson, "Zero-crossing measurements for analysis and recognition of speech sounds," IEEE Trans. Audio Electroacoust., vol. AU-19, pp. 235-242, Sept. 1971.

66. D.A. Wasson and R.W. Donaldson, "Speech amplitude and zero crossings for automated identification of human speakers," IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-23, pp. 390-392, Aug. 1975.

67. A.E. Rosenberg, "Automatic speaker verification: a review," Proc. IEEE, vol. 64, pp. 475-487, Apr. 1976.

68. A.E. Rosenberg, "Evaluation of an automatic speaker verification system over telephone lines," Bell Syst. Tech. J., vol. 55, pp. 723-744, July-Aug. 1976.

VI IMAGE CODING, STORAGE, TRANSMISSION AND RECONSTRUCTION

VI-1 Images and Electronic Information Services

The general tenor of the comments made in Section V-1 for speech apply also to images. It is not known at this time whether speech or images will ultimately contribute more bits to the totality of EIS data traffic; nor is it known which will emerge as the preferred means of presenting information to users.

Interest in image processing and transmission is growing at an astounding rate. A recent survey [1] of works from selected English language journals published in 1975 and excluding papers in computer graphics (synthesis and manipulation of pictoral information), optical and photographic processing and visual perception included 354 separate items.

This chapter includes only aspects relating to the conversion of stationary images (including color images) to and their reconstruction from digital data. However, the discussion includes picture properties and characterizations which bear on computer graphics and animation, image enhancement and restoration, pictoral pattern recognition, shape and texture analysis, and picture grammars.

VI-2 Image Properties, Characterizations and Distortion Measures

A monochrome image may be represented as $I(x,y)$ when x and y are rectangular spatial co-ordinates and I is light intensity (brightness). If $S(x,y)$ and $R(x,y)$ denote, respectively, the incident light intensity and surface reflectivity then [2,3]

$$I(x,y) = S(x,y)\ R(x,y) \qquad (6\text{-}1)$$

Incident light S typically varies slowly over space, while the reflectivity varies rapidly. If (6-1) is rewritten

$$\log I(x,y) = \log S(x,y) + \log R(x,y) \qquad (6\text{-}2)$$

and if log I(x,y) is high-pass filtered, then variations in S are removed while those in R remain. A remarkable restoration of an image dulled in places by shadows can be effected [2,4].

Images are highly structured. If a rectangular piece of paper were divided into 256 x 256 equal-sized rectangles (called picture elements or pels) and if each pel were shaded by an amount independent of that by which the other pels were shaded, the appearance of a meaningful pattern would be very unlikely [5].

Images are composed of areas of various shapes and sizes; in any one area the brightness is approximately constant. The probability distribution of area size was estimated [6] for four monochrome images quantized to 256 points on each edge and eight brightness levels. Image samples included a head and shoulders (low detail), cameraman (medium detail), crowd scene (high detail) and cloud formation (large areas of black and white). For all four images the fraction of areas consisting of n pels decreased as $1/n^2$. Many areas comprised few pels; in fact, approximately half of the areas consisted of a single pel. Some areas included a large number of pels. The average number of pels per unit area varied from 29 for the crowd scene to 83 for the cloud scene.

Statistics describing image structure are meagre. The first order amplitude probability of the intensity typically has two peaks, one in the black and the other at the white end of the scale [6,7]; however, in some instances intensity is more nearly uniform. The average intensity and normalized correlation function

$$\rho(x,y) = E[I(u,v)\ I(x-u,\ y-v)] \tag{6-3}$$

where E[ ] denotes expected value are reasonably independent of x and y; ρ is reasonably well approximated by either (6-4) or (6-5) below where α and β are constants which depend on the image class [7-12].

$$\rho(x,y) = e^{-(\alpha x + \beta y)} \tag{6-4}$$

$$\rho(x,y) = e^{-\sqrt{(\alpha x)^2 + (\beta y)^2}} \tag{6-5}$$

The model in (6-5) is appealing because of its spatially isotropy; (6-4) is convenient for analysis involving rectangular co-ordinates.

The discussion thus far has emphasized monochrome images. A color image depends on light wavelength λ as well as spatial variables x and y. Any intensity function $I(x,y,\lambda)$ can be reconstructed from three appropriately selected colors [13,14]. Colors red (R), green(G) and blue (B) are nearly always selected; thus

$$I(x,y,\lambda) = A_R(\lambda)\ R(x,y) + A_G(\lambda)\ G(x,y) + A_B(\lambda)\ B(x,y) \tag{6-6}$$

where A_R, A_G, A_B depend on the actual filter attenuation functions peculiar to the three color regions.

Color images are often represented in terms of co-ordinates which are linear combinations of the R, G, B functions. Thus,

$$\begin{aligned} Y &= 0.30\ R + 0.59\ G + 0.11\ B \\ I &= 0.60\ R - 0.28\ G - 0.32\ B \\ Q &= 0.21\ R - 0.52\ G + 0.31\ B \end{aligned} \tag{6-7}$$

Inversion of (6-7) yields

$$R = 1.00\ Y + 0.96\ I + 0.62\ Q$$
$$G = 1.00\ Y - 0.27\ I - 0.65\ Q \qquad (6\text{-}8)$$
$$B = 1.00\ Y - 1.11\ I - 1.70\ Q$$

In (6-7) and (6-8), which is the NTSC representation [15], Y is the (monochrome) intensity signal, and I and Q carry the color information; all three signals are functions of x and y. The I and Q information content is much less than that for Y.

IV-3 Evaluating Image Quality

How many bits are required to represent an image such that reconstruction will yield a replica whose distortion from the original image is acceptable? The answer provides a lower bound to the cost of digitally storing or transmitting images, and also provides a benchmark against which to compare alternative image coding schemes.

Unhappily, the answer to the above question is unavailable to us for two reasons. First, a distortion measure having the property that an increase in image distortion is always accompanied by a decrease in perceived quality is not yet available. Second, an accurate statistical description of images is not available, as noted earlier. (For a good discussion of the need for distortion measures and statistical models, see [5]).

With reference to Fig. 6-1, signal-to-noise ratio SNR is defined as follows,where I and $\hat{I}$ denote, respectively, the original prefiltered image and its reconstructed replica.

$$SNR = E[(I(x,y,\lambda) - \hat{I}(x,y,\lambda))^2]/E(I^2(x,y,\lambda)) \qquad (6\text{-}9)$$

Because of its intuitive appeal and analytical convenience SNR is

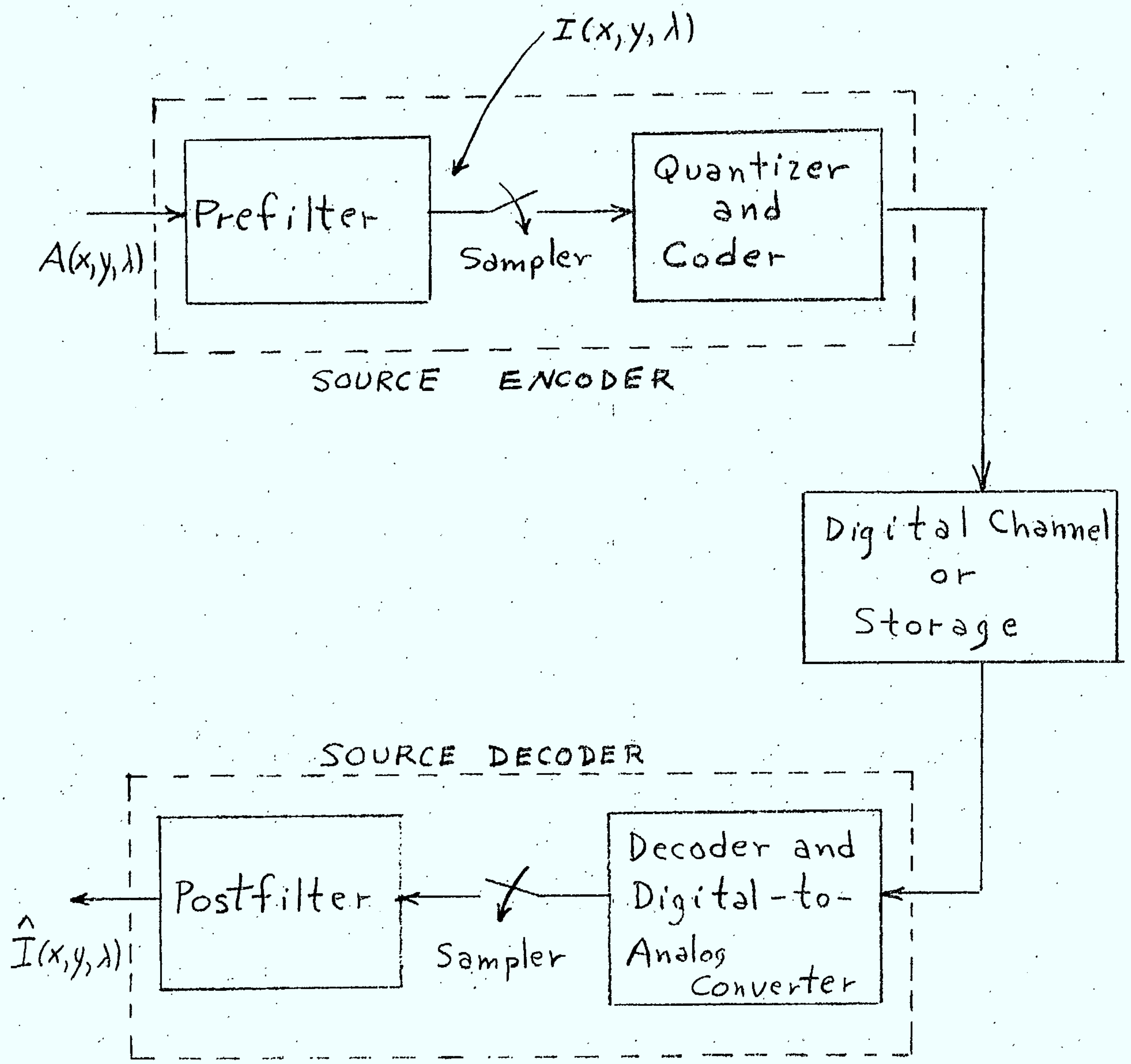

Fig. 6-1: Image coding and transmission/storage system

often used as a performance measure even though its relationship to perceived image quality is not monotonic.

An alternative and somewhat more appealing distortion measure was evaluated by Mannos and Sakrison [7] who weighted the difference between the original and reconstructed image in accordance with the amplitude of the difference signal's frequency components. Various weighting functions were selected in accordance with the known properties of the human visual system.

If distortions due to prefiltering the original image are to be included in performance calculations, then the mean-square error

$$MSE = E[A(x,y,\lambda) - \hat{I}(x,y,\lambda)]^2/E[A^2(x,y,\lambda)] \qquad (6\text{-}10)$$

where $A(x,y,\lambda)$ is the original unfiltered image, can be used. As with speech, for Nyquist-rate sampling

$$MSE \simeq \iiint_{F_c} S_A(f,g,h)\, df dg dh + \frac{1}{SNR} \qquad (6\text{-}11)$$

where $S_A(f,g,h)$ is the (three-dimensional) power spectral density of the image and F_c includes all frequencies removed by the prefilter. For a prefilter which passes all color frequencies f as well as all spatial frequences $g \leqslant W$ and $h \leqslant W$,

$$\iiint_{F_c} S_A(f,g,h)\, df dg dh = 8\int_{f=0}^{\infty}\int_{g=W}^{\infty}\int_{h=W}^{\infty} S_A(f,g,h)\, df dg dh \qquad (6\text{-}12)$$

Final evaluation of image processing or transmission schemes requires human observers to participate in tests in which various recontructed images are rank ordered, or compared in pairs. Subjective testing is a specialized topic; the interested reader should consult the literature [7,10,16,17].

VI-4 Coding and Reconstructing Monochrome Images

Monochrome images can be coded on either a line-by-line or an area basis.

In lin-by-line (raster-scan) coding, the image intensity, which depends on two spatial variables, is converted into a one-dimensional function of time by scanning the image horizontally, at a fixed rate. The one-dimensional signal is then encoded by means of PCM, DPCM or ADPCM in the same way that speech is encoded.

An alternative approach is to encode the image area by representing image intensity I(x,y) as a two-dimensional orthonormal series, retaining and quantizing only those coefficients u_{ij} whose energy exceeds a threshold, and later reconstructing the image from these retained coefficients. Thus, I(x,y) is approximated by $\hat{I}(x,y)$ as follows, (* denotes complex conjugate):

$$\hat{I}(x,y) = \sum_{i=0}^{N} \sum_{j=0}^{M} u_{ij}\, \phi_{ij}(x,y) \qquad 0 < x \leq a,\quad 0 < y \leq b \tag{6-13}$$

$$u_{ij} = F\{\int_0^a \int_0^b I(x,y)\, \phi_{ij}^*(x,y)\,dxdy\} \tag{6-14}$$

$$\int_0^a \int_0^b \phi_{ij}^*(x,y)\, \phi_{kl}(x,y)\,dxdy = \begin{cases} 1, & i = k \text{ and } j = l \\ 0, & i \neq k \text{ or } j \neq l \end{cases} \tag{6-15}$$

where F denotes the non-linear quantization operator.

The operation implied by (6-13) to (6-15) is called transform encoding [5,8,11,12]. Area encoding schemes are appealing because statistical dependencies in two spatial directions are used to remove image redundancies. Line-by-line coding does not normally utilize redundancies in the y (vertical) direction. Potential coding gains inherent in area encoding as compared with line-by-line coding has been determined for Gaussian images [18].

Orthogonal function sets $\{\phi_{ij}(x,y)\}$ used to encode images include the Fourier, Hadamard, and Karhunen-Loeve Functions [5,7, 8, 11,18-23]. To reduce coder costs, the image is usually divided into blocks of n x n samples, where n = 4, 8, or 16. Blocks are coded individually, stored or transmitted, and later reconstructed and reassembled.

Use of results from various sources [8,24,25] allows the normalized mean-square error MSE between an original and reconstructed image as well as SNR to be expressed as follows:

$$MSE = \sum_{i=N+1}^{\infty} \sum_{j=M+1}^{\infty} \sigma_{ij}^2 + \frac{1}{SNR} \tag{6-16}$$

$$SNR = \sum_{i=0}^{N} \sum_{j=0}^{M} \left(\frac{A_{ij}}{L_{ij}^2} + 4B_{ij} P_{ij}\right) \frac{\sigma^2_{ij}}{\sigma^2} \tag{6-17}$$

$$\sigma^2 = E(I^2(x,y))$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E[|u_{ij}|^2]$$

where A_{ij} and B_{ij} are constants approximately equal to unity, $\sigma^2_{ij} = E(|u_{ij}|^2)$, L_{ij} and P_{ij} denote respectively the number of quantizer output levels and bit-error probability (assuming random errors) applicable to the various coefficients, and σ^2 is the mean-square image intensity. Normally $P_{ij} = p$ for all i, j.

Distortions arise from three separately identifiable sources:

1. removal (filtering) of some coefficients
2. quantization of the coefficients
3. channel transmission errors or memory read/write errors

The comments in Chapter 5 regarding optimal balancing of the three types of distortions are appropriate here, as are the comments regarding the potential availability of quality-cost trade-offs.

Coefficients u_{ij} having large variance in (6-17) contribute more to degradations than do those of smaller variance, unless each coefficient is coded (adaptively) with a resolution dependent on its energy. Because statistics vary over subpicture blocks, use of quantization schemes which are adaptive over various subpictures is indicated [18-21].

Much effort has been devoted to quantizer optimization [24-28], with fewer studies [22] being devoted to consideration of the effects of channel transmission errors on reconstructed image quality.

For either Fourier or Hadamard basis functions, non-adaptive coding with optimum (but fixed) quantization of each subpicture's coefficients requires approximately 2 bits/pel if the reconstructed images are to be comparable to the original. Adaptive schemes require 1 bit/pel, but are considerably more expensive. Random bit-error probabilities of $p \simeq 10^{-4}$ or better are virtually unnoticed in reconstructed images. Implementation considerations favour nonadaptive quantization and Hadamard basis functions since $\phi_{ij}(x,y) = \pm 1$ for any (x,y) pair. Considerations based on area properties of images and psychovisual effects also motivate Hadamard functions [5].

Returning to line-by-line coding, PCM requires 6 or 7 bits/pel if the reconstructed images are to appear comparable with the pre-filtered originals. Previous sample nonadaptive DPCM requires 4 to 5 bits/pel [21,22,25,29,30]. Using additional feedback samples, including the one on the line above to reduce the prediction error reduces the number of bits to 3.5 bits/pel [25,29,20]. If the quantizer output levels are entropy-coded, an additional 0.5 bits/pel may be saved [11,22,29].

A reasonably thorough study of line-by-line adaptive encoding, similar to the Cohn-Melsa [31] study, is lacking. The differential adaptive one-bit quantizer with overshoot suppression and over-sampling yielded reconstructed images of reasonable quality [32]. On the basis of Cohn and Melsa's [31] results for speech, one might expect that a four- or five- level adaptive quantizer with entropy coding and adaptive prediction would yield good quality reconstructed images at 2 or even 1 bit/pel.

Random bit-error probabilities $p \simeq 10^{-4}$ or better are virtually undiscernable on reconstructed images coded line-by-line on a well designed system [29,30,33].

VI-5 Coding and Reconstructing Color Images

Relatively few studies have been conducted on color image coding, one of the reasons being that color imagery test facilities are not readily available.

Although color adds additional information or uncertainty to an image, it also adds a very rich psychophysical dimension [34,35], with the result that color images convey more information to the user than do monochrome images of the same bit rate. Compare, for example, black and white vs. color television. Color signals occupy the same bandwidth as monochrome signals, however color is obviously more pleasing to watch and is more informative. For example, color facilitates determination of the combatants in a football game, and whether or not the game is being played on artificial turf.

The RGB signals in (6-8), or any non-singular linear combination of these can be coded for storage or transmission. Bushan [36] used the RGB signals, and found green to be subjectively the most sensitive to distortion, and blue the least sensitive.

Coding the YIQ signals is advantageous because these are the standard NTSC signals delivered to the home television receiver. The Y signal yields a monochrome image, and contains most of the information; fewer than 0.4 bits/pel are needed to code each of the I and Q signals [37-39]. Pratt [38] and Seecheran [40] each obtained good quality reconstructed images using fewer than 1.75 bits/pel. Seecheran used Hadamard coding, while Pratt used Karhunen-Loeve coding on the Y signal and Fourier coding of the I and Q signals.

A novel approach was investigated by Soubigou [41] who used three-dimensional Fourier transformation to code the intensity function $I(x,y,\lambda)$, which was reconstructed with good quality, using 2 bits/pel.

The Y signal, as well as the I and Q signals can be encoded using PCM, DPCM or ADPCM. In the absence of adequate data, we estimate that the bit rate required for color images is equal to or less than that for monochrome images of comparable quality. Existing evidence indicates that bit-error probabilities of 10^{-4} or better are virtually imperceptible in reconstructed color images [42].

VI-6 Concluding Remarks

Unless transform coding is used, and current implementation costs would likely be prohibitive in many applications, at least two bits/pel are required to code an image for subsequent high-quality reconstruction. A spatial sampling rate of at least 256 x 256 points is needed, with the result that each image requires at least $2^{17} \simeq 1.3 \times 10^5$ bits/image. Even at this bit-rate, ADPCM coding or Hadamard transform coding would be required.

The fact that images, like speech require channels having random error probabilities $p \lesssim 10^{-4}$ again motivates the suggestion that electronic information system data channels should perhaps be designed for $p \simeq 10^{-4}$. As noted in Chapter 7, error-correcting channel codes are available for messages requiring more accurate transmission.

It seems unlikely that coding schemes capable of reconstructing good quality images of moderate detail will have bit rates much below 10^5 bits/image. More efficient coding schemes would necessarily be based on a better understanding of image structure, in the same way that improved efficiencies for speech coding are based on knowledge and understanding of the human speech production mechanism (See Section V-4).

In this regard there is a difference; speech is produced by humans for perception by humans, whereas image sources such as gardens, cloud scenes and animals occur naturally and may be perceived by humans. There is no human image production mechanism, unless an artist is sketching or painting a scene which he is observing. Should we observe the way in which artists construct paintings and drawings in order to determine the basic building blocks of images? We might learn something, or we might be dismayed by the (apparently) different perceptions of a scene by different artists. Or are these perceptions really so different? Different individuals undoubtedly perceive spoken utterances differently. In fact, it may be that the intersection of different observers' interpretations of various specific scenes contains useful clues for efficient image representation. Better understanding of how humans perceive images, as well as available data on image statistics and perceptual effects of various distortions would undoubtedly be useful.

In many applications display of a visual concept rather than accurate reconstruction of a specific image is required. It would be useful to be able to construct the appropriate images from stored concept or information segments by methods similar to those used to generate speech from stored text segments. How should the stored segments be defined and described? How should these be converted into meaningful visual images? The answers to these questions, which involve computer graphics [43,44] are not known. The representation of an image as a weighted sum of basis functions has obvious appeal. The difficulty is to define the basis functions. Once these are specified, weighting coefficients for a set of images might be obtainable by trial and error.

Differential and transform methods can be extended to encode movies if the intensity function I includes time t as an independent variable; thus $I = I(x,y,t)$ and $I = I(x,y,\lambda,t)$ for monochrome and color images, respectively. Coding movies, mainly for picturephone applications has been studied [45-48].

VI-7 References

1. A. Rosenfeld, "Picture processing: 1975," Computer Graphics and Image Proc., vol. 5, pp. 215-237, June, 1976.

2. T.G. Stockholm, Jr., "Image processing in the context of a visual model," Proc. IEEE, vol. 60, pp. 828-842, July, 1972.

3. H.C. Andrews and B.R. Hunt, Digital Image Restoration, Englewood Cliffs, N.J.: Prentice-Hall, 1977.

4. A.V. Oppenheim and R.W. Scaefer, Digital Signal Processing, Englewood Cliffs, N.J.: Prentice-Hall, 1975.

5. H.J. Landau and D. Slepian, "Some computer experiments in picture processing for bandwidth compression," Bell Syst. Tech. J., vol. 50, pp. 1525-1540, May-June 1971.

6. A. Nishikawa, R.J. Massa and J.C. Mott-Smith, "Area properties of television pictures," IEEE Trans. on Inform. Theory, vol. IT-11, pp. 348-352, July 1965.

7. J.L. Mannos and D.J. Sakrinson, "The effects of a fidelity criterion on the encoding of images," IEEE Trans. Inform. Theory, vol. IT-20, pp. 525-536, July 1974.

8. A. Habibi and P.A. Wintz, "Image coding by linear transformation and block quantization," IEEE Trans. Commun. Technol., vol. COM-19, pp. 50-62, Feb. 1971.

9. L.E. Franks, "A model for the random video process," Bell Syst. Tech. J., vol. 45, pp. 609-630, April 1966.

10. T.S. Huang, "The subjective effect of two-dimensional pictorial noise," IEEE Trans. Inform. Theory, vol. IT-11, pp. 43-53, Jan. 1965.

11. J.B. O'Neal, Jr., and T.R. Natarajan, "Coding istropic images," in Conf. Rec., Int. Conf. Commun., Philadelphia, Penn., 1976, pp. (47-1)-(47-6).

12. E.R. Kretzmer, "Statistics of television signals," Bell Syst. Tech. J., vol. 31, pp. 751-763, July 1952.

13. R.W.G. Hunt, The Reproduction of Color, New York, New York: Wiley, 1967.

14. C. Wyzecki and W.S. Stiles, Color Science, New York: Wiley, 1967.

15. A.B. Carlson, Communication Systems, 2nd ed., New York: McGraw-Hill, 1975, Appendix C.

16. J. Yan and R.W. Donaldson, "Subjective effects of channel transmission errors on PCM and DPCM voice communication systems," IEEE Trans. Commun., vol. COM-20, pp. 281-290, June 1972.

17. W.S. Torgerson, Theory and Methods of Scaling, New York: Wiley, 1958.

18. D.J. Sakrison and V.R. Algazi, "A comparison of line-by-line and two-dimensional coding of random images," IEEE Trans. on Inform. Theory, vol. IT-17, pp. 386-396, July 1971.

19. M. Tasto and P.A. Wintz, "Image coding by adaptive block quantization," IEEE Trans. Commun., vol. COM-19, pp. 957-971, Dec. 1971.

20. P.A. Wintz, "Transform picture coding," Proc. IEEE, vol. 60, pp. 809-820, July 1972.

21. A. Habibi, "Comparison of n^{th} order DPCM encoder with linear transformations and block quantization techniques," IEEE Trans. Commun., vol. COM-19, pp. 948-956, 1971.

22. A. Habibi, "Hybrid coding of pictorial data," IEEE Trans. Commun., vol. COM-23, pp. 614-624, May 1974.

23. A. Habibi and R.S. Hershel, "A unified representation of differential pulse code (DPCM) modulation and transform coding systems," IEEE Trans. on Commun., vol. COM-22, pp. 629-696, May 1974.

24. A.J. Kurtenbach, "Quantizing for noisy channels," IEEE Trans. Commun. Tech., vol. COM-17, pp. 271-302, April 1969.

25. J.B. O'Neal, Jr., "Predictive quantizing systems (differential pulse code modulation) for the transmission of television signals," Bell Syst. Tech. J., vol. 45, pp. 689-721, May-June 1966.

26. J. Max, "Quantizing for minimum distortion, IEEE Trans. Inform. Theory, vol. IT-6, pp. 7-12, March 1960.

27. R.C. Wood, "On optimum quantization," IEEE Trans. Inform. Theory, vol. IT-15, pp. 248-252, March 1969.

28. J.J.Y. Huang and P.M. Schultheiss, "Block quantization of correlated Gaussian random variables," IEEE Trans. Commun. Syst., vol. CS-11, pp. 289-296, Sept. 1963.

29. D.J. Carron, R.C. Brainard and J.O. Limb, "Intraframe coding for picture transmission," Proc. IEEE, vol. 60, pp. 779-791, July 1972.

30. K.Y. Chang and R.W. Donaldson, "Non-adaptive DPCM transmission of monochrome pictures over noisy communication channels," IEEE Trans. Commun., vol. COM-24, pp. 173-183, Feb. 1976.

31. D.L. Cohn and J.L. Melsa, "The residual encoder - an improved ADPCM system for speech digitization," IEEE Trans. Commun., vol. COM-23, pp. 935-942, Sept. 1975.

32. L. Weiss, I.M. Paz and D.L. Schilling, "Video encoding using an adaptive delta modulator with overshoot suppression," IEEE Trans. Commun., vol. COM-23, pp. 905, 920, Sept. 1975.

33. N. Scheinberg and D.L. Schilling, "Techniques for correcting transmission errors in video adaptive delta modulation channels," IEEE Trans. Commun., vol. COM-24, pp. 1064-1070, Sept. 1976.

34. N.S. Anderson and P.S. Fitts, "Amount of information gained during brief exposures to numerals and colors," J. Exper. Psychol. vol. 56, pp. 362-369, March 1956.

35. G.A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," Psychol. Rev., vol. 63, pp. 81-97, March 1956.

36. A.K. Bushan, "Experiments in coding color pictures," in Picture Bandwidth Compression, T.S. Huang and O.J. Tretiak, eds., New York: Gordon and Breach, pp. 697-726, 1972.

37. J.O. Limb, C.B. Rubenstein and K.A. Walsh, "Digital coding of color picturephone signals by element-differential quantization," IEEE Trans. Commun. Tech., vol. COM-19, pp. 992-1005, Dec. 1971.

38. W.K. Pratt, "Spatial transform coding of color images," IEEE Trans. Commun. Tech., vol. COM-19, pp. 980-992, Dec. 1971.

39. J.O. Limb and C.B. Rubenstein, "Plateau coding of the chrominance component of color picture signals," IEEE Trans. Commun., vol. COM-22, pp. 812-820, June 1974.

40. C.A. Seecheran, "Color image coding using linear transformation and block quantization," Dept. Elect. Eng. M.A.Sc. Thesis, Univ. B.C., Vancouver, Canada, June 1974.

41. A. Soubigou, "Three-dimensional transform encoding and block quantization of still color and monochrome moving pictures," Dept. Elect. Eng. Ph.D. Thesis, Univ. B.C., Vancouver, Canada June 1975.

42. W.K. Pratt, "Binary symmetric channel error effects on PCM color image transmission," IEEE Trans. Inform. Theory, vol. IT-18, pp. 636-643, Sept. 1972.

43. A. Rosenfeld, Picture Processing by Computer, New York: Academic Press, 1969.

44. W.M. Newman and R.F. Sproull, Principles of Interactive Computer Graphics, New York: McGraw-Hill, 1973.

45. B.G. Haskell, F.W. Mounts and J.C. Candy, "Interframe coding of videotelephone pictures," Proc. IEEE, vol. 60, pp. 793-800, July 1972.

46. B.G. Haskell and R.L. Schmidt, "A low-bit-rate interframe coder for videotelephone," Bell Syst. Tech. J., vol. 54, pp. 1475-1495, Oct. 1975.

47. J.O. Limb, R.F.W. Pease and K.A. Walsh, "Combining intraframe and frame-to-frame coding for television," Bell Syst. Tech. J., vol. 53, pp. 1137-1173, July-Aug. 1974.

48. F. Giorda and A. Racciu, "Bandwidth reduction of video signals via shift vector transmission," IEEE Trans. Commun., vol. COM-23, pp. 1002-1004, Sept. 1975.

VII CODING, STORAGE, TRANSMISSION AND RECONSTRUCTION OF FACSIMILIE, TEXT AND TEXT SEGMENTS

VII-I Facsimilie, Text, and Electronic Information Services

Of special interest for many electronic information service applications are images whose intensity I(x,y) assumes one of two nominal values, namely black (B) or white (W). Text, line drawings and maps are included in the class of two-level images called facsimilie (FAX).

Actual FAX documents are not strictly black or white, particularly at boundaries where the B-W and W-B transitions proceed through a continuum of gray levels. For this reason, the FAX intensity function I(x,y) is quantized to two levels before being coded for storage or transmission.

Techniques developed for coding monochrome images, including differential or transform coding methods can be employed for coding FAX. However, FAX documents have special properties which are particularly suited to special schemes, including run-length coding.

VII-2 Properties of FAX Information Sources

Fax documents differ from monochrome images such as human faces, crowd scenes, buildings, gardens, and cloud scenes in several ways.

First, the fact that the information in a FAX image lies mainly in the B-W and W-B transitions often makes the source bandwidth as determined by the source's two-dimensional Fourier transform much wider than that for conventional images. While an image is often adequately represented by 256x256 samples, a page of text of 70- characters line-width requires 1000x1000 spatial samples.

Second, FAX documents normally contain much more white than black [1-4].

Third, if FAX documents are raster-scanned, runs of black alternating with white occur. The length of each run tends to be statistically independent of other run-lengths [2]. The probability of the run-lengths seems to fall off as a power of the length [2]. If text is scanned parallel to the lines, those scans which fall between lines of text yield an all-white output, in which case special run-length codes become attractive [3].

VII-3 Run-Length Coding

Because black and white run-lengths tend to be statistically independent, it is possible to calculate run-length entropy H as follows, where P_i is the probability that a run comprises i contiguous sample points of one intensity [5,6].

$$H = - \sum_{i=1}^{\infty} P_i \log_2 P_i \qquad (7-1)$$

It is not necessary to distinguish between black and white runs in (7-1) since these alternate. It follows immediately [5,6] that if the run-lengths are coded into binary digit sequences, then H is the average number of digits required per run, and H/E(L) where

$$E(L) = \sum_{i=1}^{\infty} i \; P_i \qquad (7-2)$$

is the required number of bits/pel. Existing evidence indicates that H/E(L) ≃ 0.10 to 0.25 depending on the actual data source [2].

Huffman codes, whose sequence lengths vary inversely with the probability of the run-lengths, virtually achieve the minimum number of bits/pel. However, implementation of variable-length sequences is often inconvenient, and channel bit-errors can cause intolerable loss of synchronization at the receiver.

Other schemes exist for coding run-lengths, and these are almost as efficient as Huffman codes while being easier to implement and easily synchronized. We mention only two of these.

The first involves a so-called B code [2], in which a run is represented by a block of binary digits consisting of sub-blocks each of which begins with either a one or zero, depending on whether or not the previous sub-block belonged to the same main block. The redundancy, defined as

$$\eta = \frac{r}{H/E(L)} - 1 \qquad (7\text{-}3)$$

was observed to be approximately 0.30 when averaged over 30 FAX images [2]. In (7-3) r denotes the actual number of bits/pel.

The white-block-skipping (WBS) code [3] involved transmitting a 0 for a block of N white pels, and a 1 followed by the actual bit pattern for an N-pel block containing at least one B pel. An algorithm for optimizing N was developed; the optimum value was $N \simeq 8$. To more efficiently transmit textual documents containing lines with no B pels, the scheme was modified and used on all documents. A line containing no B pels was indicated by 0. In any line with at least one B pel, a 1 preceeded the regular WBS code. For a 26-line page of double-spaced text; 0.18 bits/pel were needed; for a business letter with five lines of text, a heading and a signature, the rate was 0.10 bits/pel; weather maps required approximately 0.40

bits/pel; a circuit diagram used 0.25 bits/pel.

Channel transmission errors are potentially very damaging to run-length coding schemes, since run-length errors are cumulative on any line. This problem can be overcome by using a buffer at both transmitter and receiver. If a line is received incorrectly as determined by checking the total number of received bits in the line, it can either be retransmitted or merely replaced by the line above.

Line-replacement for error correction uses the image's vertical redundancy, which is ignored during coding, rather effectively. With line-replacement, bit-error probabilities as low as 10^{-3}, which implies one line in four incorrect, can be tolerated in some situations [2]. A 10^{-4} bit-error probability would be acceptable for most FAX applications. Burst errors are normally less harmful than random errors, since bursts affect fewer lines than do an equal number of random errors.

References dealing with run-length coding on a more abstract and generalized basis are available [7-9].

VII-4 Continuous Tone Images on Bilevel Displays

It is often desirable for economic reasons, and sometimes necessary to quantize a continuous-tone monochrome image into two brightness levels prior to its being coded, stored or transmitted. Recent developments in display device technology [10] motivate binary quantization.

Studies [11-15] indicate the feasibility of using various techniques, including ordered dither [11,12] and edge enhancement [13-15] to adaptively quantize each pel. A comparison [15] of four successful techniques which are relatively easy to implement indicates that ordered dither in combination with edge emphasis produces the perceptually most pleasing reconstruction. The image quality appears comparable to that of a medium-quality Xerox copy of a multitone monochrome image [15].

A related problem faced by the publishing industry involves display of multitone monochrome images using halftones [16-18]. Recent efforts here have involved improvement in perceptual quality by appropriate selection of dot size, shape and location in the region assigned to each pel [18]. The feasibility of generating color halftone images has been demonstrated [19].

VII-5 Direct Coding of Text

If text or other FAX data is spatially quantized to 1000x1000 points and then run-length coded, a total of $(10^6)r$ bits/page is needed to store or transmit the document, where r is the number of bits/pel. Even if r = 0.18 for a page of text [3] or 0.10 for a letter [3], approximately $1.8x10^5$ bits/page are required, respectively.

A typical 8½"x11" page of text contains 70 characters per line and 27 lines/page for a total of 1890 characters. A typical business letter of 7 lines might contain a total of 600 characters. If each character were encoded directly by a 6-bit binary number, a double-spaced page of text or a letter would require approximately $1.1x10^4$ bits/page or $3.6x10^3$ bits/page, respectively. The savings for direct coding as compared with run-length coding is 16 for a page of double-spaced text and 28 for a typical letter. Such savings obviously encourage direct coding rather than FAX transmission by organizations wishing to use electronic mail or electronic filing facilities.

In comparing direct coding vs FAX the effects of channel transmission errors must be considered. For FAX transmission, bit-error probabilities $p \approx 10^{-4}$ are of little consequence. However, requirements for direct transmission involve much lower error probabilities, typically 10^{-10} or 10^{-12} [21,22]. For many channels

$$p \approx \exp(-K/R) \qquad (7\text{-}4)$$

where K is the channel's carrier-to-noise ratio and R the bit rate. If K is such that $p \approx 10^{-4}$ for R=Ro, then R=Ro/3 is needed for $p = 10^{-12}$. Therefore, if 6-bit character sequences are to be transmitted with no channel coding, they must be transmitted at one-third the FAX rate, in which case a typical business letter would be transmitted 9 times as fast as a FAX letter, and a page of double-spaced text 5.3 times as fast. In fact, channel coding would be used on encoded characters to bring their bit-error rate from 10^{-4} to 10^{-12}. A triple-error-correcting code [5,23] would be required (perhaps (23,12) Golay code) which would add some cost to the coding/decoding system.

VII-6 Coding Text Segments

Most text, including English prose, is highly structural. Shannon [24] estimated that English prose is 75% redundant, in the sense that knowledge of 25% of the text permits the reader to deduce (with considerable difficulty) the remaining 75%. It is rather easy to demonstrate that text is at least 50% redundant by covering the bottom half of a line of English prose or by deleting every other letter; the information can be obtained from the remaining text.

A straight-line approximation to the probability Q_N of an English word of rank N is [25]

$$Q_N \simeq 0.1/N \qquad (7-5)$$

where words are ranked in order of their probability of occurence. The word length L_N, of the N^{th} ranked word is [25]

$$L_N \simeq 2.45\ N^{0.167} \qquad (7-6)$$

Included in L_N is one space adjacent to the word.

From (7-5) and (7-6) it follows that a 10,000 word vocabulary has an average word length of 5.75 characters.

A vocabulary consisting of all single letters, the 200 most common digrams and trigrams and the 400 most common words was used to encode text [25]. A test on actual text showed that encoding could be done using 3.18 rather than 6-bits/character. Other tests of a similar nature are reported elsewhere [26].

Recently, techniques based on sequential decoding methods have been applied to source coding [27-29]. These are of limited utility because of their complexity and because they code continuous strings rather than blocks of text. The retrieval of a text segment for modification is virtually impossible, since the resulting modification would have to alter not only the desired block of text, but succeeding blocks as well.

More obvious compaction schemes exist and are being implemented [30]. For example, 6-bits/character will encode any date using 54 bits; 12 for the day, 18 for the month, and 24 for the year. To individually encode all dates from zero to 2000 AD (7.3×10^5 days) requires 20 bits. Suppressing repeated characters and avoiding empty file space is another obvious compaction strategy [30].

Compaction saves storage space, transmission time, memory access time and queueing delays, since there are fewer bits to store, transmit, search and queue.

VII-7 Optical Character Recognition

Although a thorough discussion of optical character recognition (OCR) is beyond the scope of our present study, it would seem inappropriate not to at least consider the topic briefly. This we now do.

The first operational step of an OCR system involves positioning the document to be read, followed by locating the first character to be scanned. The entire document must then be scanned and discharged. Although difficult technological problems are involved in these initial steps these problems have been overcome in existing systems. Ultimately, the difficulty lies in recognizing the text.

As mentioned earlier in connection with speech recognition and speaker verification, pattern recognition involves two separate operations; namely, feature extraction followed by decision making [31-35]. Decisions can be on the basis of the observed patterns and either stored statistical data regarding the apriori probabilities of the patterns and conditional distributions of the features, or on syntactical (structural) analysis of the features, or on both statistical and syntactic factors.

An indication of the results obtainable using a simple feature extractor and a statistical classifier is provided by

Hussain, Toussaint and Donaldson [36]. Hussain et al., [36] used as features the amount of black in each of the 25 square regions obtained by dividing the original size-normalized 20x24 binary-valued array of character data points into 4x4 squares. Bayes classification yielded accuracies of 72.5% on Munson's [37] hard-to-read 147 handprinted upper case character sets obtained from Fortran coding sheets. The results compare favourably with other statistical schemes which use more complex features [37,38]. Additional work by the author on typewritten texts showed accuracies of 99% on single type fonts, 95% on a set of three fonts, and 90% on ten fonts.

The syntactic approach, which is faced with the difficulty of segmenting patterns into features (called primitives), was used by Pavlidis and Ali [39]. Segmentation of handwritten numerals was accomplished by approximating each character's boundaries by piecewise-linear segments which were used as primitives. A classification tree examined various primitives at each stage to yield recognition votes of approximately 85% on Munson's [37] handprinted numerals.

Ultimately, recognition schemes will likely use statistical methods to help classify primitives, and a combination of statistical and decision tree methods akin to sequential decoding [40] or Viterbi decoding [41] for decision making. Alternatively, statistical methods might be used to divide characters into classes (for example, A, R and P might constitute one class and D, O and Q another class), with primitives then being used with tree classifiers to make final decisions [42]. In any event, both statistical and syntactic methods will likely be applied to OCR.

VII-8 Conclusions and Summary

In considering various FAX coding schemes, including run-length coding, it is apparent that direct coding of characters requires approximately one-twentieth the number of FAX bits, depending on the source document and the amount of compression used with either technique. This result, together with potential savings in paper, will ultimately encourage use of electronic filing, in conjunction with other electronic information services.

In view of the fact that higher bit-error probability on a data channel implies higher data rate, the argument for designing channels with 10^{-4} bit-error probabilities with error-correcting channel codes being available when required again appears.

Although text compression reduces storage space, transmission delays due to queueing and the time to search and access data, it is not clear how much overhead cost is warranted for text compression.

The possibility of improved OCR accuracy by combination of statistical and syntactic pattern recognition methods is very real, in spite of segmentation difficulties inherent in the syntactic approach. The variety of type fonts useable as well as the possibility of handprinted characters being processed on OCR machines is thereby potentially increased, as is the recognition accuracy.

VII-9 References

1. J. Capon, "A probabilistic model for run-length coding of pictures," IRE Trans. Inform. Theory, vol. IT-5, pp. 157-163, Dec. 1959.

2. H. Meyr, H.G. Rosdosky and T.S. Huang, "Optimum run length codes," IEEE Trans. Commun., vol. COM-22, pp. 826-835, June 1974.

3. T.S. Huang and A.B.S. Hussain, "Facsimilie coding by skipping white," IEEE Trans. Commun., vol. COM-23, pp. 1452-1460, Dec. 1975.

4. T.S. Huang, "Run length coding and its extensions," in Picture Bandwidth Compression, T.S. Huang and O.J. Tretiah, Ed. New York: McGraw-Hill, 1972.

5. R.G. Gallager, Information Theory and Reliable Communication. New York: Wiley, 1968.

6. T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

7. D.C. Van Voorhis, "An extended run-length encoder and decoder for compression of black/white images," IEEE Trans. Inform. Theory, vol. IT-22, pp. 190-199, March 1976.

8. T.S. Huang and H.U. Koller, "Coding of multilevel graphics," IEEE Trans. Commun., vol. COM-23, pp. 598-606, June 1975.

9. T.S. Huang, "Bounds on the bit rate of run length codes," IEEE Trans. Inform. Theory, vol. IT-21, pp. 707-708, Nov. 1975.

10. A. Sobel, T. Henion, L.N. Heymich, H. Kressel, B.J. Lechner, and I. Reingold, "Introduction" - Special Issue on Display Devices and Systems, IEEE Trans. Electron Dev., vol. ED-22, pp. 629-635, Sept. 1975.

11. B. Lippel and M. Kurland, "The effect of dither on luminance quantization of pictures," IEEE Trans. Commun. Technol., vol. COM-19, pp. 879-888, Dec. 1971.

12. J.O. Limb, "Design of dither waveforms for quantized visual signals," Bell Syst. Tech. J., vol. 48, pp. 2555-2582, Sept. 1969.

13. J.F. Jarvis and C.S. Roberts, "A new technique for displaying continuous tone images on a bilevel display," IEEE Trans. Commun., vol. COM-24, pp. 891-898, Aug. 1976.

14. T.H. Morrin, "A black-white representation of a grey scale picture," IEEE Trans. Computers, vol. C-23, pp. 184-186, Feb. 1974.

15. J.F. Jarvis, C.N. Judice and W.H. Ninke, "A survey of techniques for the display of continuous tone pictures on bilevel displays," Computer Graphics and Image Processing, vol. 5, pp. 13-40, 1976.

16. T.S. Huang, "Digital transmission of halftone pictures," Computer Graphics and Image Processing, vol. 4, pp. 195-202, June 1974.

17. S. Arnemann and M. Tasto, "Generating halftone pictures on graphic terminals using run length coding," Computer Graphics and Image Processing, vol. 2, pp. 1-11, March 1973.

18. R.L. Gard, "Digital picture processing for the publishing industry," Computer Graphics and Image Processing, vol. 5, pp. 151-170, June 1976.

19. C.H. Hertz and T. Orhang, "The ink jet plotter: A computer peripheral for producing hard copy imagery," Computer Graphics and Image Processing, vol. 5, pp. 1-12, March 1976.

20. K. Knowlton and L. Harmon, "Computer-produced grey scales," Computer Graphics and Image Processing, vol. 1, pp. 1-20, 1972.

21. W.J. Miller, "Technology assessment for electronic mail handling," in Conf. Proc., Int. Conf. Commun., Philadelphia, Penn., 1976, pp. (15-10) - (15-12).

22. M.A. Robbins, "Error objectives in the electronic mail system," in Conf. Rec., Int. Conf. Commun., Philadelphia, Penn., June 1976, pp. (15-7) - (15-9).

23. S. Lin, An Introduction to Error Correcting Codes. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

24. C.E. Shannon, "Prediction and entropy of printed English," Bell Syst. Tech. J., vol. 30, pp. 50-64, Jan. 1951.

25. H.W. White, "Printed English text compression," Proc. IEEE., vol. 55, pp. 390-396, March 1967.

26. F. Rubin, "Experiments in text file compression," Commun. ACM., vol. 19, pp. 617-623, Nov. 1976.

27. T.C. Ancheta, "Syndrome source coding and its universal generalization," IEEE Trans. Inform. Theory, vol. IT-22, pp. 432-436, July 1976.

28. F. Jelinek, "Tree coding of memory-less time discrete sources with a fidelity criterion," IEEE Trans. Inform. Theory, vol. IT-15, pp. 584-590, Sept. 1969.

29. M.E. Hellman, "Convolutional source encoding," IEEE Trans. Inform. Theory, vol. IT-21, pp. 651-656, Nov. 1975.

30. J. Martin, Computer Data Base Organization. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

31. K.S. Fu and A. Rosenfeld, "Pattern Recognition and Image Processing," IEEE Trans. Computers, vol. C-25, pp. 1336-1345, Dec. 1976.

32. K. Fukunega, Introduction to Statistical Pattern Recognition. New York: Academic Press, 1972.

33. K.S. Fu, Ed., Applications of Syntactic Pattern Recognition. New York: Springer, 1976.

34. K.S. Fu, Sequential Methods in Pattern Recognition and Machine Learning. New York: Academic Press, 1968.

35. L. Kanal, "Patterns in pattern recognition: 1968-1974," IEEE Trans. Inform. Theory, vol. IT-20, Nov. 1974.

36. A.B.S. Hussain, G.T. Toussaint and R.W. Donaldson, "Results obtained using a simple character recognition procedure on Munson's handprinted data," IEEE Trans. Computers, vol. C-21, pp. 201-205, Feb. 1972.

37. J.H. Munson, "Experiments in the recognition of handprinted text," in Conf. Proc., AFIPS Fall Joint Comp. Conf., Washington, D.C.: Thompson, 1968, pp. 1125-1138.

38. A.L. Knoll, "Experiments with characteristic loci for recognition of handprinted characters," IEEE Trans. Electron. Comput., vol. EC-18, pp. 366-372, Apr. 1969.

39. T. Pavlidis and F. Ali, "Computer recognition of handwritten numerals by polygonal approximation," IEEE Trans. Syst. Man. and Cybernetics, vol. SMC-5, pp. 610-614, Nov. 1975.

40. J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering. New York: Wiley, 1965.

41. G.D. Forney, "The Viterbi algorithm," Proc. IEEE, vol. 61, pp. 268-278, Mar. 1973.

42. R.W. Donaldson and G.T. Toussaint, "Use of contextual constraints in recognition of contour-traced handprinted characters," IEEE Trans. Computers, vol. C-19, pp. 1096-1099, Nov. 1970.

VIII DESIGN AND EVALUATION OF USER TERMINALS

VIII-1 Problem Definition

The interface between an electronic information system and the user is an extremely important determinant of overall system utility. Much remains to be learned regarding the analysis, design and evaluation of interface terminals.

In considering terminals it is probably most natural to do so on three levels, as follows:

1. The physical level, which involves the actual hardware and software facilities for conversion of human sensory stimuli to and from electrical signals.

2. The symbol or language level, which deals with the textual, graphic or verbal symbols used to construct requests for information as well as information displays.

3. The concept level, which involves the communication of ideas between the system and its users, or between geographically separated users.

For our purposes, it is convenient to discuss terminals in terms of user considerations, hardware, and software. Actually, the close relationship of the user to the interface terminal limits the utility of any topical subdivision.

We emphasize that our discussions are brief and incomplete, and are intended to give direction to further studies, the need for which is clearly indicated.

VIII-2 User Considerations

How should a terminal be designed to best meet user requirements? What are the users' requirements? We really don't know.

We are aware of some user requirements, and know how to meet them. For example, when a terminal is used interactively, in the sense that a user-system dialogue occurs, a response to a user's comment should occur within two seconds [1-3]. In fact, a user-controlled response time would be desirable. In situations requiring minimal mental effort from the user, responses should be rapid (less than 2 sec.). When the user must do a lot of thinking, a slower response is desirable, in order that the user not feel rushed. The two-second response rule is waived only following task closure, which concludes a dialogue related to a specific task. Closure occurs, for example, following the dialling of a telephone number or the completion of a series of related questions. Response behaviour expected of terminals is not much different than that expected of other humans [1,4-9].

Short response times during dialogue are required because of the limited capacity of human short-term memory, which is capable of holding approximately seven items [10]. Unless these items are integrated into long-term memory by rehearsal or cognition they are lost [11-13].

Limitations of short-term memory also explain why English prose presented letter-by-letter at a user-controlled rate is read at approximately 20 words per minute (wpm) whereas word-by-word presentations yield reading rates of 110 wpm. In reading a page of text, 300 wpm is normal. Evidently, humans prefer to read by organizing text into cognitive "chunks". Such organization occurs whether or not visual, auditory, tactile, or kinesthetic sensory stimuli are used to represent letters. Readers quickly learn to recognize words as units rather than as individual letters [14].

Blind persons seeking high reading rates use Braille II which contains many contractions [15]. Braille I, in which each letter is represented by a unique pattern of raised dots in a 3 x 2 matrix permits reading rates of 50 wpm; Braille II rates can exceed 90 wpm, particularly if several fingers are used to read in word-chunks.

Humans' ability to identify sensory stimuli is relevant to terminal design. When Pollack [16-18] asked listeners to identify pure tones which differed only in frequency, the maximum amount of information transmitted was about 2.5 bits per stimulus, independent of whether there were 6, 7, 8, 9, 11, or 14 different tones. This corresponds to about 5 perfectly identifiable stimuli. Expansion of the range of frequencies used, as well as several hours of practise, failed to change these results to any great extent. In a similar experiment in which the tones differed only in intensity, Garner [19] found that the maximum amount of information transmitted was about 2.2 bits per stimulus.

By increasing the number of dimensions along which the tones could vary, Pollack was able to increase the information transmitted. When 5 co-ordinates along each of the dimensions of frequency, intensity, duration, direction, repetition rate, and per cent time on were used, subjects received 7.2 of possible 13.9 bits per stimulus. Use of three co-ordinates along each of these dimensions resulted in a transfer of 6.7 of a possible 9.5 bits per stimulus. Use of only two co-ordinates per dimension resulted in a transfer of 5.2 of a possible 6 bits per stimulus. Addition of the two dimensions of frequency and intensity of noise to the above six resulted in a transfer of 6.9 of a possible 8 bits per stimulus. Pollack concluded that the way to reduce the information bottleneck in an absolute judgement situation is to make stimuli multidimensional, and use only a few co-ordinates along each dimension.

Experiments by others support this conclusion.

Andersen and Fitts [20] found that the information transmitted by sequences of different colored numerals was greater than that transmitted by sequences of different numerals of the same color, or by sequences of one numeral of different colors.

Ericsen [21] found that a redundant use of dimensions can increase the amount of information transmitted in an absolute judgement situation. Identification of 20 squares which varied in size, color, and hue resulted in a transfer of 2.8, 3.0 and 2.3 bits per stimulus, respecively, by his subjects. Perfectly correlated variations in size and hue, size and brightness, and hue and brightness resulted in a transfer of 3.6, 3.0 and 3.7 bits per stimulus, respectively. Perfectly correlated variations along all three dimensions resulted in a transmission of 4.1 of a possible 4.5 bits per stimulus. In his review, Miller [10] describes several experiments where increases in information transmitted in absolute judgement situations has been increased by enlarging the dimensionality of the stimulus ensemble. It has been stated [1,8] that for error-free recognition by normal individuals, the number of co-ordinates associated with various visual dimensions should be limited as follows: color - 6; geometric shape - 10; line-width - 2; line type - 5; intensities - 2.

More recently it has been shown that the time required for viewers to select items from a field is reduced if those to be selected are of a color different from those not to be selected [22-24].

Alternatively, if those items to be selected blink at 3 Hz or if all except those to be selected blink, selection is again facilitated [25].

All of these results apply to display design. For example, consider the display of the following file:

EMPLOYEE NAME	AGE	SEX	JOB CODE	DEPT	YEARS WITH COMPANY	ANNUAL SALARY
A B CUTHBERT	50	M	5	2	5	30
D E FINCH	40	F	10	13	20	40
.						
.						
.						
X Y ZEFFERT	30	F	15	7	10	20

Color could be used to delineate the appropriate lines following the question, "Which employees of more than 5 years standing have an annual salary of less than $20,000.00?" Use of blinking would further delineate the appropriate items indicated by the subsequent enquiry "Which of these employees exceed age 40?"

As noted in Chapter 6, distortions in reconstructed speech and images produce effects which, for equal amounts of noise power, are subjectively dissimilar [26-32]. Subjective evaluation procedures are now well established although often tedious to administer, particularly when a large number of variables are involved [26-33].

A large body of psychophysical data is available to quide designers of user terminals. However, a coherent and convenient description of how the various factors combine to affect user utility in some easily defined and measurable way is unavailable.

We conclude this section by noting that tasks should be assigned to users and machines in an optimal way, with due regard for task continuity [2,5,6]. Humans are best suited for formulation of goals, use of intuition and dealing with exceptions. Machines are best for remembering details and performing repetitive calculations with great rapidity and accuracy. Optimal assignment of tasks facilitates a symbosis where user and machine work together to exploit the capabilities of the other [34].

VIII-3 Hardware

Terminal hardware provides the physical medium whereby human sensory information is converted to and from electrical signals. As noted in Section 2-2, terminals normally include a keyboard, visual display, hardcopier, buffer storage and control logic. Thus, terminal inputs are normally via the human tactile and kinesthetic senses, and output is visual.

In human-human communication speech is often used. Recent advances in speech processing as detailed in Chapter 5 indicate that terminal designers should begin to consider inclusion of voice input and output facilities. Hardware for voice communication is relatively inexpensive, particularly insofar as maintenance is concerned, since there are no moving parts. A microphone, preamplifier and analog-to-digital converter is required for speech input, and a digital-to-analog converter, audio amplifier and speaker for speech output. However, additional hardware is needed for effecient coding of speech, and software is required if recognition of words or verification of speakers is to occur at the terminal itself. Normally, recognition and verification would occur at a central cite, in which case software costs would be shared among many terminals.

Voice input and output is convenient because it is familiar to the user, it leaves his hands free for other tasks, it permits him physical mobility, and allows for use of an easily expanded vocabulary.

Unless microphone shielding were available, voice input to terminals would be difficult in noisy environments. Unless earphones were available, voice output would likely be disruptive if several workers were within hearing range of the terminal.

Considerable effort has been devoted to development of visual displays [35-39]. It is probably true that for most applications the CRT is still best in terms of overall speed, spatial resolution, dynamic intensity range, linearity, noise characteristics and spatial response, in spite of its size, fragility, high voltage and high cost. Furthermore the CRT is still being improved [40]. However, flat plasma panels [41] promise to become more suitable for some applications. Light emitting diodes (LED's) [42] and liquid crystals [43] are of interest in displaying text and digits and are used in electronic watches and calculators. Some novel visual displays have been proposed, including one where a matrix of tiny half-black/half-white balls are magnetically rotated to produce a black-white image [44]. Extension to color is possible, at least in principle.

Current problems in plasma panels, which consist of a matrix of individual cells filled with ionizable gas, include cell addressing, luminance and power conversion efficiences, and dynamic range. LED's suffer from high power comsumption and lack of yield uniformity. As noted in Section 7-4, some (apparent) dynamic range improvement is realizable by quantizer threshold adjustment during scanning for conversion of continuous-tone images to two levels [32]. Perhaps these techniques could be extended to multi-level images. The same kind of quantizer threshold adjustments can be used to reduce defects in solid-state image sensors [45]. In displaying multitone images on bilevel displays, half-tone techniques [46,47] in which a pel's intensity is represented by the number of ON cells in the pel's region of definition can also be used.

High resolution scanning of images currently requires specialized and rather expensive hardware [48] such as flying spot scanners, vidicons or image dissectors. FAX and OCR scanners are also relatively expensive.

There is considerable interest in adapting existing terminals to applications for which they were not specifically designed. For example, an inexpensive color display was required by the author for on-line viewing of color images. The problem was solved by interfacing a standard color television receiver to a NOVA 840 minicomputer. The interface consisted of an LSI 256 x 256 matrix of 8-bits intensity for each of three color components (either RGB or YIQ - see Section 6-5). The matrix is randomly accessible from the computer and raster-scanned by the television monitor.

It is appropriate now to design terminals which involve a variety of input and output sensory nodes. Some attempts in this area are already in evidence. Thus, existing interactive graphics terminals include as inputs joysticks, track-balls, light pens, and electronic writing tablets [1,2,49,50]. Discussions regarding the modification of cable television systems for limited two-way use is of continuing interest [51-54]. Transmission from the user's premises would likely involve low-speed keyboard generated data.

VIII-4 Software

Terminal software is needed for assembly, storage, retrieval and updating of data as expressed by symbol sequences used to transfer information between the physical terminal and the user. Software is also needed to facilitate transmission of data to and reception from another location [55,56]. As noted in Chapter 4, software costs continue to increase relative to other system costs. Disciplined software engineering, including use of program modularity, is essential for cost control.

In those situations where maintenance of a two-second response time is problematical, text compaction [57,58] as described in Section 7-6 may be useful, along with data storage procedures for minimization of search time. One approach is to search items in order of their probability of usage. It has been suggested that 80% of the accesses often involve 20% of the data [59]. A variety of other approaches are available to reduce access time [60-63]

In addition to reducing response time, data compaction and rapid access techniques minimize storage costs by permitting more efficient use of high speed storage which is always at a premium. Data compaction also reduces queueing and transmission delays which, together with accessing and processing of data accounts for the time delay between a request for information and a response. Minimization of the processing time involves careful design and coding of the processing algorithm.

Extensive software is needed to implement man-machine dialogues which are often classified either as user- or machine-initiated [1,5]. User-initiated dialogues involve specialized languages and trained operators. Dialogue mnemonics are often used to save time, storage and data channel bandwidth. A typical example is an airline reservation system. An operator might request details regarding A.P. Jones' current reservations by keyboard entry of the symbols shown in Fig. 8-1.

Fig. 8-2 shows a typical file display. The operator could modify the file or request additional information, including details regarding in-flight meals and entertainment, as well as time and locations of any stop-overs. The dialogue is quick and unambiguous but is useful only when it is cost-effective to train terminal operators.

Computer-initiated dialogues often use menu-selection, which is easily used but slow. Fig. 8-3 illustrates a sequence of displays which might appear following a request for the numbers and corresponding titles of university courses. Following the user's selection of computer science in Fig. 8-3(a), the list in Fig. 8-3(b) appears. Selection of an item from this second list would result in the display of the appropriate courses as well as their meeting time and location.

A variety of languages are available for use with terminals [64-73]. FORTRAN (Formula-Translation) [67] is convenient for programming numerical computations, COBOL (Computer Oriented

```
*                 163        16 DEC        JONES, A P          ENTER
File display      Flight     Today's       Passenger's         Depression of ENTER key
command           number     date          name                results in execution of
                                                               file display command
```

Fig. 8-1 Request for file

```
JONES, A P

1.        105  E  (1)     03 JAN     VYR      9:30A      5:15P

2.         21  F  (2)     07 JAN     JFK      7:05P      8:55P
```

- Flight sequence number
- Flight number
- Travel class
- Number of seats reserved
- Date of departure
- Destination
- Takeoff time
- Arrival time

Fig. 8-2 Displayed file

AREA OF INTEREST

1. Anthropology
2. Architecture
3. Asian Studies

.
.
.
.

11. Computer Science

.
.
.
.

37. Management Science

38. Mathematics

.
.
.
.
.
.
.
.
.
.
.

74. Zoology

(a)

COMPUTER SCIENCE SUB-AREAS

1. Advanced Programming
2. Artificial Intelligence
3. Assembly Language Programming
4. Automata Theory
5. Combinatonics
6. Computer Architecture
7. Data Structures
8. Graphics
9. Introductory Programming
10. Modelling and Simulation
11. Numerical Computation
12. Programming Languages
13. Systems Programming

(b)

Fig. 8-3 Menu-Selection example

Business Language) [68] for manipulating business data, and BASIC [64] for on-line arithmetic calculations. Many special-purpose languages are available, including WYLBUR [69] for text-editing (see Section 2-2) and GLUE [70] for on-line psychological experiments.

Each language serves at least one specific need and requires some user training. Experience with one language facilitates the learning of a second. It is not unreasonable to require a user to invest some time in learning a language which will enable his effective communication with an electronic information system. Operation of an automobile or a programmable calculator requires some training, as does development of skills for inter-personal communication.

Most existing software provides for keyboard-based man-machine dialogues in the form of text. Future software will likely be needed to co-ordinate multi-modality inputs and outputs.

As indicated earlier, the really difficult software problems involve data-base organization and management as well as manipulation and synthesis of images. Software is needed to facilitate responses to questions not previously anticipated in designing the data base. Worthy of mention here is the English query language SEQUEL [71] which handles queries against relational data bases. The question "How many employees earn more than their department managers?" can be answered by accessing a data base of the type described in Section 8-2. Better understanding of human information processing capabilities, limitations, and methods would be useful in dealing with data-base technology and image synthesis.

A more immediate objective is the development of modular software usable with a variety of terminals and applications.

Software may be located either at user terminals or another point in the computer communications network. As noted in Chapter 3, this matter of location warrants study at this time.

VIII-5 Conclusions and Suggestions for Further Study

Design and evaluation of user terminals requires an understanding of human information processing, electronics, software, engineering, speech processing, image processing, computer technology and EIS applications. The need for an integrated understanding of these diverse fields often frustrates terminal designers.

The basic difficulty, however, is the absence of a well-defined design methodology. Consider a digital speech transmission system which involves many design variables including speech bandwidth (often pre-specified), number and arrangement of quantizer levels, sampling rate, and coding strategy. A specific design objective can be articulated: maximize the output signal-to-noise ratio (SNR) for a given data rate or, equivalently, minimize the data rate subject to maintaining a given SNR. Although maximum SNR does not guarantee maximum subjective quality, the design criterion is reasonable. Subjective tests can be used to finalize selection of system parameters.

There is no quantitative performance criteria for terminals. Performance goals would probably depend to some extent on applications whose relationship to terminals is only partially understood. What is to be done?

In the absence of objective performance measures, one might proceed as follows. First, delineate rather general criteria against which terminals can be assessed. Second, weight these in accordance with the application(s) and user(s). Third, specify additional suitably weighted criteria specific to the application(s) and user(s).

The following general criteria are suggested, along with the admonition that additional study is needed to confirm these or alternative criteria as good choices.

1. Ease of use: How much user training is needed? Does the terminal adapt to user familiarity by permitting language contractions? Is the terminal easy to communicate with?

2. Versatility: Is the terminal easily adapted to other users or other applications? Are new features easily added? How many different applications will the terminal serve?

3. Cost: Here we include capital cost, rental cost, cost of modifications or additions for new situations, maintenance costs, reliability, durability, obsolescence, and efficiency in terms of overall performance.

4. User enjoyment/acceptance: Does the terminal fatigue the operator unnecessarily? Is the operator's interest maintained? Ultimately, a terminal which is enjoyable to use will be used, probably with reasonable effectiveness.

5. Environmental compatability: Is the terminal's physical size acceptable? Is it too noisy? Are special facilities such as low ambient lighting necessary?

The above criteria are not mutually exclusive. Whatever criteria are used, there remains the problem of quantifying and measuring utility.

A final comment on versatility is warranted. Since we aren't really certain how to optimize terminals, they should probably possess some flexibility which permits the user to adjust parameters and select features. Those features and parameter values chosen might ultimately lead to optimization of terminals from a user's point of view.

VIII-6 References

1. J. Martin, Design of Man-Computer Dialogues. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

2. W.B. Rouse, "Design of man-computer interfaces for on-line interactive systems," Proc. IEEE, vol. 63, pp. 817-857, June 1975.

3. R.B. Miller, "Response time in man-computer conversational transactions," in AFIPS Conference Proceedings, Fall Joint Computer Conference, 1968. Washington, D.C.: Thompson Book Co., 1968.

4. D.W. Davies, C.R. Evans and D.M. Yates, "Human factors in interactive teleprocessing systems," in Conf. Rec., Second Int. Conf. Comp. Commun., Stockholm, Sweden, Aug. 1974, pp. 491-496.

5. T.C.S. Kennedy, "The design of interactive procedures for man-machine communication," Int. J. Syst. Science, vol. 6, pp. 309-334, May 1974.

6. D.L. Hebditch,"Terminal systems for real people," in Conf. Rec., Second, Int. Conf. Comp. Commun., Stockholm, Sweden, Aug. 1974, pp. 497-504.

7. H. Sackman, "Experimental analysis of man-computer problem solving," Human Factors, vol. 12, pp. 187-201, 1970.

8. J.D. Foley and V.L. Wallace, "The art of graphics in man-machine conversation," Proc. IEEE, vol. 62, pp. 462-471, April 1974.

9. G.D. Hornbuckle, "The computer graphics user/machine interface," IEEE Trans. on Hum. Factors Electron., vol. HFE-8, pp. 17-20, Mar. 1967.

10. G.A. Miller, "The magical number seven plus or minus two: some limits on our capacity for processing information," Psychol. Rev. vol. 63, pp. 81-97, 1956.

11. P.H. Lindsay and D.A. Norman, Human Information Processing. New York: Academic Press, 1972.

12. V. Neisser, Cognitive Psychology. New York: Appleton, 1967.

13. W. Kintsh, Learning, Memory and Conceptual Processes. New York: Wiley, 1970.

14. R.W. Donaldson, "Communication via the auditory, tactile and kinesthetic senses," IEEE Trans. Inform. Theory, vol. IT-13, pp. 11-21, Jan. 1967.

15. Zahl, Blindness. Princeton, N.J.: Princeton Univ. Press, 1950.

16. I. Pollack, "The information of elementary auditory displays," J. Acoust. Soc. Amer., vol. 24, pp. 745-749, 1952.

17. I. Pollack, "The information of multidimensional auditory displays II, J. Acoust. Soc. Amer., vol. 25, pp. 765-769, 1953.

18. I. Pollack and L. Fitts, "Information of multidimensional auditory displays," J. Acoust. Soc. Amer., vol. 26, pp. 155-158, 1954.

19. W.R. Garner, "An informational analysis of the absolute judgements of loudness," J. Exptl. Psych.,vol. 46, pp. 373-380, 1954.

20. N.S. Andersen and P.S. Fitts, "Amount of information gained during brief exposures to numerals and colors," J. Exptl. Psych., vol. 56, pp. 362-369, 1958.

21. C.W. Ericsen, "Multidimensional differences and accuracy of discrimination," USAF: WADC Tech. Rept., pp. 54-165, 1954.

22. M.C. Cahill and R.C. Carter, Jr., "Color code size for searching displays of different density," Human Factors, vol. 18, pp. 273-280, June 1976.

23. S.L. Smith, "Color coding and visual separability in information displays," J. Applied Psych., vol. 47, pp. 358-364, 1963.

24. S.L. Smith, "Display color coding for a visual search task," J. Exptl. Psych., vol. 64, pp. 434-460, 1962.

25. S.L. Smith and N.C. Goodwin, "Blink coding for information display," Human Factors, vol. 13, pp. 783-790, 1971.

26. T.S. Huang and M. Chikhaoni, "The effect of BSC channel errors on PCM picture quality," IEEE Trans. Inform. Theory, vol. IT-13, pp. 270-273, Apr. 1967.

27. J. Yan and R.W. Donaldson, "Subjective effects of channel transmission errors on PCM and DPCM voice communication systems," IEEE Trans. Commun., vol. COM-20, pp. 281-290, June 1972.

28. J.L. Mannos and D.J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," IEEE Trans. Inform. Theory, vol. IT-20, pp. 525-536, July 1974.

29. J.A. Gardner and S.M. Soliday, "Human factors in digital telecommunications," Human Factors, vol. 16, pp. 146-153, Feb. 1974.

30. S.M. Soliday and J.M. Gardner, "Picture quality judgements in digital television systems," Human Factors, vol. 16, pp. 139-145, Apr. 1974.

31. T.S. Huang, "The subjective effect of two-dimensional pictoral noise," IEEE Trans. on Inform. Theory, vol. IT-11, pp. 43-53, Jan. 1965.

32. J.F. Jarvis, C.N. Jundice and W.H. Ninke, "A survey of techniques for the display of continuous-tone images on bilevel displays," Computer Graphics and Image Processing, vol. 5, pp. 13-40, Mar. 1976.

33. J.P. Guilford, Psychometric Methods. New York: McGraw-Hill, 1954.

34. J.C.R. Licklider, "Man-computer symbiosis," IRE Trans. Human Factors Electron., vol. HFE-1, pp. 4-11, Mar. 1960.

35. A. Sobel, T. Henion, L.N. Heynick, B.J. Lechner, and I. Reingold, "Introduction (to the special issue on display devices)," IEEE Trans. Electron. Dev., vol. ED-22, pp. 629-635, Sept. 1975.

36. Proc. IEEE (Special Issue on New Materials for Display Devices), vol. 61, pp. 801-1056, 1973.

37. IEEE Trans. Electron. Dev., (Special Issue on Display Devices), vol. ED-20, Nov. 1973.

38. W.M. Newman, "Trends in graphic display design," IEEE Trans. Computers, vol. C-25, pp. 1321-1326, Dec. 1976.

39. W.M. Newman and R.F. Sproull, Principles of Interactive Computer Graphics. New York: McGraw-Hill, 1973.

40. A.M. Morell, H.B. Law, E.G. Ramberg and E.W. Herald, Advances in Image Pickup and Display: Color Television Picture Tubes (Supplement I), New York: Academic Press, 1974.

41. J. Smith, "Experimental storage display panels using DC gas discharges without resistors," IEEE Trans. Electron. Dev., vol. ED-22, pp. 642-649, Sept. 1975.

42. R.N. Bhargava, "Recent advances in visible LED's," IEEE Trans. Electron. Dev., vol. ED-22, pp. 691-701, Sept. 1975.

43. K. Nakada, T. Ishibashi, and K. Toriyama, "A design of multiplexing liquid-crystal display for calculators," IEEE Trans. Electron. Dev., vol. ED-22, pp. 725-729, Sept. 1975.

44. L.L. Lee, "A magnetic-particles display," IEEE Trans. Electron. Dev., vol. ED-22, pp. 758-765, Sept. 1975.

45. P. Stucki, "Solid-state area scanning arrays as interface devices between optical and digital computing systems," IEEE Trans. Computers, vol. C-24, pp. 370-380, Apr. 1975.

46. S. Arnemann and M. Tasto, "Generating halftone pictures on graphic terminals using run length coding," Comp. Graphics and Image Proc., vol. 2, pp. 1-11, Mar. 1973.

47. C.H. Hertz and T. Orhaug, "The ink jet plotter: A computer peripheral for producing hard copy imagery," Comp. Graphics and Image Proc., vol. 5, pp. 1-12, Mar. 1976.

48. R.T. Chien and W.E. Snyder, "Hardware for visual image processing," IEEE Trans. Ccts. Syst., vol. CAS-22, pp. 541-551, June 1975.

49. G.J. Richie and J.A. Turner, "Input devices for interactive graphics," Int. J. Man-Mach. Studies, vol. 7, pp. 639-660, Sept. 1975.

50. I.E. Sutherland, "Three-dimensional data input by tablet," Proc. IEEE, vol. 62, pp. 453-461, Apr. 1974.

51. P. McManamon, "Technical implications of teleconference service," IEEE Trans. Commun., vol. COM-23, pp. 30-38, Jan. 1975.

52. P. Baran, "Broadband interactive services to the home: Part I - potential market demand," IEEE Trans. Commun., vol. COM-23, pp. 5-15, Jan. 1975.

53. P. Baran, "Broadband interactive services to the home: Part II - impasse," vol. COM-23, pp. 178-184, Jan. 1975.

54. I.T. Frisch, "Technical Problems in nationwide networking and interconnection," IEEE Trans. Commun., vol. COM-23, pp. 78-88, Jan. 1975.

55. J. Martin, <u>Systems Analysis for Data Transmission</u>. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

56. P.G. Neumann, "System design for computer networks," in <u>Computer Communication Networks</u>, N. Abramson and F. Kuo, Ed., Englewood Cliffs, N.J.: Prentice-Hall, 1973, pp. 29-81.

57. H.E. White, "Printed English compression by dictionary encoding," Proc. IEEE, vol. 55, pp. 390-396, Mar. 1967.

58. F. Rubin, "Experiments in text file compression," Commun. ACM, vol. 19, pp. 617-624, Nov. 1976.

59. I.J. Barton, S.E. Creasy, M.F. Lynch and M.J. Snell, "An information-theoretic approach to text searching in direct access systems," Commun. ACM, vol. 17, pp. 345-350, June 1974.

60. D. Severance and R. Duhne, "A practitioner's guide to addressing algorithms," Commun. ACM, vol. 19, pp. 314-327, June 1976.

61. W.A. Burkhard and R.M. Keller, "Some approaches to best match file searching," Commun. ACM, vol. 16, pp. 230-236, Apr. 1973.

62. R.A. Wagner, "Common phases and minimum-space text storage," Commun. ACM, vol. 16, pp. 148-152, Mar. 1973.

63. M.C. Harrison, "Inplementation of the substring test by hashing," Commun. ACM, vol. 14, pp. 777-779, Dec. 1971.

64. P.M. Sherman, Programming and Coding Digital Computers, New York: Wiley, 1963.

65. S.S. Kuo, Assembler Language for FORTRAN, COBOL and PL/I Programmers, Reading, Mass.: Addison Wesley, 1974.

66. P. Wagner, "Computer languages - the first 25 years," IEEE Trans. Computers, vol. C-25, pp. 1207-1226, Dec. 1976.

67. D.D. McCracken, A Guide to FORTRAN IV Programming, New York: Wiley, 1965.

68. D.D. McCracken, A Guide to COBOL Programming, New York: Wiley, 1963.

69. R. Fajman and J. Borgett, "WYLBUR: An interactive text editing and remote job entry system," Commun. ACM, vol. 16, pp. 314-322, May 1973.

70. T.R.G. Green and D.J. Guest, "An easily implemented language for computer control of complex experiments," Int. J. Man-Mach. Studies, vol. 6, pp. 335-360, May 1974.

71. M.M. Astraham and D.D. Chamberlin, "Implementation of a structured English query language," Commun. ACM, vol. 18, pp. 580-587, Oct. 1975.

72. P. Naur, "Programming languages, natural languages and mathematics," Commun. ACM, vol. 18, pp. 676-682, Dec. 1975.

73. J.M. Smith and P.Y.T. Chang, "Optimizing the performance of a relational algebra data base interface," Commun. ACM, vol. 18, pp. 568-579, Oct. 1975.