Statistical Characterisation and Stochastic Modelling of 1-Layer Variable Bit Rate H.261 Video Codec Traffic

presented in 1994 to

The School of Electronic Engineering
Dublin City University

by

A. P. MacFhearraigh, BEng

for the award of

Master of Engineering

Supervisors :

Mr. Thomas Curran, PhD,
and
Mr. Dmitri D. Botvich, PhD.

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Engineering is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed : Adrian Mac Fhearraigh

Date : 9th February 1996

Abstract

The Integrated Services Digital Network(ISDN) is under re-design to provide flexibility which will ensure efficient network utilisation in the provision of broadband services. The main broadband services envisaged for provision on the Broadband ISDN(B-ISDN) are : Videophone; Videoconferencing; Television and High Definition TV. The B-ISDN will be a packet switched network where the packets(cells) will be transferred by the Asynchronous Transfer Mode(ATM) concept. Unlike voice and data services, the impact video services will have on the B-ISDN is unknown and hence loss of information is difficult to predict. Present videophone terminals are based on the CCITT H.261 Video Coding standard hence the picture quality is variable because video codec traffic is transmitted at a constant rate. To maintain a constant quality picture the codec output data must be transmitted at a variable rate or alternatively, for constant rate video codecs extra information must be made available to achieve constant picture quality. This latter technique is 2-Layer video coding where the first layer transmits at a constant rate and the second layer at a variable rate. The ATM B-ISDN promises constant picture quality video services, therefore to achieve this aim the impact variable rate video sources will have on the network must be determined by network simulation, thus variable rate video source models must be derived. To statistically characterise and stochastically model 1-Layer VBR(Variable Bit Rate) H.261 Video Codec traffic, here a videophone sequence is analysed by two alternative strategies : Talk-Listen and Motion Level. This analysis also found that 2-Layer H.261 Video Codec traffic can be stochastically modelled via a 1-Layer VBR H.261 Video Codec traffic model. Numerous hierarchical stochastic models with the ability to capture the statistical characteristics of long video sequences, in particular the short-term and long-term autocorrelations, are presented. One such model was simulated and the resulting simulated traffic was analysed to confirm the advantage hierarchical stochastic models have over non-hierarchical stochastic models in modelling video source traffic.

Acknowledgement

I would like to take this opportunity to convey my appreciation to some people with whom I have worked over the past two years. Firstly I must thank both my supervisors, Dr. Tommy Curran and Dr. Dmitri Botvich. I would like to thank Tommy for his patience and understanding during my period of injury and afterwards. I am also thankful to him for giving me the opportunity to undertake this research and for continuously urging me to insert a lot of energy into this work. This paid off. I would like to thank Dmitri for his continuous inspiration and enthusiasm. I found his knowledge of mathematics enlightening. He has been a teacher to me by helping me broaden my knowledge of mathematics, in particular applied statistics and probability. I must also commend Sean Murphy on his hard work in producing sample video coded data without which this postgraduate research would have been impossible. A concerted effort through discussion and ideas by all made this research possible and exciting.

I would also like to thank the numerous friends I have at Dublin City University which range from undergraduates, postgraduates, technicians and lecturers. They have all contributed to my successful duration in the university. I have valued their friendship and will always.

Lastly, I must praise Tommy for his immense effort to create a telecommunications centre of excellence at Dublin City University. He is making great inroads to involve young intelligent Irish telecommunications engineers in European projects which will give Irish research institutions exposure to the world. I applaud his determination.

Matamaitic agus Fisic go Brách.

Table of Contents

1. Introduction 1

2. The Broadband Integrated Services Digital Network 5

2.1 General 5

2.1.1 The Integrated Services Digital Network(ISDN) 5

2.1.2 ISDN Enhancement 8

2.2 Transmission Modes on the B-ISDN 10

2.2.1 General 10

2.2.2 Synchronous Transfer Mode(STM) 11

2.2.3 Asynchronous Transfer Mode(ATM) 11

2.2.3.1 The ATM Packet(Cell) 13

2.2.3.2 ATM Networking 15

2.2.3.3 Error Control Strategies 16

2.2.3.4 Traffic Control 18

2.2.3.5 ATM Call Set Up 20

2.3 Aspects of ATM Quality of Service 21

3. CCITT H.261 Video Coding Standard 24

3.1 General 24

3.2 The H.261 Video Codec 25

3.2.1 The Source Coder 26

3.2.1.1 Source Coder Input Signal Format 26

3.2.1.2 The Transformer 28

3.2.1.3 The Quantiser 28

3.2.1.4 Picture Memory 29

3.2.1.5 Loop Filter 29

3.2.1.6 Coding Control 30

3.2.1.7 RM8(Reference Model 8) 30

3.2.2 Video Multiplex Coder 31

3.2.3 Transmission Buffer 33

3.2.4 Transmission Coder 34

3.3 Adaptation of the CCITT H.261 Video Coding Standard for B-ISDN ATM 34

4. Video Source Analysis 39

4.1 Video Sequence Pre-Video Coding Analysis 39

4.1.1 General39

4.1.2 Videophone Sources 40

4.1.3 Videoconference Sources 41

4.1.4 Television Sources 42

4.2 Video Sequence Post-Video Coding Analysis 44

4.2.1 The Source 45

4.2.2 Examples of Video Coded Data 47

4.2.3 The Talk-Listen Strategy 50

4.2.3.1 Talking Sequence Analysis 50

4.2.3.1.1 The Talking Duration 50

4.2.3.1.2 The Bits per Talking Frame 54

4.2.3.2 Listening Sequence Analysis 57

4.2.3.2.1 The Listening Duration57

4.2.3.2.2 The Bits per Listening Frame 60

4.2.3.3 Synopsis 63

4.2.4 The Motion Level Strategy 64

4.2.4.1 High Motion Sequence Analysis 66

4.2.4.1.1 High Motion Duration 66

4.2.4.1.2 The Bits per High Motion Frame 68

4.2.4.2 Low Motion Sequence Analysis 71

4.2.4.2.1 Low Motion Duration 72

4.2.4.2.2 The Bits per Low Motion Frame 73

4.2.4.3 Synopsis 76

4.3 The Long-term Autocorrelation of 1-Layer H.261 Video Codec VBR Data 78

4.4 On the parameters to describe H.261 Video Codec Data 79

5. VBR Video Source Models 81

5.1 Origin of Short-term and Long-term Correlations in 1-Layer H.261 Video Codec VBR Data 81

5.2 A VBR Video Source Model Review 82

5.2.1 General82

5.2.2 Autoregressive Models 83

5.2.2.1 First Order Autoregressive Model 83

5.2.2.2 Superposition of two First Order Autoregressive Models 83

5.2.3 Markov Chain Models 84

5.2.3.1 Birth-Death Continuous Time Video Model(Linear Case) 84

5.2.3.2 Two Dimensional "Birth-Death" Continuous Time Video Model(Linear Case) 86
5.2.3.3 Birth-Death Discrete Time Video Model(Non-Linear Case) 88
5.2.3.4 Discrete Autoregressive Video Model : DAR(1) model 89
5.2.4 Discussion 90
5.3 A VBR Video Source Hierarchical Model 90
5.3.1 The Natural Hierarchy of a Video Source 90
5.3.2 A Formal Description of the Natural Hierarchy 92
5.3.3 Method of Hierarchical Model Simulation 92
5.3.4 Hierarchical Video Source Models 93
5.3.4.1 General 93
5.3.4.2 Continuous Time Models 93
5.3.4.3 Discrete Time Models 95
5.3.4.4 Structurally Stable Fluid Video Models 97
5.3.5 Discussion of Models 99
5.4 On the Modelling of a 2-Layer Video Source 100

6. VBR Video Model Simulation 102

6.1 Simulation Requirement 102
6.2 Simulation Software Description and Design 102
6.3 Discrete Autoregressive Video(DAV) Model Simulation 104
6.3.1 The Talk-Listen Strategy 105
6.3.2 The Motion Level Strategy 113
6.4 Discrete State Non-correlated Video Model Simulation 123
6.4.1 The Talk-Listen Strategy 123
6.4.2 The Motion Level Strategy 124
6.5 Impact of Duration Distributions in the Higher Layer 126
6.6 Discussion 128

7. Conclusions and Recommendations 132

7.1 Conclusions 132
7.2 Recommendations for the future 135

References 138

Appendices

1. Introduction

Computer networks are growing in magnitude from LANs to MANs to WANs but to enable inter-communication between all networks, all must adhere to the Open Systems Interconnect(OSI) Reference model[1, 51]. Application of the OSI reference model to any network design and operation will ease the integration process of all networks. This ultimate network is aptly named the Integrated Services Digital Network(ISDN). The ISDN reference model[1, 52] embodies the OSI reference model where it incorporates several information functional groupings : User, Control and Management; into each layer of the OSI reference model. Presently, end terminals can only communicate with an end terminal resident on the same network or on limited networks. ISDN will nullify this problem by making network types transparent to the user allowing cross-network communication. ISDN will also provide more user control in communication link set-up and service selection.

The ISDN standard was expected to provide a combination of narrowband services(i.e. services which require small bandwidths($\leq$ 64kbits/s) for data transmission e.g. speech, facsimile and small data files) and broadband services(i.e. services which require large bandwidths($>>$ 64kbits/s) for data transmission e.g. videoconferencing, videophone, television and very large data files) but the transfer mode recommended for implementation on the ISDN was determined to be inflexible particularly for the provision of broadband services. To overcome this problem a packet oriented transfer mode was recommended, the Asynchronous Transfer Mode(ATM), for implementation on the Broadband-ISDN(B-ISDN). This transfer mode, as it suggests, is asynchronous and hence provides a better platform to achieve statistical gains and design a more efficient network. Its greatest advantage will be in the provision of video services.

The objective of this research therefore is to statistically characterise and stochastically model video sources for application to ATM B-ISDN network simulations. Video source models are necessary to determine the impact such sources will have on network resources(buffers, bandwidth, delay, switch capacities, etc.) and to determine whether any advantage can be had from statistical characteristics inherent in video data. The most important parameter to determine in any case is the Quality of Service the network can guarantee video services i.e. the Probability of Data Loss. This is crucial to the provision of video services due to the serious effects data loss has on video coding techniques.

The following chapter, Chapter 2, brings the B-ISDN into perspective. The chapter begins with a very brief introduction to the evolution of digital networks. This evolutionary process led to the ISDN concept which is described to a certain degree of detail. The ISDN was the forerunner to the B-ISDN, hence it was necessary to describe to some extent the characteristics and properties of the ISDN. Also this information is relevant because B-ISDNs will inherit the larger majority, if not all, of ISDN attributes. This eventually leads to a description of required ISDN enhancements to upgrade ISDNs to B-ISDNs in order to accommodate broadband services. Data transfer modes considered for implementation on B-ISDNs are then introduced and discussed. Emphasis here is on introducing ATM by describing the ATM packet in relation to the ATM Reference Model, how such entities are to be networked, proposals on how errors within a packet are to be controlled and proposals on how cell traffic will be policed. Finally, aspects to maintaining the quality of service in an ATM B-ISDN are introduced with particular emphasis on resources and conditions required to maintain constant quality pictures in video services.

Chapter 3 introduces the video coding end terminal that will support video services on ATM B-ISDNs. The chapter initially introduces video coding techniques that were previously applied in the analysis of video sequences and also introduces future video coding techniques under consideration. The primary objective of the chapter though is to introduce and describe the present CCITT H.261 Video Coding Standard. Each aspect of the video coder section of the video codec architecture is explained. The coder employs two coding techniques referred to as Intraframe and Interframe coding, hence the hybrid codec description. The former coding technique takes advantage of spatial redundancy in a picture while the latter coding technique takes advantage of both spatial and temporal redundancy in a picture sequence. The compression algorithm employed in both coding techniques is the Discrete Cosine Transform(DCT) which transforms the spatial domain into the spatial frequency domain. The coding process which selects either coding technique is quite intelligent but is not part of the CCITT H.261 Video Coding standard. The decision to apply either coding technique is based upon a decision threshold proposed in Reference Model 8(RM8)[50]. This video coding reference model was developed as a tool for use by CCITT SGXV Specialist Group on Coding for Visual Telephony to evaluate proposed features to be included in the H.261 Recommendation. Lastly, recommended adaptations to the present CCITT H.261 Video Coding standard are introduced. These adaptations are proposed methods to achieve constant quality pictures by implementing H.261 Video Codecs. The present H.261 Video Coding standard as it stands produces variable quality pictures. Therefore, if ATM B-ISDNs are to guarantee constant picture quality video services, these CCITT H.261 Video Coding standard adaptations must be seriously considered.

Video sources are analysed in Chapter 4 to determine whether such sources can be statistically characterised in any way. Video sources are initially analysed prior to being processed by a 1-Layer VBR(Variable Bit Rate) H.261 Video Codec to determine whether the contents of the video sequence will have an impact on the statistical characteristics of video traffic. In the case of videophone sequences(head and shoulders/waist-up sequences) it is thought talking and listening scenes exhibit different statistics. Several videophone sequences were recorded to carry out this analysis. Attention is then turned to post-processing analysis i.e. 1-Layer VBR H.261 Video Codec traffic is analysed. Two analysis strategies are employed to analyse the video traffic : Talk-Listen strategy and Motion Level strategy. The Talk-Listen strategy analysis approach is influenced by pre-processing analysis while the Motion Level strategy is influenced by a combination of both pre- and post-processing analyses. A video sequence is analysed by both strategies to determine statistical characteristics - mean, standard deviation, probability distribution and autocorrelation - of the duration and traffic within each state(Talk, Listen, High Motion and Low Motion). The analysis also reveals long-term autocorrelation phenomena within video traffic. Moreover, variable bit rate video sources show different short-term and long-term autocorrelation behaviour. The long-term autocorrelation effects are topics of interest and are primarily thought to originate from the alternating Talk-Listen and Motion Level processes. The chapter finally discusses the effectiveness of recommended parameters to characterise video data in comparison to the parameters determined by the analysis strategies employed.

Chapter 5 introduces stochastic models that will implement statistical characteristics determined by analysis in Chapter 4. The chapter first defines the possible origin of long-term autocorrelation effects in video traffic and provides a clear indication that a hierarchical stochastic model is required to capture long-term phenomena in video traffic. A hierarchical stochastic model is appropriate to match the hierarchical structure of video source data transmission. Previous video source models are reviewed, discussed and considered for implementation to model 1-Layer VBR H.261 Video Codec traffic. All models were either Autoregressive or Markov Chain models on the frame level. Markov Chain models gave the option of Discrete State Continuous Time or Discrete State Discrete Time models where Autoregressive models could only offer Continuous State Discrete Time models. Moreover, models should be considered in discrete time because a video source will transmit a frame at discrete instants. On the other hand, Discrete State Continuous Time Markov Chains are very often more convenient for queuing analysis. One important feature of these reviewed video source stochastic models is the non-hierarchical structure in all models. These models neglect the long-term effects in video coded data but are adequate for the short-term autocorrelation modelling task. Hierarchical video source stochastic models are then presented. A wide range of models are forwarded for

application, in both Continuous and Discrete Time, to the video source modelling process. The chapter lastly demonstrates a method to simulate 2-Layer H.261 Video Codec traffic via a 1-Layer VBR H.261 Video Codec stochastic model.

A hierarchical stochastic model under the influence of statistical characteristics determined in Chapter 4 is simulated in Chapter 6. The chapter initially explains the need for a software simulation model followed by a description of the software design. The hierarchical model simulates 1-Layer VBR H.261 Video Codec traffic controlled by statistical parameters determined separately by both analysis strategies : Talk-Listen and Motion Level. Short-term autocorrelation model coefficients were also estimated by various techniques. Model results were statistically analysed and compared to those of Chapter 4 to determine the effectiveness of the applied model and modelling process. The hierarchical stochastic model for both analysis strategies was further investigated to ensure its ability to capture long-term autocorrelations inherent in video traffic.

Chapter 7 concludes the report by discussing the results produced by the hierarchical stochastic models under the influence of statistical parameters determined by either the Talk-Listen strategy or the Motion Level strategy. Comparisons of model performance for both strategies are made on all levels - mean, standard deviation and autocorrelation characteristic. The objective of this analysis is to determine the best stochastic model by indicating the most successful analysis strategy. The chapter finally mentions recommendations for future video source analysis to support and improve H.261 Video Codec traffic statistical characterisation and stochastic modelling.

2. The Broadband Integrated Services Digital Network

2.1 General

To introduce the Broadband Integrated Services Digital Network(B-ISDN) it is a prerequisite to introduce its immediate predecessor, the Integrated Services Digital Network (or Narrowband ISDN) which has the task of laying the foundations for its broadband counterpart.

2.1.1 The Integrated Services Digital Network (ISDN)

The ISDN network concept has naturally evolved over the past 20-30 years. The initial step on the ladder toward ISDN was taken when the decision was made to digitise the analogue voice and data communications systems approximately 25 years ago. This investigation was endorsed then by the International Consultative Committee on Telegraphy and Telephony(CCITT), a sub-committee of the UN treaty organisation ITU, which is now known as ITU-T(International Telecommunications Union-Telecommunication Standardization Sector). From this investigation there evolved the Integrated Digital Network(IDN) which combined the technologies of digital switching and digital transmission. Digital transmission was achieved by 8kHz sampling pulse code modulation(PCM), at 8bit quantisation achieving a rate of 64kbps(kilobits per second). Voice was the main determining factor for a 64kbps channel rate due to audible voice having a bandwidth of 4kHz and being the predominant source for the communications networks at that period. Hence with the advent of the ISDN concept, 64kbps channels became the basic building blocks to enable the network to provide further services other than voice and data.

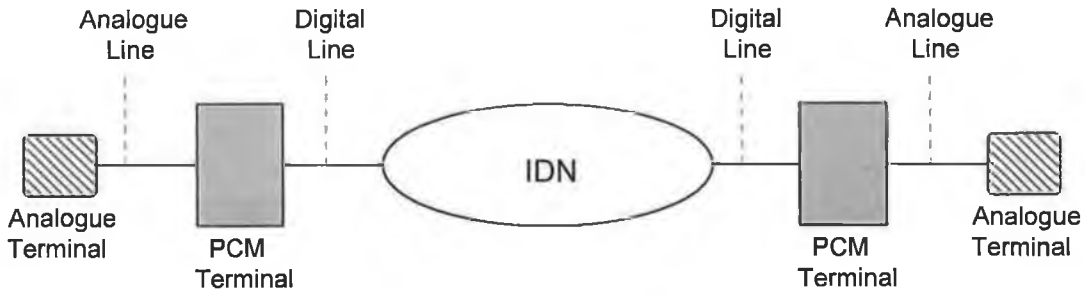


Fig. 2.1

In the IDN case, the end user would still access the network via an analogue terminal which in turn would connect to a PCM terminal(Fig. 2.1). In ISDN this is

not necessary, the analogue elements will be replaced by totally digital elements allowing a completely digital system(Fig. 2.2). This is the first component of the ISDN definition which is built on three fundamental system concepts :

1. Digital connectivity(end-to-end) for all information transfer across the network;
2. Common channel signalling interchange for establishing or terminating all communications and for controlling resources;
3. Capability of providing multiple service capability for the user at the user-network interface point.

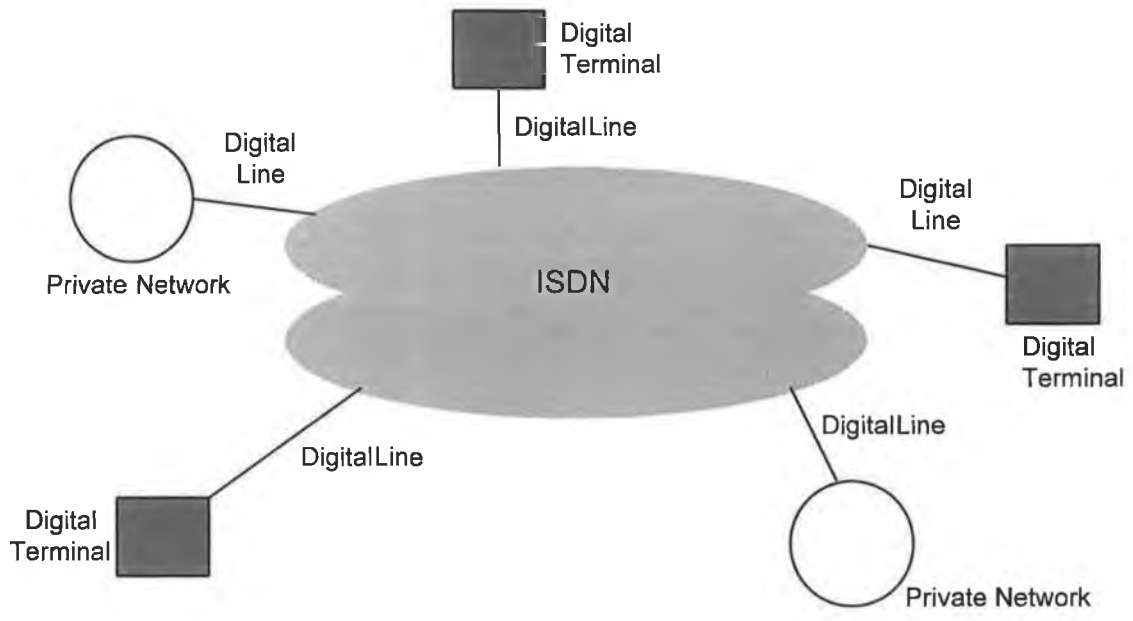


Fig. 2.2

To uphold the three fundamental system concepts, universally compatible telecommunications equipment must be established. Therefore ISDNs must adhere to international standards, to ensure a functional ISDN. Another main aspect of the ISDN is user control. The user will be given more control over the network by having access to signalling functions.

As already mentioned, the main building block in the ISDN is a 64kbps channel rate which is a fundamental component in determining ISDN access rates. The 64kbps channel rate building block concept created the following ISDN interface rates : *Basic Rate Access*[12] and *Primary Rate Access*[13]. The Basic Rate interface consists of 2B + D channels where B and D channels are 64kbps and 16kbps transmission rate channels respectively. The Primary Rate interface provides 30B + D channels in Europe while in the United States it is 23B + D channels but unlike the

D channel rate for Basic Interface Rate, the D channel rate for Primary Rate interface is increased to 64kbps.

The B-channel is defined as the main user channel and is the main information transporter on the ISDN. The D-channel provides a control information path between user and network and can also be used for packet switched data. Hence, Basic Access Rate provides 144kbps interface bandwidth to the network and Primary Access Rate, a 2.048Mbps(Megabits per second) interface bandwidth(this is in Europe, 1.544Mbps in the United States of America).

Network efficiency(i.e. efficient use of network resources e.g. bandwidth, links, buffers, etc.) is ensured by optimal network design. This can be achieved by multiplexing lower transmission rate channels onto higher transmission rate channels which are referred to as H channels[14] in the ISDN architecture. H channels provide the following channel rates :

H_0-channel rate	:	384kbps;
H_1-channel rate	:	1536kbps in the United States(H_{11}), 1920kbps in Europe(H_{12}).

These higher rate channels are intended for a variety of user information streams, e.g.

- fast facsimile;
- video e.g. videophone, videoconferencing;
- high speed data;
- high quality audio;
- multiplexed lower transmission rate information streams;
- packet switched information.

H channels though do not carry signalling information to set up a circuit switched connection on the ISDN.

The major thrust of ISDN is the incorporation of a wide spectrum of user services beyond those currently available. A key element of ISDN will be integrated access to all communications services needed by the user and is designed to give the user greater flexibility and more control over their communications facilities. The ISDN services proposed span a broad range of features, ranging from information retrieval, transaction processing, home office, electronic mail, security services and low bitrate video services but the fundamental role of ISDN is to provide a fully digital communication facility for sophisticated voice and data applications[1],[2].

2.1.2 ISDN Enhancement

Due to increasing improvements in Optical Fibre quality, i.e. decreasing impurity levels and improved opto-electronic sources and detectors, optical fibre transmission characteristics have improved enormously. Optical fibre losses have continually decreased and hence network overheads are decreasing, i.e. the repeaterless distance of an optical fibre link is always increasing. The bandwidth available in an optical fibre is also in the Gigabit range(nearing the Terabit range) and therefore, when ISDN achieves "fibre to the home" the bandwidth available will more than likely be under-utilised. Considering the advances in VLSI(Very Large Scale Integration) technology more advanced services can be visualised for implementation on the ISDN e.g. higher quality videotelephone, videoconferencing, fibre distributed television services, etc., which will utilise the bandwidth provided by optical fibre. Such services are bandwidth hungry and are termed broadband sources. The present ISDN standard has a very inflexible structure for the provision of broadband services therefore to provide broadband services a Broadband-ISDN standard is required.

The Broadband-ISDN(B-ISDN) will become the ultimate integrated services digital network. This network will provide combined broadband services(video) and narrowband services(voice and data). The network must be a high-speed telecommunications system which can cater for video services in particular. Video synchronisation is important and thus must be preserved to prevent loss of service or degradation in picture quality. The network must also be robust to ensure effective and efficient operation.

In the ultimate ISDN architecture - a B-ISDN architecture - the user will only utilise the capacity needed at any instant in time, and will be charged accordingly. Thus, the traffic will be variable in nature therefore bandwidth allocation must be dynamic. This is the case for video traffic which can be highly variable depending on the video sequence content i.e. motion or non-motion video sequence. Dynamically allocated bandwidth allows for statistical gains to be made which in turn will improve network efficiency.

Optical fibres are now capable of supporting transmission rates in the upper Gbits/s range subsequently making the idea of providing broadband services a reality. Broadband User Network Interfaces(BUNIs) have been recommended based upon the Synchronous Digital Hierarchy(SDH)[54, 55] standard which is a world multiplexing standard that supports switched broadband telecommunication services. SDH originated from SONET(Synchronous Optical NETwork)[56], a standard created to carry the North American hierarchy of bit rates: 1.5/6/45 Mbits/s. This standard was also defined to carry the European hierarchy of bit rates: 2/34/140 Mbits/s but to avoid a multitude of Plesiochronous Digital Hierarchies(PDHs)

forming worldwide, as was occurring, the SDH standard was defined with a primary rate of 155.52 Mbits/s to carry the lower transmission rates of 1.5/2/6/34/45/140 Mbits/s. SDH therefore incorporates the B, H_0 and H_1 channels for information transportation in the present ISDN in addition to the higher rate channels, H_2 and H_4. These additional higher rate channels have the following bit rates :

H_{21}-channel	:	32,768 kbps;
H_{22}-channel	:	43 - 45 Mbps (an integer multiple of 64 kbps);
H_4-channel	:	132 - 138.240 Mbps (an integer multiple of 64 kbps).

SDH is also defined to support higher interface rates at integer multiples of 155.52 Mbits/s, in a 155.52 Mbits/s x 4*n* sequence where $n = 1,2,....$ and is only limited by technology[56]. Payloads are successively layered by the multiplexing process into lower-order and higher-order Virtual Containers(VCs) which include overhead functions for management and error monitoring. Synchronous Transport Modules(STMs)(see Fig. 2.3) then transport the VCs on the network to their destination[57]. This approach therefore removes payload error monitoring to the network boundaries while the network only concentrates on assurring the delivery of each transmitted STM to its desired destination fulfilling the required Quality of Service.

There are advantages to transporting data on the B-ISDN by implementing the SDH standard[56]:

i)Improved flexibility and services - flexibility of the payload structure, direct access to 64kbits/s is possible from high bit rates, circuit provisioning is incorporated in transmission products manageable remotely down to customer terminals, plenty of overhead channels are provided for management allowing faster response and more complex operations;

ii)Improved transmission performance - better circuit availability, reduction in transmission errors due to an optical network, improved circuit quality due to better monitoring and maintenance, faster and more widespread restoration and protection.

The above advantages that the SDH offer meet the requirements for the provision of broadband services on the B-ISDN along with additional advantages of *iii)reduction of costs in equipment*, *iv)reduction of costs in operation* and *v)use of advanced technology*.

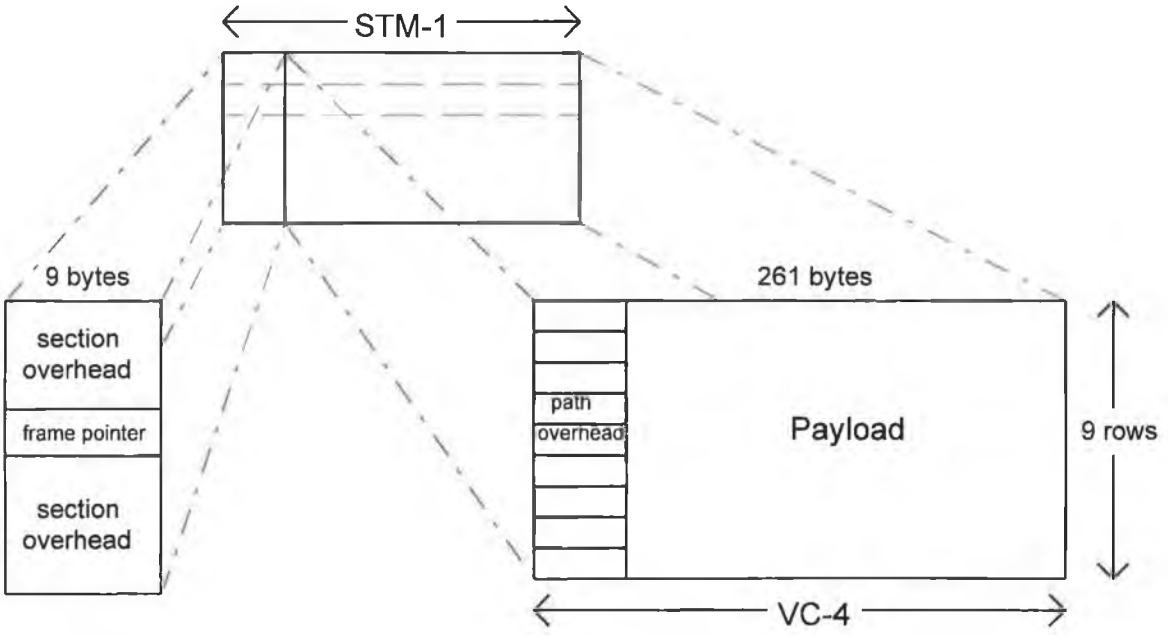


Fig. 2.3 *SDH Synchronous Transport Module(STM), Virtual Container(VC) and overheads for the primary rate 155.52 Mbits/s(STM-1 transmitted every 125μs).*

The transmission links in the B-ISDN will be Virtual Paths(VPs) and Virtual Channels(VCs). To improve switching and network efficiency the VP and VC hierarchy concept was introduced - VCs are contained within VPs. Within the network architecture, switching elements must be improved to reduce the transit delay especially for high volumes of traffic that will be presented to the B-ISDN. To overcome delays imposed by individual VC switching, batch switching has been introduced to form VP switching. In this case the VCs within an incoming pre-switched VP will have the same destination post-switched VP and if required the VCs will also be individually switched. The transmission time gains expected from this switching strategy will be significant and probably reduce the number of switching nodes traversed in an end-to-end link. This is highly attractive for video services where delay is a problem when synchronisation must be maintained [3, 4, 5, 6, 7].

2.2 Transmission Modes on the B-ISDN

2.2.1 General

A principle of B-ISDNs is to support both switched and non-switched connections, and connections made in a B-ISDN must also support both circuit-mode and packet-mode services[15]. Therefore to meet these specifications two transmission techniques, or transfer modes, were under consideration for B-ISDN :

1. Synchronous Transfer Mode(STM),

2. Asynchronous Transfer Mode(ATM).

Both transfer modes are described in more detail in the following sections.

2.2.2 SynchronousTransfer Mode (STM)

STM is comparable to Time Division Multiplexing(TDM) in present communication systems. It is based upon allocation and subdivision of bandwidth within a channel by allocating time slots within a channel, or frame, to a service for the duration of a call. The same time slots are allocated to the user throughout the period of a call. STM is strongly influenced by digital telephony, as is the ISDN channel structure, therefore STM is the foundation of the digital transmission hierarchies present in the ISDN. The STM was the initial transfer mode under consideration for B-ISDN because the Basic Access Rate could be extended by adding higher bit-rate channels i.e. $i\mathrm{H}_4+j\mathrm{H}_3+k\mathrm{H}_2+l\mathrm{H}_1+m\mathrm{H}_0+n\mathrm{B}+\mathrm{D}$, where $i,j,....,n$ define the number of higher bit-rate channels required to maintain a service available on the ISDN. It was assumed that this would allow an early introduction of B-ISDN services but the flexibility of the STM structure was questioned for future dynamic bandwidth needs because STM techniques do not support bursty services effectively or efficiently. Hence this transmission mode is more suited to providing fixed bit-rate services and higher priority services on the B-ISDN [2].

2.2.3 Asynchronous Transfer Mode (ATM)

Unlike STM, ATM does not assign time slots to a channel. Capacity is dynamically assigned on demand and this capacity is placed in fixed size packets, known as cells. By adopting this allocation process, advantage can be taken of statistical gains among bursty services or pauses in traffic while at the same time guaranteeing acceptable continuous bit rate performance. It is also very flexible by supporting a wide range of services enhancing network integration and utilisation. Table 2.1 illustrates this flexibility in ATM where it supports four approved service classes[3].

Table 2.1

	Class A	Class B	Class C	Class D
Timing relation between source and destination	Required	Required	Not Required	Not Required
Bit Rate	Constant	Variable	Variable	Variable
Connection Mode	Connection Oriented	Connection Oriented	Connection Oriented	Connection-less

Examples of the four approved service classes are as follows :

Class A : Constant Bit Rate video, voice;
Class B : Variable Bit Rate video;
Class C : Connection oriented data transfer;
Class D : Support of connectionless data transfer.

From Table 2.1, video sources are represented by classes A and B. The present H.261 Video Coding Standard[3] is a Class A service while the 2-Layer H.261 Video Codec, a proposed adaptation of the present H.261 Video Coding Standard[3], is a combined Class A and Class B service because the second layer of the video codec transmits a variable bit rate.

STM in comparison with ATM would only manage to support two service classes : a service class similar to service class A above(circuit-mode) and a connectionless class A(packet-mode). Hence with the ability to support a greater diversity of service classes, ATM over-shadowed STM in the process to determine a transfer mode for B-ISDN.

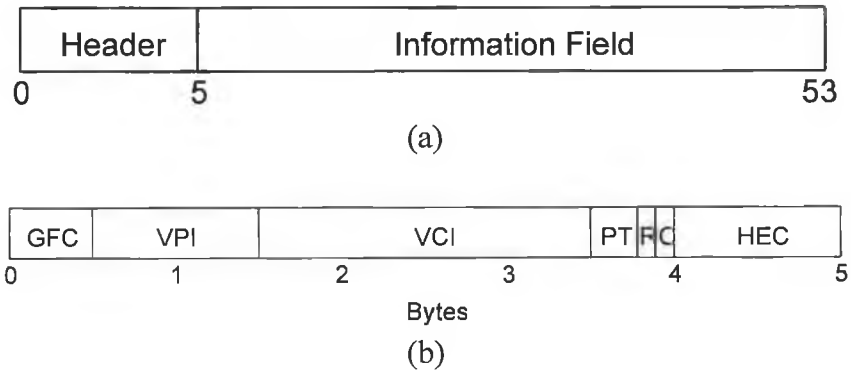


Fig. 2.4 *(a) the cell, (b) the header and its various fields*

2.2.3.1 The ATM Packet(Cell)

The ATM is a connection-oriented packet-switching concept were the packet, known as the cell, is a 53 byte long entity, which consists of an information field and a header(Fig. 2.4a). The information field, only for user traffic, is 48 bytes long. The header is 5 bytes long and mainly consists of network instructions and header error control(Fig. 2.4b).

The header component of the cell is interrogated in compliance with the B-ISDN ATM Protocol Reference Model(Fig. 2.5) by the network. Fig. 2.4b illustrates the header contents and its various fields[3] which are as follows :

- Generic Flow Control(GFC) - A 4 bit field present only at the UNI(User-Network Interface), it is intended to be used to arbitrate usage between multiple terminals sharing the same access connection. This field is otherwise an extension to the VPI field in NNIs(Network-Node Interfaces);
- Virtual Path Identifier(VPI) - An 8 bit field which identifies the path a cell must be multiplexed through in an end-to-end connections;
- Virtual Channel Identifier(VCI) - A 12 bit field that identifies the channel a cell must be multiplexed through in an end-to-end connection;
- Payload Channel Identifier(PT) - 2 bits;
- Reserved(R) - 1 bit reserved for future network applications;
- Cell loss priority(C) - 1 bit high or low priority indicator;
- Header Error Control(HEC) - An 8 bit field to prevent loss of header data, and subsequently the cell, the header instructions must be error encoded.

Each layer of the B-ISDN ATM Protocol Reference Model have specific functions but the ATM and Adaptation layers of the Reference Model take specific interest in the contents of the cell. The function of these layers and their relationship with the contents of cells are as follows[3] :

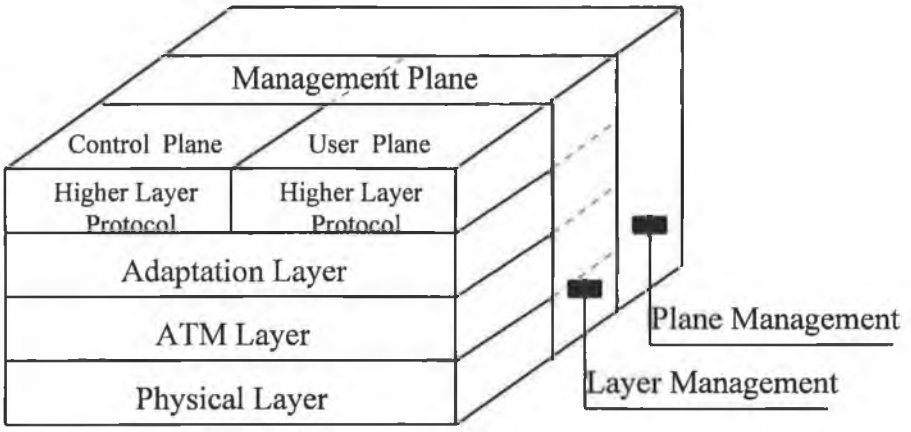


Fig. 2.5 *B-ISDN ATM Protocol Reference Model.*

ATM Layer

The ATM layer functions are :

- cell delineation;
- merging and recovering of cell boundary information;
- appending and checking HEC;
- cell rate decoupling by insertion and suppression of idle cells;
- multiplexing/demultiplexing of cells from/to different VCs and VPs.

All the contents of the cell header are of interest to this layer of the B-ISDN ATM Protocol Reference Model because the cell header only contains network information i.e. VPI, VCI, HEC, etc. Therefore, a network node will only consist of a Physical layer and an ATM layer of the reference model along with their corresponding management and control functions, remaining transparent to the information field. For every node traversed by a cell, the header contents will be read and analysed and on re-transmission the fields within the header will be altered and updated if necessary. Cell header fields which will be changed and updated in particular are the VPI, VCI and HEC fields.

Adaptation Layer

The Adaptation layer is subdivided into two sublayers :

- Convergence Sublayer(CS);
- Segmentation And Reassembly(SAR) sublayer.

The lower of the two is the SAR sublayer which operates on a cell by cell basis and depends on the class of service(see Table 2.1) being implemented. The SAR sublayer

is only interested in the information field of the cell, referred to as the SAR-PDU(Protocol Data Unit), which contains a Sequence Number(SN), Sequence Number Protection(SNP) and a SAR-PDU Payload. The SN has 4 bits and is used to detect lost or misdelivered cells and SNP, which also has 4 bits, is used to protect the SN. SN may also take on a specific value to indicate a special purpose SAR-PDU.

The higher sublayer, the CS sublayer, operates on information units also called Protocol Data Units(CS-PDU). Functions performed in this sublayer are as follows :

- Error correction is performed on the transported information, bit interleaving may also be implemented to further protect against errors;
- Clock recovery service for the receiver for synchronous dependent services;
- Explicit time indication by insertion of time stamp pattern in CS-PDU payload;
- Sequence number processing to detect lost or misdelivered cells.

Overall, the essential functions performed by the Adaptation layer are :

- cell assembly and disassembly(production of constant or variable length blocks of user information);
- end to end synchronisation by a Time Stamp(TS) field;
- detection of lost or inserted cells by a Sequence Number(SN) field;
- protection of information blocks against bit errors by a cyclic redundancy check(CRC) field;
- compensation for the cell delay variation.

The main function of the Adaptation layer is to insert/extract the transmitted/received higher layer data in the reference model into/from the information field of the cell.

2.2.3.2 ATM Networking

Emphasis to this stage in the ATM description has been on channel structure and the cell and bandwidth allocation technique. To achieve an effective and efficient network, multiplexing and switching elements must also be included. These elements will as usual be located within nodes on the network and will be implemented at the ATM layer of the ATM Protocol Reference Model(Fig. 2.5). Due to ATM being a new concept the present day multiplexing and switching architecture will not succeed in the ATM multiplexing and switching task. The ATM concept requires these elements to be fast and have large bandwidths to prevent nodal congestion. The speed is necessary to reduce propagation time for delay sensitive services e.g. video. B-

ISDN is expected to provide a wide range of services therefore these network components must exhibit the bandwidth required.

The ATM network efficiency can be controlled by an effective multiplexer. To achieve fast multiplexing, a parallel load and serial transmit architecture has been accepted while to provide bandwidth a buffer has also been incorporated into the architecture. The buffer provides *virtual* bandwidth through absorbing data from bursty traffic sources which would experience data loss otherwise and subsequently smoothes the traffic flow for further stages within the network. It is at this point statistical gains can be made by the network. The multiplexer has the capacity to serve N deterministic sources but by taking advantage of statistical characteristics of bursty deterministic sources, a statistical gain of G might be achieved. Therefore the multiplexer will statistically serve $G*N$ sources without degrading the quality of service guaranteed to the N deterministic sources. The multiplexer permits such statistical gains to be achieved due to the presence of the buffering resource which consequently prevents traffic losses. To minimise the multiplexing contribution to the propagation delay an optimum buffer dimension must be reached to ensure statistical gain and negligible propagation delay.

To minimise propagation delay in the switching process, fast switching techniques must be employed. The switching architecture presents a challenge due to the dynamic nature of bandwidth allocation within ATM. If large switch processing speeds and an effective architecture can be designed to carry out the switching procedures on a VP and a VC level, switch efficiency will result. Therefore switches will have acquired the large bandwidths required for the wide range of services expected on the B-ISDN.

2.2.3.3 Error Control Strategies

Virtually error free transmissions are achieveable due to the SDH multiplexing standard, optical fibre being an excellent transmission medium and lastly the user determining the bandwidth usage by exerting control over the transmitted data, but to safeguard against possible losses and traffic degradation, error control techniques must still be implemented to maintain the Quality of Service offered by the network. Present error control strategies are adequate for present network requirements but for application on the ATM B-ISDN their efficiency and effectiveness come into question. Considering one particular popular error control strategy, which requests re-transmission of all data from the instant the data was in error, the application of this error control technique is unviable for implementation on B-ISDNs. It's unviable due to the high volume of traffic expected on a B-ISDN which would cause delays during the re-transmission process to recover the error. In addition, this error control strategy would require extremely large nodal memories to support error free transmission for such high volumes of traffic at very high

transmission rates. Other error control strategies exist where re-transmission is not required but there is higher demand on network node processing. They are referred to as Forward Error Correcting(FEC) strategies.

More robust error control strategies are demanded and are essential for implementation on B-ISDNs because network speed and high traffic volumes expected from the wide range of services on the network do not favour presently applied re-transmission error control techniques. To maintain network speed and efficiency, fast and effective FEC strategies must be employed. Block and trellis coding are the most effective FEC techniques available at present. Therefore, due to the packet nature of ATM, block coding techniques are considered to be the most efficient error control strategies where cells are packed into blocks for error checking. For overall network performance, by incurring small propagation delays as a result of extra error control processing, FEC strategies are seen to be the most effective error control strategies to maintain network efficiency.

The above error control strategies are proposals strictly applied by the network to the contents of the cell header. Error control in the information field is solely carried out by the interacting end terminals. In the case of video, loss of data is detrimental to the service quality. In particular, loss of synchronisation in cell stream reception due to cell loss or misdelivered cells will cause severe degradations in picture quality and will remain unrectified until picture synchronisation has been reset. The loss of synchronisation has such a large impact because video services are synchronous dependent due to the coding nature of video images and sequences(see Chapter 3). Loss of synchronisation is also caused by cell delay jitter which is a function of network loading.

To combat loss of video information several techniques have been proposed and are continuously assessed[9, 10, 11]. Two of the main error recovery approaches under consideration are : i) the use of FEC codes and ii) error concealment by use of visual redundancy. Block coding is considered to be the best solution for FEC coding due to the packet nature of data transmission on B-ISDN but the main reason for it's consideration is due to the unreasonable error code lengths that would be necessary for alternative FEC coding techniques to retrieve lost individual cell information. Analysis on applying this proposed error recovery approach in comparison with error concealment by visual redundancy has not been strongly considered. Error concealment by visual redundancy is where loss of pixel intensity information is retrieved by averaging neighbouring pixel intensities. More precisely this technique takes advantage of the field structure in transmitted images(single pictures in a video sequence are transmitted in two fields - one odd and one even - which are interlaced to create a picture). The technique retrieves a lost pixel intensity, (x,y), in the odd field by averaging the intensities of pixels $(x,y+1)$ and $(x,y-1)$ in the even field, and vice versa. This error recovery technique is ideal for small information loss i.e. loss

of pixel data or even loss of data for one line of pixels, but when more than one line or a block of data in an image is lost as a result of cell loss other techniques must be employed. A recovery technique considered for this case is error concealment by data from the previously transmitted picture i.e. the lost data in the present picture is replaced by the corresponding data in the previous picture and motion compensated if necessary. Analysis in [10, 11] found the performance of both error concealment error recovery techniques to be exceptional.

To nullify the cell delay jitter at the receiving video codec, it has been proposed to implement an elastic buffer to absorb the cell delay jitter[9]. Due to the variational nature of cell delay jitter there is the possibility that later transmitted cells may be received prior to earlier transmitted cells. Therefore to re-synchronise the cell order before the decoding process, cells are time stamped to preserve cell synchronisation. But to maintain the video sequence rate there is a finite cell delay that can be tolerated, and if cells are not received within the maximum delay, they are effectively lost. Error recovery in such circumstances will be carried out by FEC and visual redundancy error concealment techniques.

The above video data error recovery techniques are quite powerful but do not rectify the error within a single picture duration. The error further propagates through the video sequence until the received data achieves full coding synchronisation again. In order to maintain picture coding synchronisation, there has been proposals to reset the coding synchronisation during every 8^{th} transmitted line in a video sequence picture[9](this is arbitrary, the synchronisation could be reset for every transmitted line). By implementing this approach, the error is prevented from propagating spatially through the picture, whereas with visual redundancy error concealment techniques, spatial error propagation will occur in addition to temporal error propagation through the video sequence. This is of consequence due to the video coding techniques which are utilised(see chapter 3). Therefore, in resetting the video coding synchronisation during a picture the error propagation is restricted to one section of the picture and only propagates temporally, not spatially except in the region where the error has occurred.

2.2.3.4 Traffic Control

The B-ISDN is designed to cater for a wide range of services. Under the ATM concept the bandwidth allocation for each service will be dynamic. Therefore, during the setting-up phase of a communications link between two(or more) users, the users must negotiate a QOS(Quality of Service) level with the network. If the users violate the agreed level of QOS, the network will activate it's traffic control procedures to protect other network users, thus information loss is imminent. To ensure a certain level of performance for the network user in such circumstances, the network must provide some levels of traffic control capabilities[3] which are as follows :

i)Connection admission control : The user will only be admitted to the network only when the network can provide and maintain the QOS required for the service along with existing network users.

ii)Policing : To ensure network users do not violate their agreement with the network on the QOS required, monitoring and control procedures must be implemented. These structures are put in place to protect network resources from malicious and/or unintentional behaviour which other established connected users require to maintain their QOS. The monitoring and control is applied to traffic volume(bandwidth) and cell routing validity which only occurs during the information transfer phase of a connection. The policing function is performed on VCs and VPs only. Several policing techniques have been proposed i.e. Peak, Jumping Window, Stepping Window, Moving Window, Leaky Bucket, Exponential Weighted Moving Average(EWMA). Only one has been considered seriously, the Leaky Bucket technique but has been adapted to improve network performance. The Leaky Bucket technique can provide the QOS that is required but if the user traffic exceeds the agreed level of QOS, cell losses are imminent. Cells which violate the requested QOS are identified and discarded by the congestion management systems which are located within the network nodes. The Leaky Bucket technique maintains the agreed QOS and protects the existing users QOS but large cell losses will result particularly for bursty services if the Leaky Bucket policing technique is employed. Therefore to improve cell loss performance an adaptation of the Leaky Bucket technique has been proposed, the Virtual Leaky Bucket technique. This technique identifies violating cells and if the network has the capacity without threatening existing users' QOS, they're not discarded. These violating cells are discarded by the congestion management system only when the QOS of existing network users is under threat.

The above policing techniques are proposed network policing functions which threaten information loss and degradation of services. In the case of video services loss of information is detrimental to data synchronisation which is necessary to preserve the communication link. Therefore video terminals may require self-policing. To preserve the communication link one method for self-policing has been proposed. The proposal suggests placing a buffer on the output channel to safeguard against violations of the connection agreement with the network[9]. The buffer will limit the output rate of the source to the agreed rate negotiated with the network but to prevent buffer overflow and information loss during periods of high data transmission, the buffer must instruct the video coder to adjust it's coding. This procedure will affect the received picture quality due to the implementation of coarser coding which instantly reduces the level of data being produced. But B-ISDN

guarantees constant picture quality video services therefore instead of placing a static length buffer on the source output, an elastic buffer may be preferred.

iii)Priority Control : To maintain a definite level of QOS, a priority cell traffic control strategy could be implemented. The ATM B-ISDN is being designed to take advantage of statistical gains that can be achieved but a probability will exist for zero or minimal statistical gain. To safeguard against QOS degradation in such situations, priority cells have been proposed to ensure continuous QOS. ATM cells have this facility where a Cell Loss Priority(CLP) bit can be allocated in the cell header(see {2.2.3.1}). The provision of a CLP bit satisfies proposed video codec architectures (see chapter 3) which are under consideration to provide video services on B-ISDNs. The video codec architecture is a layered structure, two layers in all(more layers could be added) where high priority information is transmitted only on one layer. The high priority information is necessary to maintain the communication link and is not dependent on the lower priority information. The lower priority information improves the picture quality produced by the high priority information which contains the main structural information of the transmitted picture. Therefore with only two state priorities available with the CLP bit, the priority traffic control technique would be more appropriate for application to two layered video codecs.

2.2.3.5 ATM Call Set up

To achieve call set up by ATM the following steps are taken[58] :

1) When a new terminal wants to communicate with another terminal it must be allocated its own signalling channel by being assigned a SVCI(Signalling Virtual Channel Identifier). Therefore to obtain a SVCI the new terminal first sends a request over a predefined channel, the MSCV(MetaSignalling Virtual Channel) which is shared by all terminals on this access. The metasignalling message informs the network that a new terminal has been connected which prompts the network to allocate a SVCI to the newly connected terminal. Note - Metasignalling is never used to setup an ATM call.

2) To make a call the user gives the terminal the directory number required and the type of service to be used(i.e. telephone, videophone, etc.).

3) The home exchange then receives the requirements from the newly connected terminal to set up a connection within the network. The exchange finds a suitable route across the network by assigning new translation values to the translation tables within the node switches satisfying the quality of service required.

4) The connection is switched through when the destination terminal answers by the process of a signalling message sent back along the route informing switches to update their translation tables to allow fast switching of cells en route to this destination.

5) On call completion, the call is cleared down by a signalling message sent back along the call route to each exchange to invoke clearance of translation table entries related to that connection.

Due to the flexibility of the B-ISDN, the above call set up sequence is only one variant of numerous call set up sequences. Call set up can also offer advanced look-ahead, early switch-through, etc.

When a call is set up, ATM cells are transported from source to destination and vice versa by packing cells into the payload section of the VC within a STM(see Fig. 2.3)[57] from the SDH standard. The SDH standard was selected due to its inherent advantages previously highlighted in {2.1.2}. From this synchronous hierarchical multiplexing standard two standard BUNI standards have been defined and are presently in place to transport ATM cells[58]. Both these standards recommend UNI transmission rates of STM-1(155.52 Mbits/s) and STM-4(622.08 Mbits/s). To avoid accidental or deliberate confusion of the SDH framing information a special scrambler is also used in the ATM layer when running over SDH.

2.3 Aspects of ATM Quality of Service

Some aspects of the ATM quality of service have already been partially discussed when considering the impact of network elements on the B-ISDN performance. Problems with information loss and delay were mentioned but not fully explained in terms of their cause and consequences. Items expected to characterise the quality of service are now outlined[3].

i) Information Loss : Information loss can be caused by a number of occurrences :

- Errors in the cell header;
- Cell insertion from other virtual channels due to cell header errors;
- Cell losses due to buffer overflows in the network;
- Bit errors occurring within the cell information field.

To combat such losses error control and traffic control techniques were discussed in {2.2.3.3} and {2.2.3.4} respectively. These techniques are required to be robust to protect data loss sensitive services. For example, the effect of cell loss on the picture quality in video services will cause degradation and loss of synchronisation.

ii)Delay : There are two delay components to be contemplated, end-to-end delay and delay jitter. In the case of video services, end-to-end delay is vital due to the strong need for synchronisation to maintain the quality of service required. In the case of data services, such a parameter is not important except for particular data cases e.g. stock market data, important medical data, etc. Delay jitter is defined as the difference in time in the arrival of cells from the ideal arrival time. Clearly this effect is caused by buffering where the buffering duration varies en route from transmitter to receiver i.e. the network load. This form of delay degrades the QOS required by voice and video services in particular. Hence the requirement to reduce the number of nodes traversed within a connection.

iii)Throughput : Even though the ATM B-ISDN is a dynamic bandwidth allocation concept, it will support two distinctly different throughput modes :

- Constant Bit Rate(CBR);
- Variable Bit Rate(VBR).

The CBR throughput mode can be associated with the STM while the VBR

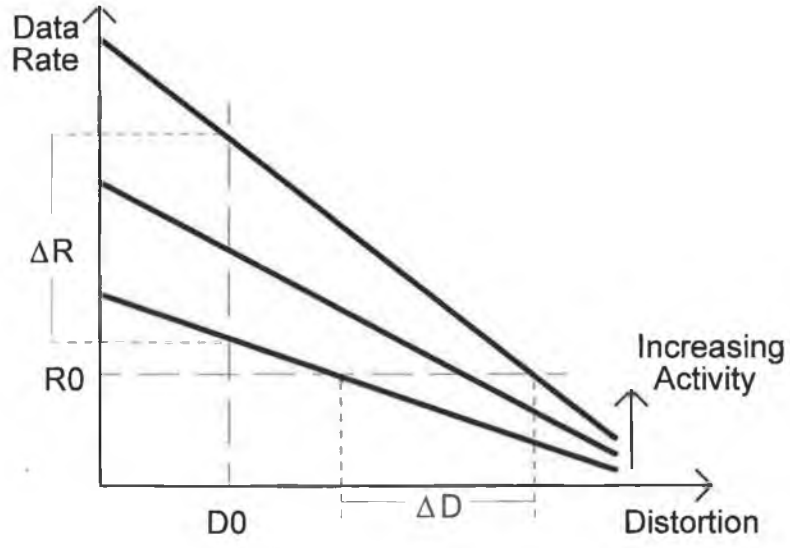


Fig. 2.6 *Rate - Distortion Graph. With a constant data rate(R0), distortion will vary(ΔD) during increasing activity while with a varying data rate(ΔR) there will be constant distortion(D0)[8, 9].*

throughput mode on the other hand is more associated with the ATM. Both throughput modes also offer separate levels of QOS to the network user. The CBR throughput mode is expected to offer a higher level of QOS than the VBR throughput mode due to the levels of control that can be applied to a CBR throughput mode. The

VBR mode is a consequence of the ATM concept where varying bandwidth services can be provided. In the case of video sources, a main aim for the ATM is to obtain near constant picture quality which can only be achieved by allowing the transmitted video data rate to vary. This is described in information theory by the Rate-Distortion function, $R(D)$ shown in Fig. 2.6, where there is a complicated trade-off between the minimum achievable coding bit rate and the distortion of the decoded images[8]. The data rate from a video codec varies, ΔR in Fig. 2.6, as a result of the activity present in the video sequence(a phenomena related to the video coding process). In the case of low bitrate video services(or CBR video sources), R0 in Fig. 2.6, the data rate is constant hence the picture quality varies due to a variation in distortion, ΔD. Therefore video services on the B-ISDN will require a VBR throughput to maintain a constant picture quality(or constant picture distortion), D0.

As already mentioned in {2.2.3.4}, prioritising data is one possible technique to guarantee a level of QOS. The two throughput modes present a method to make this possible in conjunction with the two priority level capabilities available from the Cell Loss Priority bit in the cell header{2.2.3.1}. Higher priority could be assigned to CBR throughput modes due to constant bandwidth allocation and ease of control. A VBR throughput mode on the other hand requires dynamically allocated capacity and is consequently more prone to higher information losses. Therefore by utilising two throughput modes, a two level QOS connection can be negotiated with the network.

3. CCITT H.261 Video Coding Standard

3.1 General

It is without doubt that the B-ISDN Network has enormous potential but it must be remembered that bandwidth remains limited. This fact becomes a reality particularly when large information sources (e.g. video sources) are being considered for transportation on the network. To increase the capacity of the network, processing techniques must be applied to take advantage of the redundancy contained within the information sources (as is applied to speech sources).

Two coding processes which take advantage of redundancy are implemented in video sources, either separately or in combination. The two separate coding processes are referred to as Interframe and Intraframe coding. The former process takes advantage of a combination of redundant information in a video sequence - from picture to picture(temporal redundancy) and within a picture(spatial redundancy), while the latter process takes advantage of redundant information within a single video picture(spatial redundancy) alone. Application of either coding process is dependent on the contents of the video sequence. Interframe coding would be applied to a video sequence containing minute motion where high redundancy gains can be achieved from picture to picture i.e. each picture in the video sequence is virtually static. Implementing Intraframe coding instead of Interframe coding in this instance would add extra overheads because of higher processing times required but more importantly Intraframe coding would have negligible redundancy gains. Video frames would be Intraframe coded without achieving any redundancy gain as a result of virtual similarity from picture to picture. Hence the reason for selecting Interframe coding in low motion video sequences.

Interframe coding redundancy gains are negligible in high motion video sequences which is quite understandable due to the non-static behaviour of the video sequence contents from picture to picture. The Interframe coding redundancy gain would remain low or constant throughout the video sequence in such circumstances. Intraframe coding operates on a single frame and during high motion video sequences can achieve higher redundancy gains than the Interframe coding process. In some high motion video sequences the Intraframe coding redundancy gain may increase but more than likely the gain will vary homogeneously from picture to picture. Subsequently Intraframe coding is more applicable to high motion video

sequences than Interframe coding. Therefore a combination of both coding processes would be applied to video sequences containing periods of both high and low motion.

Several video coding algorithms have been created to take advantage of redundancy within video source picture sequences. The algorithms are either one or a combination of the coding processes : Interframe and Intraframe. Following is a list of coding processes applied to video coding[3, 8, 16, 17, 18, 19, 20, 21, 22, 23, 24, 10, 9, 11, 25, 26, 27, 28, 29, 30] :

1. Conditional Replenishment;
2. Differential Pulse Code Modulation (DPCM);
3. Discrete Cosine Transform (DCT);
4. Sub-Band Coding;
5. Wavelet Transform.

Besides the five video coding techniques mentioned above, other powerful video coding schemes are being developed i.e. Fractal based video coding[31] and Object based video coding[32] algorithms. In fractal coding, the objective is to identify the smallest pattern from which the full image can be re-constructed by pattern repetition i.e. the image is compressed to a single pattern and is de-compressed by repeating the pattern until the full image has been re-constructed. The main drawback with fractal coding at present is slow compression time but it's main advantage is an extremely fast de-compression time. Object based coding leads to a segmentation of images based on detecting changed regions in a video picture. These regions are then split into moving objects and the background is uncovered. A further region type is that of areas which do not comply to the model, referred to as model failure regions. The region shapes are then independently approximated and transmitted to the decoder with any motion and colour information necessary to synthesise the entire image.

The most widely implemented video coding processes at present are those which take greater advantage of Interframe and Intraframe redundancy within video sequences, referred to as hybrid coding processes. Such performances are true from processes 3-5 mentioned above. The CCITT H.261 Video Coding Standard implements the Discrete Cosine Transform video coding algorithm. This standard defines the future B-ISDN video service terminal operation.

3.2 The H.261 Video Codec

The CCITT H.261 Video Coding Standard was recommended to code and decode moving picture components of audio-visual services at p x 64kbps rates, where p = 1,.....,30. This would enable the provision of a video service on ISDN at Basic Access rate, guaranteeing minimum picture quality (p = 1 or 2). To have

maximum picture quality ($p = 30$) would require an access rate of 1920kbps - an ISDN H_{12} transmission rate line would be required for this purpose. The codec architecture is outlined in Fig. 3.1 and each codec element is described in detail thereafter.

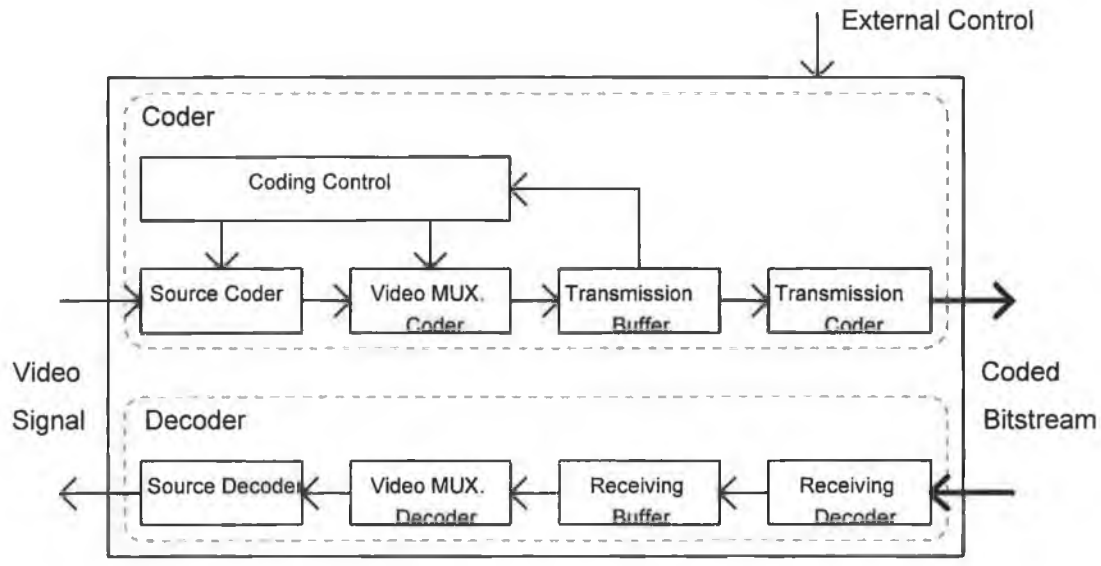


Fig. 3.1 *H.261 Video Codec architecture[3]*

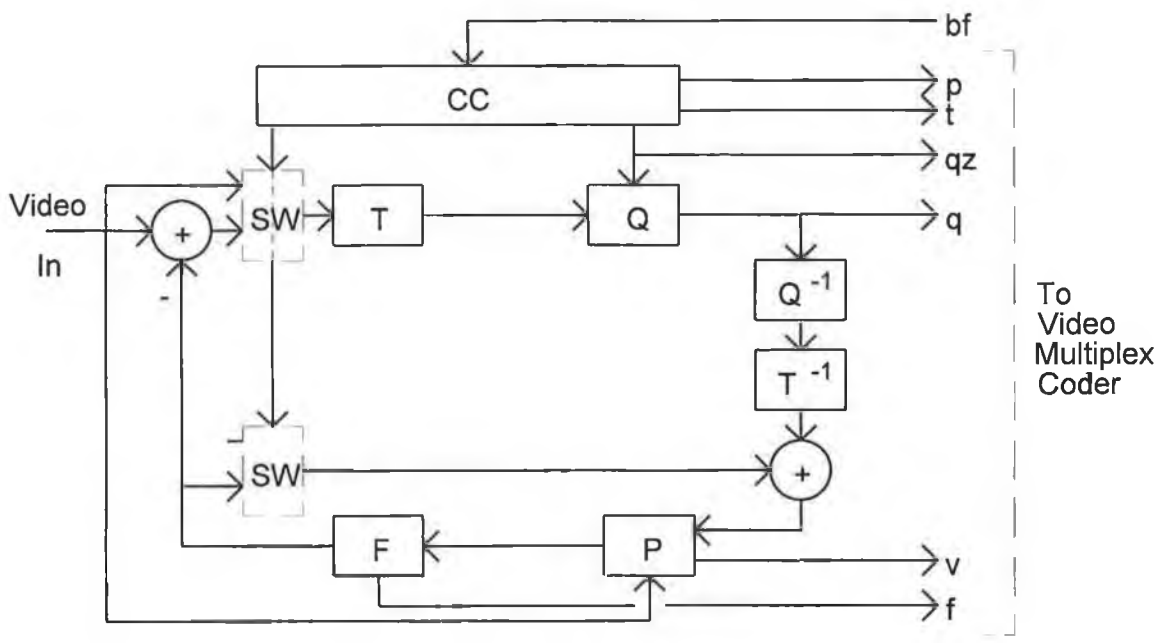
3.2.1 The Source Coder

The Source Coder is the main processing element of the codec. Subsequent processing merely alters the data for transmission. The objective of the source coder, or any communications coding process, is to take advantage of redundancy within the source information. When redundant information is eliminated from the transmitted information stream, efficient bandwidth utilisation is achieved. The source coder is a hybrid coder which implements inter-picture prediction, to capitalise on temporal redundancy(picture-to-picture redundancy), and transform coding of the inter-picture prediction error to take advantage of spatial redundancy. This is not an entirely true statement because the coding is applied at a macroblock(see {3.2.2}) level, not at the frame level(see {3.2.1.4}) but for simplicity, assume the coding process is on the frame level for the moment. The inter-picture prediction error is determined by implementing feedback illustrated in the source coder architecture(Fig. 3.2). The source coder is further described as follows in more detail.

3.2.1.1 Source Coder Input Signal Format

The input video signal has two recommended picture formats - common intermediate format(CIF) and Quarter-CIF(QCIF). Both picture formats are non-interlaced(unlike television pictures where two fields interlace to create one picture) and occur at a picture rate of 30pictures/sec. The codec samples' the video signal at

the same rate and from each picture produces three components - 1 luminance and 2 chrominance components. CIF pictures are scanned to produce a 352 pels(Picture ELementS) per line x 288 lines per picture luminance component and 176 pels per line x 144 lines per picture chrominance components. QCIF pictures have half the number of pels and lines per luminance and chrominance components. All codecs are required to operate on QCIF pictures, although some can operate on both picture formats.



T : Transformer
Q : Quantiser
P : Picture memory / Motion Compensation
F : Loop filter
CC : Coding Control
SW : Switch

p : flag for INTRA/INTER coding
t : flag for transmitted or not
qz : quantiser indication
q : quantising index for transform coefficients
v : motion vector
bf : transmission buffer feedback
f : switching on/off of the loop filter

Fig. 3.2 *Source Coder architecture[3]*

3.2.1.2 The Transformer

The transformer operates on 8 pel x 8 line blocks and transforms their contents from the spatial domain to the spatial frequency domain via a two dimensional Discrete Cosine Transform(DCT). This process produces a 64 DCT coefficient word set where each coefficient ranges in magnitude from -2048 to +2047. These coefficients carry the video information required for transmission, but all are not required due to the sensitivity of the human eye. The essential coefficients for human eye detection and perception are contained within the lower frequencies of the spatial domain. Therefore, picture reproduction requires only half, or less than half, of the DCT coefficients. The higher frequency coefficients are discarded.

The DCT function is given by :

$$F(u,v) = 1/4\,C(u)C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y)\cos[\pi(2x+1)u/16]\cos[\pi(2y+1)v/16]$$

where

$$u,v,x,y = 0,\ldots,7;$$
$$C(u) = 1/\sqrt{2} \text{ for } u = 0, \text{ otherwise } 1;$$
$$C(v) = 1/\sqrt{2} \text{ for } v = 0, \text{ otherwise } 1.$$

(3.1)

x and y are the spatial co-ordinates in the pel domain, u and v are the co-ordinates in the transform(spatial frequency) domain. At the receiving end of the video connection the DCT coefficients are inverse DCT transformed to give pixel values in the range -256 to +255 by the following function :

$$f(x,y) = 1/4 \sum_{u=0}^{7} \sum_{v=0}^{7} C(u)C(v)F(u,v)\cos[\pi(2x+1)u/16]\cos[\pi(2y+1)v/16]$$

(3.2)

where all the variables are as given in (3.1).

3.2.1.3 The Quantiser

From the transformer the DCT coefficients are passed to the quantiser where they are quantised to the nearest step value between the range -2048 and +2047. The first DCT coefficient, known as the DC coefficient, contains the majority of low frequency information which is required to make the coded image structurally perceptible. Therefore when Intraframe coding is initiated without Interframe coding(this happens when there is a scene change in the video sequence or if the rate of motion in the sequence increases), the DC coefficient is nominally the transform value, linearly quantised with a stepsize of 8. The remaining DCT coefficients are

also linear quantised but with even stepsizes varying from 2 to 62(31 various stepsizes in all) from block to block. The selected stepsize for application to a block, except the DC coefficient, is not applied to the full DCT coefficient magnitude range. A constant quantisation level is applied to an area referred to as the Central Dead-Zone(mainly applied to the region around a coefficient magnitude of 0). This implies extremely positive and extremely negative low frequency DCT coefficients are important(high frequency DCT coefficients are not transmitted because their effects are not perceptible to the human eye). In the hybrid coding situation i.e. Interframe and Intraframe coding combined, the DC coefficient quantisation stepsize reverts to the quantisation stepsize applied to the whole block.

Post-quantiser processing and the input video signal are the reason for quantisation stepsize variation from block to block. If activity within the picture sequence increases, the inter-picture prediction processing creates a larger error for the transformer to process which in turn produces larger magnitude coefficients. If the quantiser at this point happened to be applying fine quantisation, large amounts of data would be transferred to the transmission buffer(Fig. 3.1) to await transmission. To prevent buffer overflow in such circumstances the feedback control line from the buffer to the coding control system of the codec initiates coarser quantisation to reduce the data entering the buffer thus preventing buffer overflow. Loss of data is prevented but at a cost to picture quality.

The inverse quantiser reverses the quantisation process but quantisation noise is introduced into the prediction process.

3.2.1.4 Picture Memory

The Picture Memory in the source coder(Fig. 3.2) compares present and previous pictures to determine motion vectors in macroblocks for transmission. The motion vector contains a vertical and a horizontal component which are integers and do not exceed +/-15. This restricts the scanning range for macroblock motion detection to the neighbouring macroblocks within the picture. The process applies to all pels within the macroblock. To improve the coding process further, a means to estimate the motion from picture to picture in video sequences(a form of forward motion prediction) would reduce data transfer between codecs. Motion vector analysis will assist in this process.

3.2.1.5 Loop Filter

The transform coding is of a block nature which would be very noticeable on picture reproduction. Therefore to combat block edges in the decoded picture and to prevent them hampering the prediction process, a two dimensional spatial filter(Fig. 3.2) is introduced into the loop to negate horizontal and vertical block edges.

3.2.1.6 Coding Control

From Fig. 3.2, coding control determines the coding process to implement, Intraframe or Hybrid coding, by controlling the switches, SW, and the quantisation stepsize to apply. These decisions solely rely upon the Transmission Buffer feedback(Fig. 3.1). The Transmission Buffer indicates to Coding Control that buffer overflow is imminent unless the coding process is altered to reduce this risk. Coding Control acts upon receiving this information by increasing the quantisation stepsize or in the case of coding a new image, caused by a scene change or high motion, makes the decision to Intraframe code only.

3.2.1.7 RM8(Reference Model 8)

Two coding techniques are employed in the H.261 Video Codec : Hybrid and Intraframe coding. Hybrid(Interframe and Intraframe) coding is applied during periods of low motion and Intraframe coding during periods of high motion in a video sequence. This statement is not entirely true because each coding process is applied only to individual macroblocks(see {3.2.3}). During motion the contents of a macroblock in the previous frame will translate, under the influence of a motion vector, to another location within the present frame. Hence, the only information required to relay to the receiving video codec is a motion vector for that particular macroblock. The decision to apply either Intraframe or Hybrid coding to the contents of a macroblock relies on the contents of the translated(motion compensated) macroblock from the previous frame, MB(i + *motion vector*, f - 1), and also on the contents of the corresponding macroblock, MB(i + *motion vector*, f), in the present frame. The decision is determined from Fig. 3.3 where

$$\mathrm{VAR} = \mathrm{MB}(i + \textit{motion\_vector}, f) - \mathrm{MB}(i + \textit{motion\_vector}, f - 1),$$

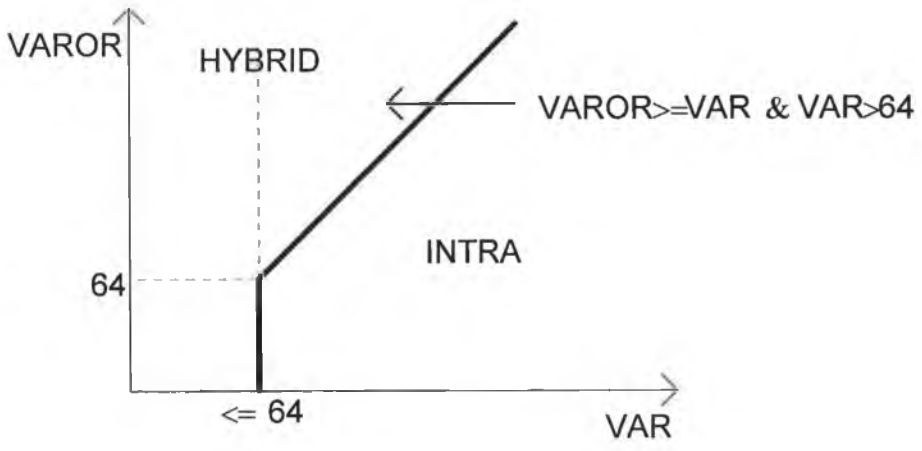


Fig. 3.3 *Hybrid/Intraframe Coding decision.*

$$\text{VAROR} = \sum_{m=1}^{M}\sum_{n=1}^{N}\left(p_{\text{MB}(i+motion\_vector,f-1)}(m,n) - mean_{p_{\text{MB}(i+motion\_vector,f)}}\right)^2,$$

$M = 16$, $N = 24$, $p(m,n)$ = pixel magnitude in macroblock at co-ordinate m,n.

VAR is basically the error between the motion compensated macroblock from the previous frame and the original macroblock in the present frame. VAROR is a measurement of the energy(mean squared error) in the original macroblock. This decision technique is not standard i.e. it is not part of the H.261 Video Coding Standard, but is implemented in the RM8(Reference Model 8)[50] video coding model. This video coding reference model was developed as a tool for use by CCITT SGXV Specialist Group on Coding for Visual Telephony to evaluate proposed features to be included in the H.261 Recommendation. Hybrid or Intraframe coding selection for a macroblock is required because during motion the luminance and chrominance information within the macroblock may change(determined by VAR and VAROR). This function is incorporated into the video Coder Coding Control to switch between Intraframe and Hybrid coding.

3.2.2 Video Multiplex Coder

Each picture is scanned by the source coder in a specific manner required by the Video Multiplex Coder. The scanning process has a hierarchical structure with four layers[3] which are as follows :

1. Picture
2. Group of Blocks(GOB)
3. Macroblock
4. Block.

1.Picture Layer

This layer consists of a picture header followed by the picture data which is transported by the following layers in the hierarchy. Headers for dropped pictures are not transmitted.

2.Group of Blocks Layer

For each video picture format, CIF or QCIF, an alternative GOB structure is implemented. For the CIF format 12 GOBs are created from a picture while from the QCIF format only 3 GOBs are created(Fig. 3.4)[3]. The GOBs are transmitted in the sequence the GOBs are numbered in Fig. 3.4.

3.Macroblock Layer

Contained within each GOB are 33 macroblocks as illustrated in Fig. 3.5[3]. The macroblocks are also transmitted as sequenced in Fig. 3.5.

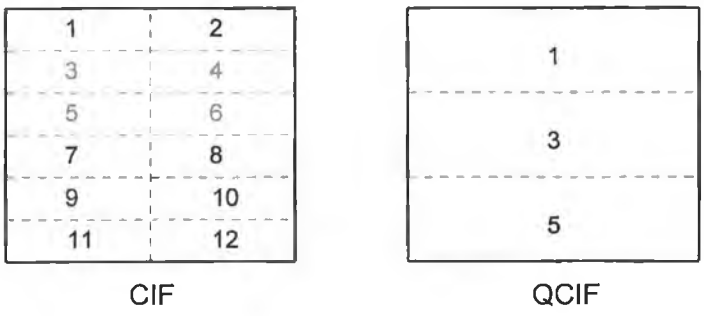


Fig. 3.4 *H.261 Video Codec picture format GOB structure*

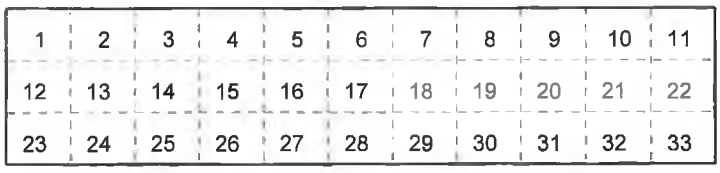


Fig. 3.5 *Macroblock structure and coding sequence within a GOB*

4.Block Layer

The block layer is the lowest layer in the video multiplex coding hierarchy. Within each macroblock there are 6 blocks - 4 luminance and 2 chrominance. Previously in {3.2.1} the source produced 1 luminance and 2 chrominance components for the picture as a whole. The 2 chrominance components are quarter the size of the luminance component, therefore, four 8 x 8 luminance blocks

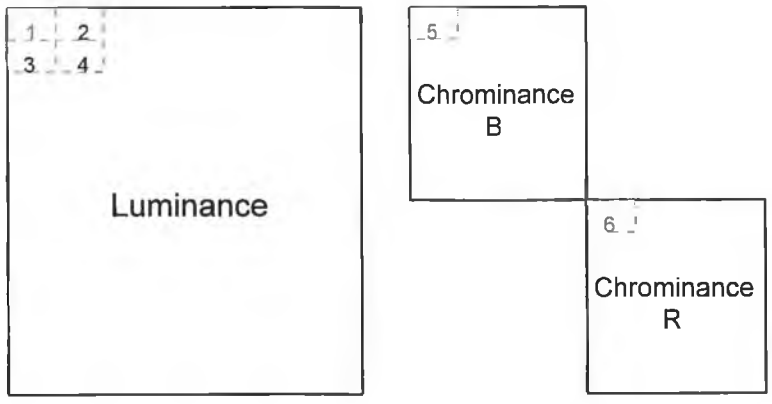


Fig. 3.6 *The luminance component(352 pels/line x 288 lines/picture) and chrominance components(176 pels/line x 144 lines/picture) of a H.261 Video Codec source*

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

Fig. 3.7 *Transmission order of block DCT coefficients*

correspond to one 8 x 8 block from each chrominance component(Fig. 3.6)[3]. Hence the reason for each macroblock containing 6 blocks. The blocks within each macroblock are video multiplex coded as sequenced in Fig. 3.6.

The DCT coefficients within each 8 x 8 block are also sequentially transmitted in a specific order known as 'zig-zag' coding, giving preference to lower frequency coefficients. The transmission order is illustrated in Fig. 3.7[3].

Finally, all the data is transmitted by implementing Run Length Coding(RLC). This coding process combines a DCT coefficient and the number of zeros which follow the coefficient into two parameters, (x, y). The x parameter denotes the quantised DCT coefficient magnitude and y denotes the number of zeros that follow the coefficient in the order of transmission given in Fig. 3.7. Variable Length Codewords(VLC), determined by Huffman Coding and stored in codebook format, replace the RLC for transmission.

The inverse to these functions are performed by the Video Multiplex Decoder.

3.2.3 Transmission Buffer

The transmission buffer is a form of transducer that monitors the output rate of the video coder. A constant bandwidth is allocated to the video codec by the network to provide a video service to the user but the inherent coding process of the codec is capable of exceeding the allocated bandwidth. Therefore the transmission buffer limits the output rate of the codec to the allocated bandwidth, or in other words the variable output rate of the Video Multiplex Coder is smoothed by the transmission buffer.

A buffer has a finite length and during periods of high data output from the Video Multiplex Coder, buffer overflow threatens. To prevent overflow occurring, the transmission buffer requests the Coding Control to adjust the coding parameters of the Source Coder i.e. quantisation stepsize. Thus, the video codec output rate is maintained and transmission buffer overflow is avoided.

3.2.4 Transmission Coder

The transmission coder has two main responsibilities : *i*) video coding delay and *ii*) error coding.

i) *Video Coding Delay* : The video encoder and decoder delays must be known to allow audio compensation delays to remain fixed for lip synchronisation when a H.261 Video Codec forms part of a conversational service.

ii) *Error Coding* : The error coding implemented is of a Forward Error Correcting (FEC) type. The transmission coder inserts a BCH (511,493) FEC code into the transmitted data. To assist the decoder to identify the block information, an error correction framing pattern is included which consists of eight BCH (511,493) frames(Fig. 3.8). Each frame is comprised of 1 bit framing, 1 bit fill indicator(Fi), 492 bits of coded data and 18 bits parity. Fi indicates to the decoder that the encoder is transmitting coded data or non-coded data, which is basically a stream of 1's. The parity is calculated against the 493 bits which includes the Fi bit.

The Receiving Decoder performs the inverse functions to the transmission coder.

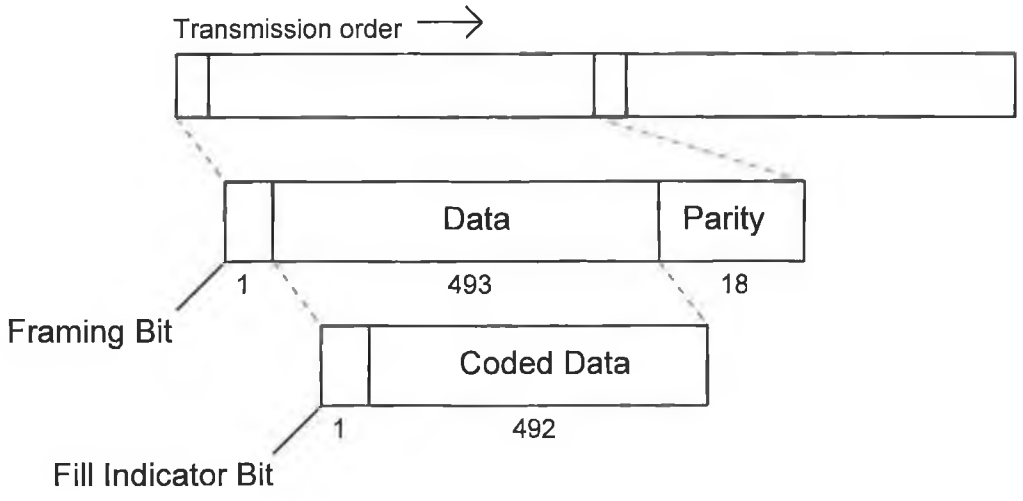


Fig. 3.8 *Error correcting frame*

3.3 Adaptation of the CCITT H.261 Video Coding Standard for B-ISDN ATM

From the description given on the H.261 Video Codec so far, it can be discerned without difficulty that the codec is a CBR(Constant Bit Rate) source. This can be deduced from the presence of the transmission buffer(Fig. 3.1), which smoothes and limits the output data rate to that of the transmission channel. By

referring to Fig. 2.6, a CBR transmission rate, R0, imposed on a naturally varying data source such as a video coder results in picture quality variation, ΔD, during active periods within a video sequence. As already explained in {3.2.2.2}, picture quality variation is due to the codec transmission buffer controlling the DCT coefficient quantisation level which directly causes the received picture quality to vary. Therefore, to realise the ATM audio-visual service aim - to obtain near constant picture quality in video services - the H.261 Video Codec must become a VBR(Variable Bit Rate) source. This is reflected in Fig. 2.6 where to maintain a constant quality picture, D0, it is necessary to allow variation in video codec output data rate, ΔR.

By changing the H.261 Video Codec method of transmission from CBR to VBR, advantage can be taken of ATM B-ISDN resources(bandwidth, buffering, speed, switching, etc.). A few coding strategies which will benefit from the ATM concept have been suggested and are under consideration. The coding strategies being considered come under two categories [3] :

1. One-layer coding;
2. Two-layer coding.

One-layer coding

One-layer coding refers to the video codec only having one output channel. The CCITT H.261 Video Coding Standard is presently a one-layer codec but data transmission is at a CBR rate. If either the data Transmission Buffer or the feedback channel from the buffer to Coding Control(Fig. 3.1) are removed from the codec, a VBR transmission rate would result because the codec is inherently a VBR codec. The alteration from a CBR source to a VBR source therefore is quite simple but the implications of the transition are enormous. For instance, to have one-layer VBR operation the quantisation stepsize must be set to a constant value. For the purposes of the following research a quantisation stepsize of 12(not standard) was applied during one-layer VBR video sequence coding.

Another aim of ATM B-ISDN is to maintain compatibility throughout the whole network for the provision of all services. VBR and CBR video codecs would be violating this aim because both codecs would be non-compatible i.e. VBR video codecs would suffer large losses during connection with CBR video codecs. Therefore, if initial audio-visual services on the B-ISDN are provided by one-layer CBR video codecs, one-layer VBR video codecs become unrealisable. To overcome the picture quality variation associated with one-layer CBR video codecs in this instance, VBR transmission must be incorporated into the codec somehow and at the same time maintain codec compatibility. This can be achieved by implementing the second coding strategy, two-layer coding.

Two-layer coding

This coding strategy has the ability to comply with the two main aims for coding real time video signals for use on the ATM B-ISDN[3] i.e.

- to obtain near constant picture quality;
- maintain compatibility with the existing CCITT H.261 specification.

A two-layer codec has two output data channels. The first layer output channel is a CBR transmitting channel and the second layer is a VBR transmitting channel(Fig. 3.9). From Fig. 3.9 the first layer is the present CCITT H.261 Video Coding Standard which preserves service compatibility on the ATM B-ISDN. To combat the variation in picture quality associated with a CBR transmission service, additional processing power has been incorporated into the video codec - a second layer. This layer transmits finely quantised(Q2) first layer transform domain quantisation noise to enhance the picture quality at the receiving codec.

From Fig. 3.9 the quantisation noise is determined in the second layer and quantised by Q2 with a quantisation stepsize of 12(this is not a standard quantisation stepsize, a quantisation stepsize of 12 was chosen arbitrarily for the purpose of coding video sequences for this research). This quantisation stepsize remains static and unlike the first layer, the stepsize does not vary throughout a video sequence. Second layer quantisation was also applied in the range -2048 to +2047.

The second layer Video Mux processing was exactly similar to the first layer Video Mux processing described in {3.2.2}.

Ideally ATM B-ISDN would prefer to offer an absolute undistorted picture QOS in audio-visual services but there exists in every telecommunication system a probability of information loss. For the two-layer H.261 Video Coder, two natural levels of priority can be provided to reduce information loss, one for each layer in the coder. The Cell Loss Priority bit provided in the ATM cell header(see {2.2.3.1}) presents an opportunity to apply a priority strategy to maintain a constant picture quality. For a two layer video coder, higher priority is assigned to the first layer, which is the present CCITT H.261 Video Coding Standard, that transmits more essential low spatial frequency DCT coefficients for picture reconstruction at the decoder(see {3.2.1.2}). The second layer does not transport the high spatial frequency DCT coefficients but transports the quantisation noise generated from the low spatial frequency DCT coefficients. Therefore the second layer is effectively an enhancement layer and is not essential for picture reproduction at the decoder, hence the reason for a lower priority assignment.

The bandwidth implications of incorporating this second layer into the CCITT H.261 Video Coding Standard are also not yet fully understood except that

bandwidth provision must be dynamic for this particular channel and the overall bandwidth requirement for a 2-Layer H.261 Video Codec terminal is increased. This process has also been hindered by the implementation of inadequate video sequences for statistical analysis which were not of substantial duration to determine the effects on network parameters such as bandwidth. The following chapters of this research thesis outline a video source model which will greatly assist in the process of determining the bandwidth implications of such a source.

Considering the coding techniques being employed on both layers and the ability of communication networks to manage CBR rates better, assigning higher priority to the CBR channel assures a higher QOS i.e. an extremely low probability of information loss(not constant picture QOS). Assigning low priority to the VBR channel allows the network, when the need arises, to discard this lower priority data in favour of the higher priority data to prevent loss of essential video picture data for picture re-construction. CBR rates are also easier to control than VBR rates, therefore VBR rates are expected to threaten higher information losses. Hence another reason for assigning a lower QOS(a higher probability of information loss) to the VBR channel of the two-layer video codec.

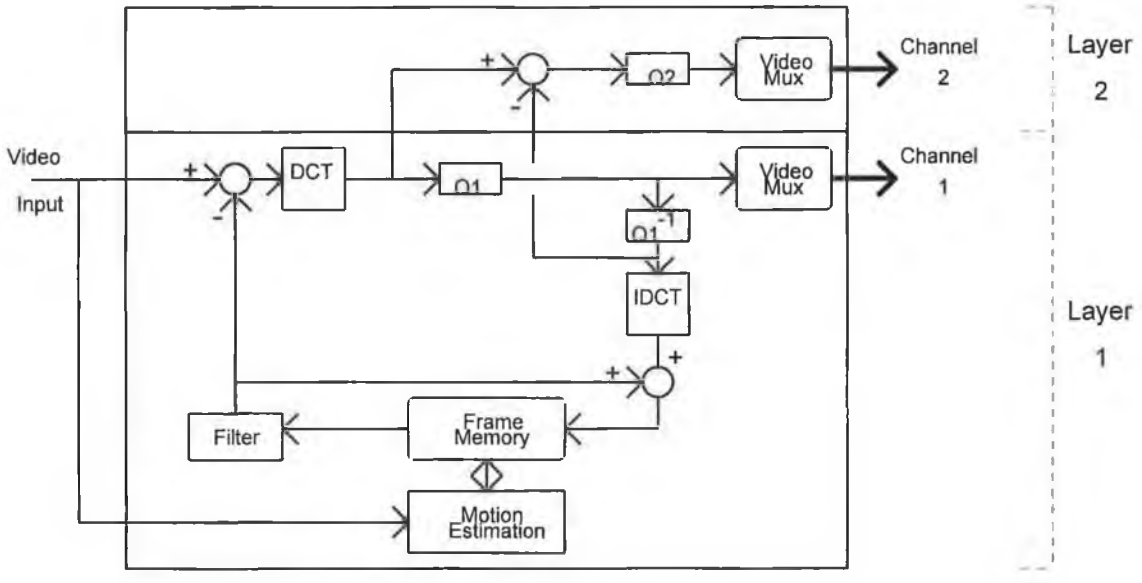


Fig. 3.9 *Two-layer H.261 Video Coder[3]*

The two-layer coding architecture in Fig. 3.9 is only one design proposal for a 2-Layer H.261 Video Coder, several others were proposed but the design in Fig. 3.9 was implemented in the following chapter when video sequences were coded to obtain statistics to characterise VBR data flows on the second layer channel. A single VBR channel H.261 Video Codec was also implemented by removing the buffer from the H.261 Video Coding Standard to obtain VBR video traffic statistics. Such research is required to generate software tools which can simulate VBR video

sources. From there another step can be taken to determine the impact such sources, one-layer and two-layer, have on the ATM B-ISDN concept by network simulation.

4. Video Source Analysis

4.1 Video Sequence Pre-Video Coding Analysis

4.1.1 General

A major attraction of the ATM B-ISDN will be the ability to provide audio-visual services of various kinds in conjunction with voice and data services(see also Table 2.1 in {2.2.3}). These various services have been compiled into two main service categories : interactive services and distributive services[53]; which are further subdivided into classes of services due to their differing characteristics outlined as follows :

Interactive Services

a) Conversational services - provide real-time bidirectional communication symmetrically or asymmetrically and in some specific cases unidirectional information flow;

b) Messaging services - provide nonreal-time user-to-user communication;

c) Retrieval services - information stored in public information centres is retrieved by the user on his demand only.

Distribution Services

a) Distribution services without user individual presentation control - broadcast services that provide a continuous flow of information distributed from a central source to an unlimited number of authorised receivers who cannot control the start and order of the presentation of the broadcasted information;

b) Distribution services with user individual presentation control - a service which distributes information from a central source providing a sequence of information entities with cyclical repetition giving users the ability to individually access the cyclically distributed information to control the start and order of presentation.

From the above service classes audio-visual services envisaged for B-ISDNs are :

1. Videophone;
2. Videoconferencing;

3. Television.

Additional services identified for B-ISDN e.g. video mail, home order shopping, advertising/notice board, entertainment-to-order, etc. depend fundamentally on the three audio-visual services listed. All three services possess various visual characteristics, discussed below, and they also differ in service provision methodology. Audio-visual services 1 and 2 are interactive conversational services while the third service is a distributive service without user individual presentation control.

The video coding standard accepted so far possessing the capabilities to support audio-visual services on the B-ISDN was introduced in Chapter 3, the H.261 Video Coding Standard. Also mentioned was the requirement to modify the existing CCITT H.261 Video Coding Standard to achieve constant picture quality services. When the codec is required to code the three diverse sources mentioned above, the output traffic is expected to be statistically diverse i.e. the mean output data rates from each source type will be significantly different as shown in [22] and [25]. Therefore, a relation may exist between the output data and the audio-visual source type, assisting audio-visual source characterisation for the modelling process. Video source models are needed for implementation in ATM B-ISDN software simulations to determine the impact on B-ISDN network parameters such as bandwidth, buffer lengths, switch capacities and delay.

4.1.2 Videophone Sources

This is the most likely public audio-visual service considered for initial implementation on the B-ISDN. A videophone service is presently in existence but the picture quality is poor as a result of the present H.261 Video Coding. The intentions of B-ISDN are to provide better quality audio-visual services as explained in Chapters 2 and 3. As mentioned in {4.1.1}, data characteristics expected from video codec sources supporting these particular audio-visual services will differ from the other two, videoconference and television sources. This can be deduced by analysing the contents of a videophone sequence.

Typical videophone sequences have the following visual characteristics :

- Head and Shoulders images;
- Localised uniform motion;
- Camera panning little to nil;
- Camera zooming is virtually nil;
- One subject per frame in the majority of cases;
- Image quality good;
- Scene changing nil.

Due to these visual characteristics, the output data from videophone codecs is expected to have negligible variation i.e. the data will vary homogeneously about a mean data level. The video coder will utilise hybrid(Interframe and Intraframe) coding(see Chapter 3) mostly to minimise the volume of data transfer due to the non-variational characteristics from picture to picture in the picture sequence. By this simple visual analysis of videophone sequence contents and considering the impact the contents have on the encoding process of the video codec, it can be deduced that a First Order autoregressive model may be appropriate to model videophone source output data[20, 35, 8].

4.1.3 Videoconference Sources

A videoconference service could very well be a fore-runner to the videophone service but not in the public domain. Both have strong similarities but a videoconference service is largely expected to have greater utilisation in the business sector. To introduce revolutionary communications services such as audio-visual services, field trials are necessary to determine network requirements and the implications, which are largely unknown, in providing audio-visual services. Field trials implementing videoconference sources are beneficial because they are in higher demand by the business sector, the business sector flavours' the service at first hand and assesses the effectiveness of the service. For monitoring purposes the majority of businesses are only operational for a particular period of the day, the business sector is a smaller user base and more than often, businesses maintain a private communications network, particularly large national companies e.g. electricity supply companies and communications companies. Public field trials on the other hand for videophone services are less controlled because monitoring would be required 24 hours a day, 7 days a week. It is also a service in smaller demand, at present, and the unknown characteristics of video traffic may threaten the quality of service guaranteed to other network users not involved in the field trials. Installing a dedicated network for public field trials would also be an expensive venture and may not yield profitable returns.

Videoconference services have potentially a higher market penetration at present mainly due to the demand for the service in the business sector. Therefore videoconference service traffic statistical analysis has been largely more vibrant[20, 34, 22, 9, 36, 18] and videoconference service field trials thus have a higher priority. In any case, videoconference sequence contents are not far removed from videophone sequence contents, described below. Hence, the traffic statistics that characterise videoconference sources will also be expected to characterise videophone sources.

The following visual characteristics are indicative of a videoconference source expected on the ATM B-ISDN :

- Head and Shoulders/Upper body image;
- Localised and non-localised uniform motion;
- Camera panning;
- Camera zooming;
- One to a few subjects per frame;
- Image quality good;
- Scene changing.

Some visual characteristics are similar in description to the visual characteristics of a videophone source but a videoconferencing system provides large output data producing functions i.e. scene changing processes, camera panning and zooming. These scene changing functions(collectively they can be looked upon as scene changing processes) can be attributed to the users of this service. Additionally the business sector require methods of communication to contact displaced business colleagues and within the duration of this communication visual aids will be utilised to relay information between both parties(graphics, picture-boards, etc.). If two or more groups are in communication, camera operation will be under human control to pan and zoom in order to capture the speaking person or the person being addressed on screen. These switching processes can only be described as scene changing.

The impact on the ATM B-ISDN by videoconferencing sources must be considered. Unlike the videophone source, the output data from a videoconference source is expected to be non-homogeneous as a result of the scene changing that will be present in the service(i.e. the output data of a videoconference source will vary homogeneously about several mean output data levels corresponding to different scenes that exist in the output data stream of a videoconference source). In this case both coding techniques, Intraframe and Hybrid(Interframe and Intraframe), are utilised by the codec resulting in a non-homogeneous output data characteristic. Also, the motion level in videoconference video sequences is expected to be higher than videophone motion levels. This will contribute a larger volume of output traffic relative to a videophone source.

The visual characteristics of a videoconference source seem to indicate a higher degree of modelling complexity to that of a videophone source. The necessity to model videoconference sources is paramount to determine the level of QOS that can be guaranteed on ATM B-ISDNs[20, 34, 22, 9, 36, 18]. Implementation of a videoconference source model in ATM B-ISDN software simulations will determine the impact such a service will have on such networks.

4.1.4 Television Sources

In comparison to the characteristics of telephony, electronic mail, videophone, videoconferencing, etc., the statistical and visual characteristics of a television

service will be extremely diverse. A television service effectively combines all these separate services into one i.e. telephony = television audio signal; electronic mail = teletext; videophone/videoconferencing = low quality television images. A need for this audio-visual service may arise in the future due to pressures on the atmospheric transmission spectrum.

The visual characteristics of a television source are as follows :

- non-Head and Shoulders images;
- non-localised motion i.e. motion varies throughout a frame and a frame sequence;
- Prolific camera panning ;
- Camera zooming;
- Number of subjects per frame vary;
- Excellent image quality required;
- Constant random scene changing.

In comparison to videophone and videoconference sources, a television source differs in every respect. The present H.261 Video Codec will be tested to the limit and possibly over-stretched in its attempt to support a television service. The output data statistical characteristics will be expected to vary heterogeneously due to the constant alternating between Intraframe and Hybrid(Intraframe and Interframe) coding techniques. That is to say the output data will vary randomly between different mean levels of data inherent in the output traffic corresponding to different scenes in the video sequence and the traffic in turn will vary homogeneously about the different levels of mean. The constant random scene changing and variation in motion will be responsible for the alternating coding techniques throughout the macroblocks of the frames. Therefore large volumes of highly varied traffic are expected as a result of these effects.

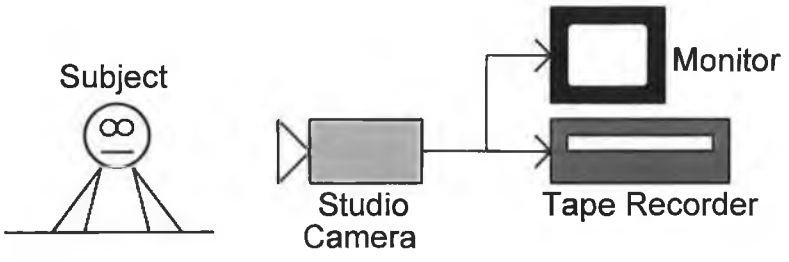
The implications of providing this audio-visual service on an ATM B-ISDN must be carefully considered even when the bandwidth requirements of the service occupy a single VP/VC between two nodes on a link connecting users to the source. Other resources on the network will be highly affected by this service e.g. buffer dimensions, the number of nodes traversed to reduce the impact of delays, network traffic control, etc. Synchronisation must be preserved to maintain a constant level of QOS. Therefore, delays must be minimal and information loss practically negligible to ensure a high level of QOS. For instance the number of nodes the data must traverse might be constant or inversely proportional to the distance the data must travel. This strategy would be necessary to maintain acceptable buffer dimensions to prevent excess delays and node congestion, otherwise data will be lost.

To safeguard against these foreseen problems network designers must utilise network simulations to determine network resource control strategies. Unlike videophone and videoconference source models, television source models will be far more complex in design. The complexity is due to the attempt to approximate statistical characteristics associated with television source data. In order to control the level of QOS for television services, a model will be required to further the analysis.

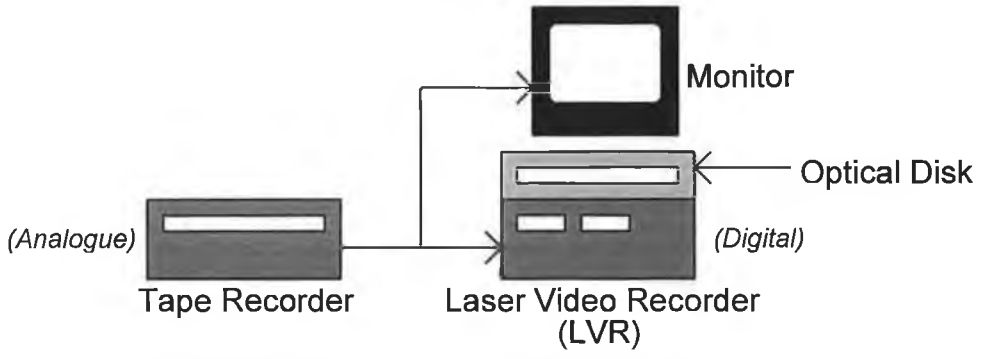
4.2 Video Sequence Post-Video Coding Analysis

The video data being analysed in the following sections was obtained by the equipment set-up shown in the following four steps:

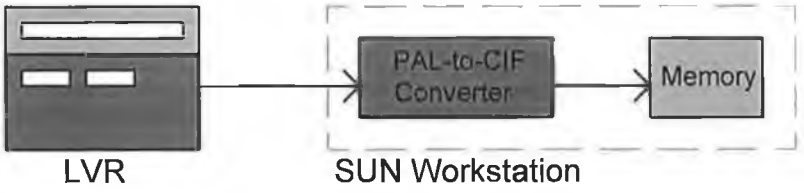
First Step



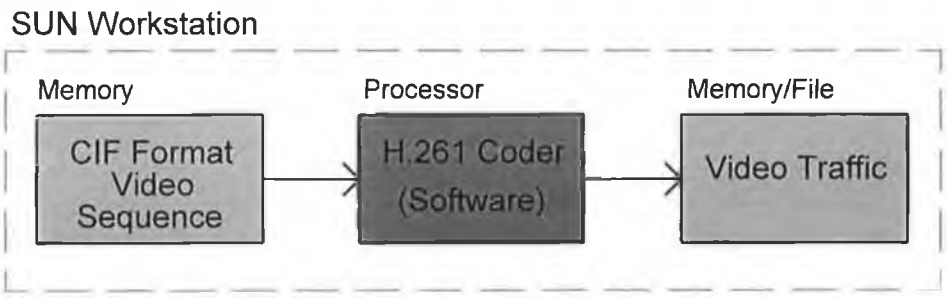
Second Step



Third Step



Fourth Step



4.2.1 The Source

The source considered in the rest of this study is a videophone source. The objective of the study will be to identify a model that can satisfactorily approximate the output data of a videophone source. A model is required to assist in the process of ATM B-ISDN simulations to determine the impact the videophone service will have on network elements such as buffer dimensions, bandwidth, multiplexing, switching, etc. and to predict the QOS that can be achieved under certain network loads. This research is not undertaken in this study, only the vital phase of characterising and modelling videophone data is undertaken here.

In any modelling process parameters must be identified that will characterise the data in question. In the videophone source case, output data is random in nature. Therefore, a videophone source traffic random variable, RV_{OD}, is required. RV_{OD} is statistically characterised by the first and second moments of the videophone source traffic(the mean(μ) and variance(ν)), a Probability Density Function with parameters μ and ν to describe the videophone source traffic distribution and a correlation characteristic to give a measure of the memory present in the random videophone source traffic, i.e. it gives a measure of the impact past data has on present data.

In the analysis of video codec traffic generally, several distributions have been identified. The majority of analysis has been carried out at the frame level(i.e. bits/frame, cells/frame)[20, 33, 19, 35, 17, 34]. Analysis has also been carried out at the pel level(i.e. bits/pel) within a frame[8], and at the transmission rate level(i.e. bits/sec)[9]. Statistical analysis on the frame level is a more natural approach because video sequence encoding is carried out on a frame by frame basis and it is presently the lowest recognised level of discretisation in video sources. The lowest level of discretisation in a H.261 Video Encoder is the macroblock level were modelling on this level would emulate realistic data transmission. From the various coding algorithms implemented(see {3.1}) the analysis revealed several video source output data distributions. Analysis on the frame, pel and transmission rate levels reveal in the majority of cases Gaussian distributed traffic. The analysis on the frame level in

other cases revealed Gamma distributed traffic[34] and more complex traffic distributions[17]. Gaussian distributed traffic was mainly produced by videophone sources at all levels(frame, pel and transmission rate) while Gamma distributed traffic was produced by videoconference sources. A television type source(a movie) produced a more complex distribution. Subsequently several video models spawned from the above analyses(see {5.3.4}). Further statistics e.g. the mean, standard deviation and peak, were reported[22, 21, 24], but without any indication to the distribution or correlation of the analysed traffic. This rendered the data statistics unusable.

The above videophone source analysis was mainly carried out on output data produced by one layer video codecs and alternative video coding algorithms. Also, the durations of the video sequences implemented in the majority of cases were short i.e. 5-20secs. because they were mainly video coding test sequences e.g. Claire, Miss. America, Salesman, Janine[35],Carphone. These video sequences could be regarded as slightly artificial videophone sequences, i.e. homogeneous backgrounds, non-natural motion, static hand and arm movement. Therefore, the full statistical characteristics of the traffic produced by videophone sequences may be neglected and the results of the analysis may not statistically characterise 2-Layer H.261 video codec traffic adequately. In other cases where longer duration videophone/videoconference sequences are implemented, further statistical characteristics are revealed[34, 21, 24, 9, 36].

Table 4.1

Videophone Sequence	Duration
Robbie	4mins. 47secs.
Sean	5mins. 34.5secs.
Paddy*	3mins. 55secs.
Donagh	5mins. 51secs.
Paul	5mins. 11secs.
Lisa	3mins. 58secs.
Rachel	2mins. 52secs.
Sinead*	2mins. 43secs.
Peter*	4mins. 42secs.

(*These sequences were recorded in a studio under special studio lighting. The other sequences were recorded in a smaller room under normal lighting conditions)

To determine a stochastic model that approximates the output data of a videophone source on the second layer of a 2-Layer H.261 Video Codec, several videophone type sequences(head and shoulders and waist-up sequences) were

recorded and coded. The sequences were recorded in the PAL standard at 25 frames/sec. and are given in Table 4.1. The video sequences succeed in capturing the expected natural relaxed behaviour of videophone users in non-artificial surroundings within a video frame.

4.2.2 Examples of Video Coded Data

Of the above videophone type sequences, Robbie was the only sequence processed by a 2-Layer H.261 Video Codec[24]. This sequence was selected based on the contents which could be described as typically human. The subject was a student who displayed active, jolly and extrovert mannerisms typical of a relaxed videophone user. The background in the sequence was also non-artificial and the lighting was virtually normal(special lighting was at a bare minimum). The contents of the other recorded sequences are similar but vary in human mannerisms because each individual is different. This was also the only H.261 Video Codec coded videophone sequence due to the long durations of the sequences and the processing resources available. Hence coding a video sequence was quite lengthy. Figs. 4.1 and 4.2 illustrate the
Robbie sequence data transmitted on channel two and on channels one and two combined, respectively. From Fig. 4.2 the output data on channel one is fairly constant while from Fig. 4.1 the data on channel two is highly variable.

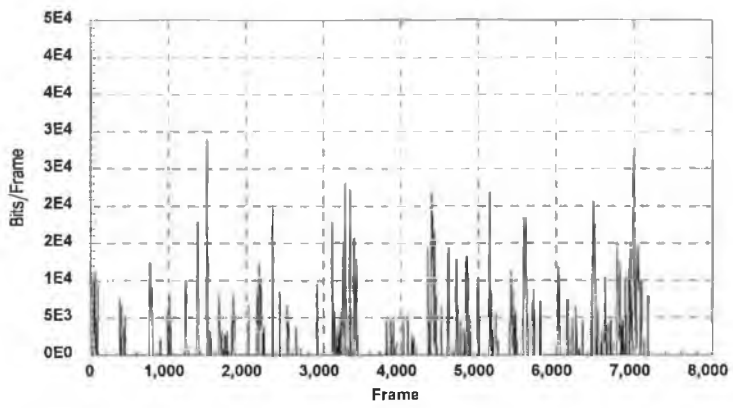


Fig. 4.1 *The data transmitted on channel two of a 2-Layer H.261 Video Codec produced by the Robbie videophone sequence.(A large transient is present in the first frame of the sequence).*

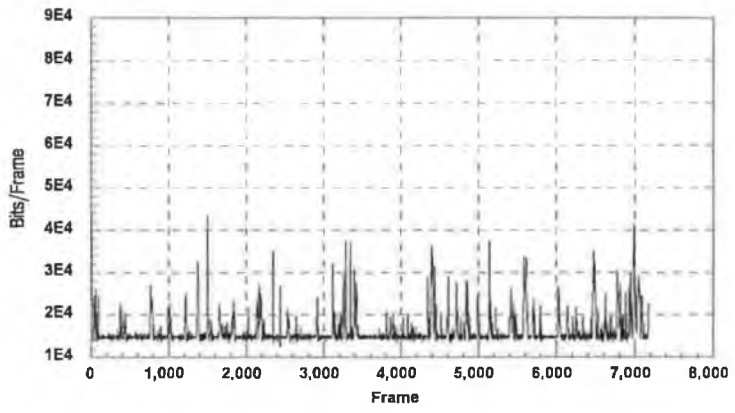


Fig. 4.2 *The data transmitted by channels one and two of a 2-Layer H.261 Video Codec combined. The data was produced by the Robbie videophone sequence.(A large transient is present in the first frame of the sequence).*

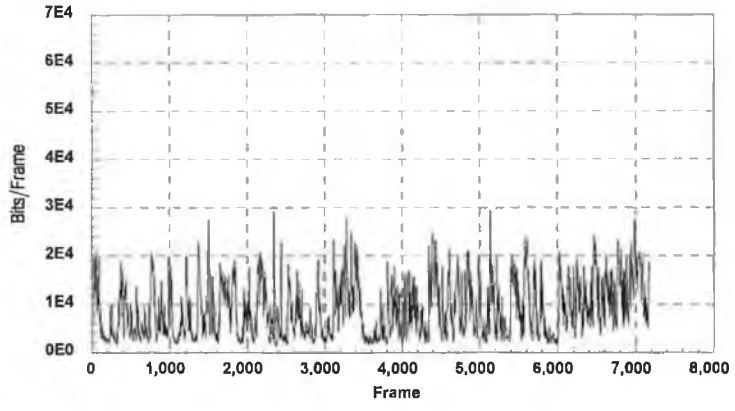


Fig. 4.3 *The output data from a 1-Layer VBR H.261 Video Codec produced by the Robbie videophone type sequence.(A large transient is present in the first frame of the sequence caused by the non-effective prediction process at the start of the coding).*

The Robbie sequence was also processed by the video codec in a one layer VBR mode. Fig. 4.3 illustrates the output data from the 1-Layer VBR H.261 Video Codec. The output data from both codec architectures were then compared to assist in characterising the source for the modelling process.

When comparing Figs. 4.2 and 4.3 noticeable differences are evident. The constant output characteristic of channel one in Fig. 4.2 is not present in Fig. 4.3, but the characteristics of the output data above this threshold are apparently preserved in both cases. Therefore, it can be deduced from this analysis that the first layer of the 2-Layer video codec compensates for low volumes of data on the CBR channel(channel 1). This compensation arises from the quantisation step-size control in the quantiser{3.2.2.2}, where the quantisation becomes finer to maintain a constant output rate. The quantisation step-size remains constant, step-size = 12, in the 1-Layer VBR video codec which results in data variation occurring below the CBR channel threshold.

In the 2-Layer video codec the quantisation level becomes an additional variable which could have major significance in the modelling process i.e. complicating the modelling process, therefore maintaining a constant quantisation level would be more attractive. From the comparisons above, the impact on the data above the CBR channel threshold in a 1-Layer VBR codec is negligible i.e. this data is virtually similar to the data on the VBR channel of a 2-Layer codec. Therefore, a 2-Layer VBR H.261 Video Codec can be modelled via a 1-Layer H.261 Video Codec model. This is made possible by the process of subtracting the constant output data of the CBR channel in a two layer codec from the output data of a one layer VBR codec.

Statistics have been gathered on the above data and are given in Table 4.2[24]. The transient peak is caused by the non-effective prediction process of the video codec at coding start-up. These statistics, the mean in particular, may not characterise the output data effectively for the modelling process. This may be deduced from Fig. 4.3 where the data displays an extremely non-homogeneous characteristic i.e. there are large transitions in the output data. To express the situation simply, the statistics given in Table 4.2 are too coarse, or too inaccurate, and will degrade the modelling process. Therefore to capture and approximate the essential characteristics of the output data in Fig. 4.3, analysis has to be made on a more detailed level. Two strategies are now introduced which attempt to do just that.

Table 4.2

Channel	Mean(kbits/frame)	Peak(kbits/frame)	Transient Peak (kbits/frame)
1-Layer VBR	8.88	29.28	67.36
2-Layer CBR+VBR	16.36	43.56	81
2-Layer VBR	1.56	28.96	41.36

(Note : To convert to bit rate i.e. bits/sec. multiply by 25 due to a frame occurring every 1/25th of a second.)

4.2.3 The Talk-Listen Strategy

This strategy adopts a Video Sequence Pre-Video Coding analysis approach to characterise the output data of videophone sources. The strategy simply isolates the talking and listening periods of a videophone sequence where it is assumed the output data for each state, talk or listen, will possess different statistical characteristics. A similar analogy would be the talk-silence characteristic in voice sources. The only difference is that data is not transmitted during silent periods in voice sources but data is continuously transmitted during the listening periods of videophone sources. The process of isolating talking and listening periods within each video sequence was achieved by meticulously identifying the border between each talking and listening period by naked eye analysis and then recording the length, in frames, of each talking and listening period.

Throughout observations of videophone type sequences and the study of subjects within the sequences, there are apparently contrasting differences between talking periods and listening periods. During talking periods the level of activity increases due to people being more animated when they are talking and particularly when there is visual contact. The output data from the video codec under these circumstances would be expected to increase as a result of the coding process. During listening periods on the other hand, the level of activity is virtually constant because when a person is listening, that person is very attentive. This effect is also expected to be reflected in the output data of the H.261 video coder. Generally these talking and listening characteristics can be observed daily by observing people. Therefore, these characteristics will be expected to have an impact on visual communication systems.

4.2.3.1 Talking Sequence Analysis

When analysing the talking periods of a videophone sequence several parameters are required : the mean, standard deviation and probability distribution of the talking duration, and the mean, standard deviation, probability distribution and autocorrelation of output data from a talking period.

4.2.3.1.1 The Talking Duration

The sample of recorded videophone sequences given in Table 4.1 were implemented in this analysis. All the sequences were meticulously analysed to isolate the talking and listening periods within each sequence. The sequences vary in duration and also in subject gender, which will permit a comparison of the sexes, i.e. do male talking and listening duration distributions differ from those of females. First though the analysis of all the recorded samples. Illustrated in Fig. 4.4 is the shifted histogram of talking durations(measured in frames) in a videophone sequence where

the shift = -3 frames(*a shift was introduced to allow a better match by a known probability density function, the shift would be re-introduced during model synthesis*). The mean(μ_{TD}) and standard deviation(σ_{TD}) of the shifted talking duration histogram are:

$$\mu_{TD} = 51 \text{ frames},$$
$$\sigma_{TD} = 54.66 \text{ frames}.$$

The histogram in Fig. 4.4 has an exponential characteristic but only after a shift of -3 frames was introduced which was necessary to approximate the talking duration histogram by an exponential probability density function. On simulating a talking duration based upon this analysis the shift must be re-introduced to ensure a true representation of the talking period duration.

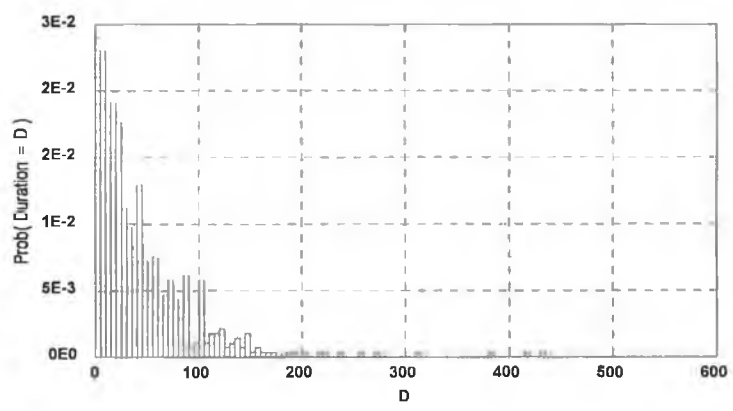


Fig. 4.4 *Videophone Sequence Talk Duration Shifted Histogram with shift = -3 frames.*

Therefore the assumption can be made that the talking duration can be represented by an exponentially distributed random variable that must have 3 frames added during any simulation to represent the true duration, in frames, of the talking duration in videophone sequences. The exponential probability density function is :

$$f(x) = \lambda e^{-\lambda x}, x \geq 0, \tag{4.1}$$

with a parameter λ and a mean = $1/\lambda$. Fig. 4.5 is an exponential curve with the parameter $\lambda = 1/\mu_{TD}$.

To measure the goodness of fit between Figs. 4.4 and 4.5, an error is calculated by the Norm Variation function :

$$error = \sum_{i=0}^{i=N} \left| v_i - \hat{v}_i \right|, \tag{4.2}$$

where N = maximum talking duration in the sample histogram, v_i = the exponential probability density function value for duration i and $\hat{v}_i$ = the shifted histogram value for duration i + *shift*. The Norm Variation function will return an error in the range, $0 \leq error \leq 2$. An error = 0 would obviously indicate a match between the measured histogram and the approximating probability density function. An error = 2 would indicate a total mismatch because the data index, i, for both the measured histogram and the approximating probability density function do not correlate i.e. i would not equal in both v_i and $\hat{v}_i$. Therefore, due to a histogram density = 1 and an approximating probability density function = 1, both add to give an error = 2. The error between Figs. 4.4 and 4.5 was calculated to be 0.337. The exponential density function of Fig. 4.5 thus is a fairly reasonable fit to the histogram of Fig. 4.4.

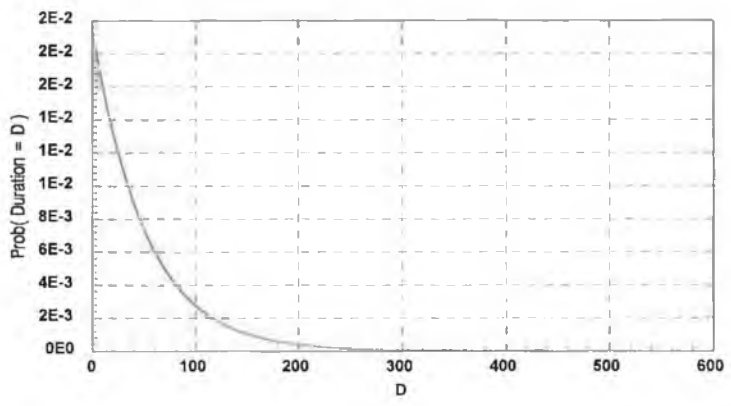


Fig. 4.5 *Exponential Probability Density Function to approximate Fig. 4.4*

A reasonable approximation to the talking duration of the full recorded sample of videophone sequences has been achieved but will the female talking duration distribution differ from that of male talking durations must be investigated. Figs. 4.6 and 4.7 illustrate the shifted talking duration histograms for female talkers and male talkers respectively. Again an exponential function characteristic is clearly evident in both histograms after shifts of -6 frames in Fig. 4.6 and -3 frames in Fig. 4.7,

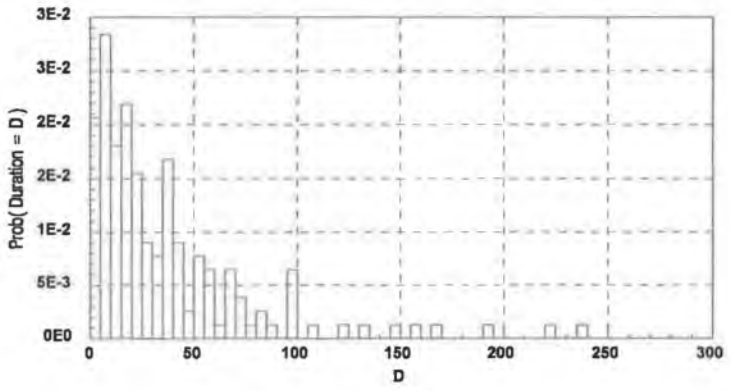


Fig. 4.6 *Talking Duration Shifted Histogram of Female talkers with shift = -6 frames.*

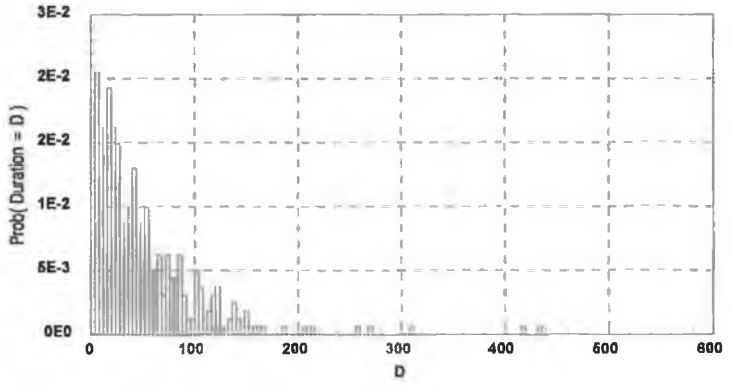


Fig. 4.7 *Talking Duration Shifted Histogram of Male talkers with a shift = -3 frames.*

respectively. On analysis their distribution functions are similar and more importantly similar to Fig. 4.4. The mean and standard deviation of both histograms are given in Table 4.3. In comparison to μ_{TD} and σ_{TD} of Fig. 4.4, the statistical parameters are approximate to each other. Therefore, this statistical analysis further strengthens the shifted histogram in Fig. 4.4 approximation by the exponential probability density function in Fig. 4.5. Hence the talking durations in videophone sequences are exponentially distributed with a shift.

Table 4.3

Statistical Parameter	Female Talk Duration	Male Talk Duration
Mean(frames)	45.00	54.00
Standard Dev.(frames)	49.24	56.4

4.2.3.1.2 The Bits per Talking Frame

The frames corresponding to talking frames were extracted from Fig. 4.3 based on the frame identification process implemented to determine the talking duration distribution in {4.2.3.1.1}. Fig. 4.8 gives the output data produced by concatenated talking frames from which a shifted histogram was obtained(Fig. 4.9).

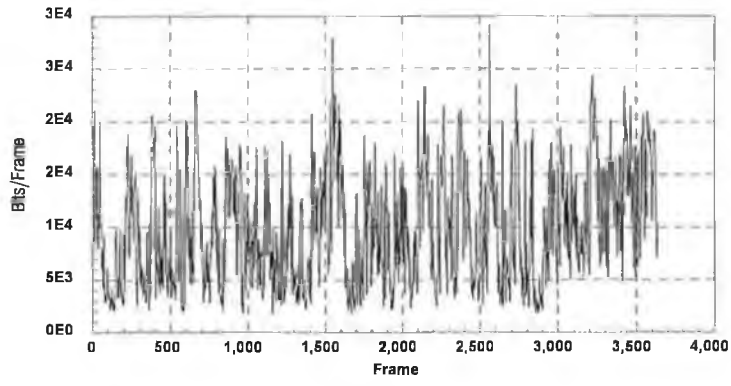


Fig. 4.8 *Concatenated Talking Duration data isolated from Fig. 4.3*

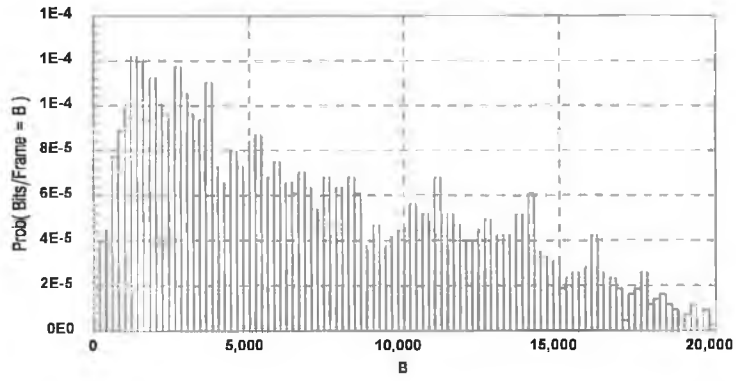


Fig. 4.9 *Bits per Talking Frame Shifted Histogram from Robbie videophone sequence with shift = -1773 bits/frame.*

The mean(μ_{td}) and standard deviation(σ_{td}) of the talking data shifted histogram in Fig. 4.9 were calculated to be

$$\mu_{td} = 8135.799 \text{ bits/frame},$$
$$\sigma_{td} = 5257.89 \text{ bits/frame},$$

after the talking data histogram had been shifted -1773 bits/frame(during any simulation of talking data based on this analysis this shift must be re-introduced). The histogram shift permits a Gamma probability density function approximation[34] which is as follows:

$$f(x) = \frac{\lambda(\lambda x)^{s-1}}{\Gamma(s)} e^{-\lambda x}, x \geq 0 \tag{4.3}$$

where $\Gamma(s)$ is the Gamma function defined as

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t}\, dt .$$

s and λ are called the *shape* and *scale* parameters, respectively, which also determine the Gamma probability density function mean($= s/\lambda$) and variance($= s/\lambda^2$). Therefore, to approximate the shifted histogram in Fig. 4.9, the parameters, $\lambda = \mu_{td}/(\sigma_{td})^2$ and $s = \mu_{td}^2/(\sigma_{td})^2$, give the values

$$\lambda = 0.000294,$$
$$s = 2.394.$$

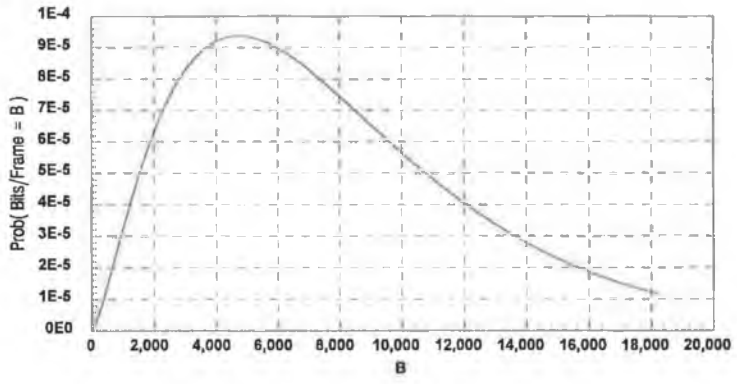


Fig. 4.10 *Gamma Probability Density Function approximation to Fig. 4.9*

Fig. 4.10 is the Gamma probability density function resulting from the parameters λ and s applied to (4.3). The variation between Figs. 4.9 and 4.10 calculated by (4.2) was 0.343 which is a fairly reasonable approximation.

There remains the question of a correlation existing between talking frames, and will the correlation be preserved across different talking period boundaries. Figs. 4.11 and 4.12 are sample autocorrelations of the videophone data in Fig. 4.9. The sample autocorrelations are over the mean talking duration, μ_{TD}, and over a duration $> \mu_{TD}$ to capture several talking periods, respectively. This latter sample autocorrelation is primarily investigative to determine whether talking durations are non-independent. This may be shown by a continuous autocorrelation characteristic across different talking duration boundaries.

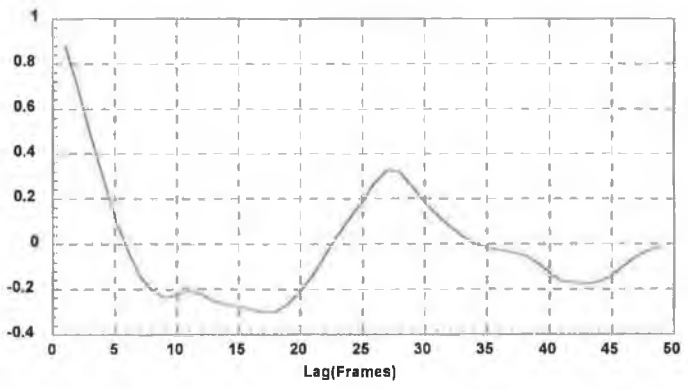


Fig. 4.11 *The Sample Autocorrelation of Talking data over the period* μ_{TD}

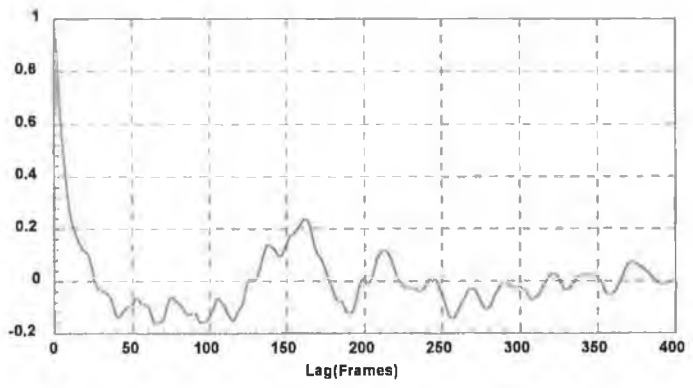


Fig. 4.12 *The Sample Autocorrelation of Talking data over a period* $> \mu_{TD}$

The autocorrelation characteristic in Fig. 4.11 exhibits primarily a Second Order Autoregressive characteristic but within a *lag* = 23 the autocorrelation has a First Order Autoregressive characteristic. Therefore, talking data within a single talking period can be modelled by First or Second Order Autoregressive processes. Fig. 4.12 unfortunately does not reflect the same characteristic to that of Fig. 4.11. The data sequence is largely uncorrelated but a certain level of correlation exists around a *lag* = 150 frames. The cause for this level of correlation at a lag of 150 frames may be attributed to similar levels of motion occurring during talking periods. This is also understandable from the impact motion has on the coding process of a H.261 Video Codec. A stronger correlation exists within a lag of μ_{TD}(*lag* = 0 - 50). Therefore, from this analysis data within a talking period is primarily independent from data within other talking periods.

4.2.3.2 Listening Sequence Analysis

When analysing the listening periods of a videophone sequence several parameters are required : the mean, standard deviation and probability distribution of the listening duration, and the mean, standard deviation, probability distribution and autocorrelation of listening period output data. From the following analysis also, output data characteristics are expected to differ highly from those of talking data as previously mentioned(see {4.2.3}).

4.2.3.2.1 The Listening Duration

A listening duration shifted histogram determined from the sample of videophone type sequences, given in Table 4.1, is shown in Fig. 4.13. The histogram

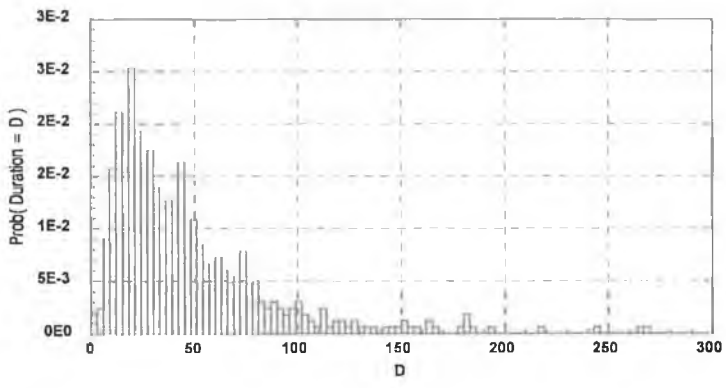


Fig. 4.13 *The Listening Duration Shifted Histogram of Videophone type sequences with a shift = -2 frames.*

was shifted -2 frames(the listening durations are measured in frames) to accommodate
an approximation by a Gamma probability density function(4.3)[34](during any listening duration simulation based on this analysis the shift must be re-introduced). The mean(μ_{LD}) and standard deviation(σ_{LD}) of the listening duration shifted histogram are :

$$\mu_{LD} = 47 \text{ frames},$$
$$\sigma_{LD} = 39.91 \text{ frames}.$$

The parameters required for (4.3), λ and s, to approximate the histogram in Fig. 4.13 are derived from μ_{LD} and σ_{LD}. They are as follows :

$$\lambda = 0.0295,$$
$$s = 1.387,$$

and the Gamma probability density function which results from these parameters is shown in Fig. 4.14. The variation between Figs. 4.13 and 4.14, determined by (4.2), is 0.296. This variation strongly recommends the Gamma probability density function of Fig. 4.14 approximation to the shifted listening duration histogram of videophone sequences in Fig. 4.13.

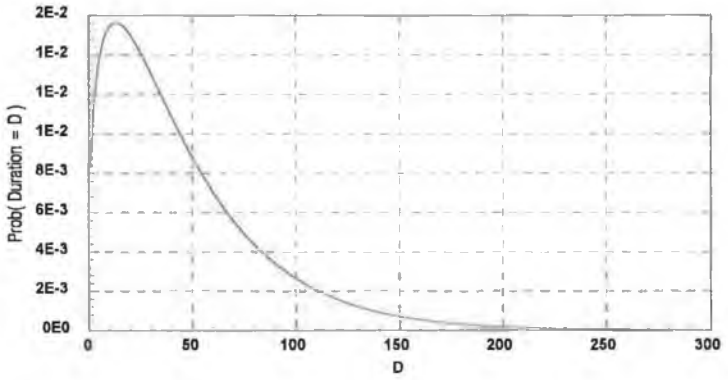


Fig 4.14 *Gamma Probability Density Function to model Fig. 4.13*

To investigate once again whether any difference exists between the sexes in their listening duration characteristics, shifted listening duration histograms for

females and males are illustrated in Figs. 4.15 and 4.16 respectively. The shifts imposed on both histograms were -2 frames and -3 frames, respectively. Statistics for

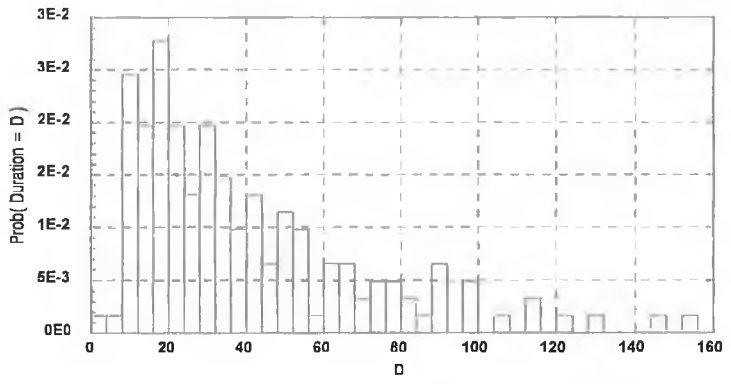


Fig. 4.15 *The Listening Duration Shifted Histogram for Female listeners with a shift = -2 frames.*

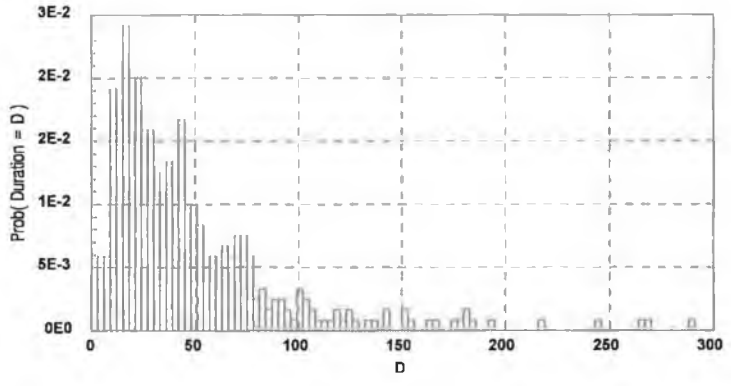


Fig. 4.16 *The Listening Duration Shifted Histogram for Male listeners with a shift = -3 frames.*

Table 4.4

Statistical Parameter	Female Listen Duration	Male Listen Duration
Mean(frames)	42.00	49.00
Standard Dev.(frames)	31.67	42.48

both histograms are shown in Table 4.4. The histograms of Figs. 4.15 and 4.16

display a similar characteristic to the histogram of Fig. 4.13 and their corresponding statistics in Table 4.4 approximate the statistical parameters μ_{LD} and σ_{LD}. Therefore, it can be deduced that the Gamma probability density function of Fig. 4.14 will approximate the listening durations of videophone sequences. Of course the shift imposed on the listening duration histogram must be re-introduced during any simulation to give the true listening duration.

4.2.3.2.2 The Bits per Listening Frame

The data corresponding to listening periods in Fig. 4.3 was extracted and concatenated for separate analysis and is shown in Fig. 4.17. A shifted listening data histogram, shifted by -1322 bits/frame, is illustrated in Fig. 4.18. The mean(μ_{ld}) and standard deviation(σ_{ld}) of the shifted histogram are as follows :

$$\mu_{ld} = 5755.35 \text{ bits/frame},$$
$$\sigma_{ld} = 5344.8 \text{ bits/frame}.$$

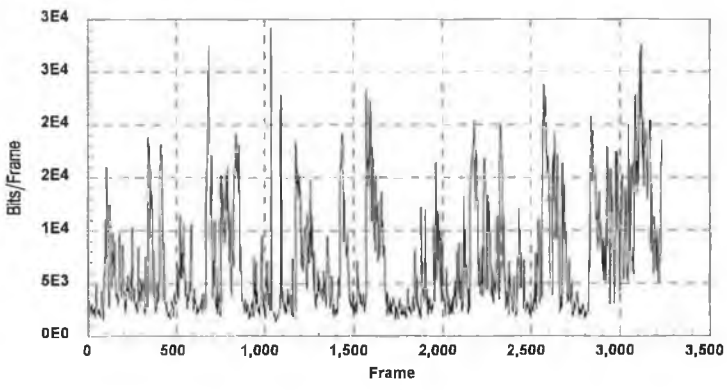


Fig. 4. 17 *Concatenated Listening Duration data isolated from Fig 4.3*

Incorporating a shift into the listening duration data histogram presented the opportunity to approximate the histogram by a Gamma probability density function(4.3) with the parameters

$$\lambda = 0.0002,$$
$$s = 1.1595.$$

Fig. 4.19 illustrates the resulting Gamma probability density function which approximates the shifted listening duration data histogram of Fig. 4.18. The variation, determined by (4.2), between the shifted histogram(Fig. 4.18) and its

approximation(Fig. 4.19) was calculated to be 0.47. As already explained, the Norm Variation function(4.2) returns values in the range 0 - 2(see {4.2.3.1.1}) which signify the degree of mis-match between two distributions. The threshold of

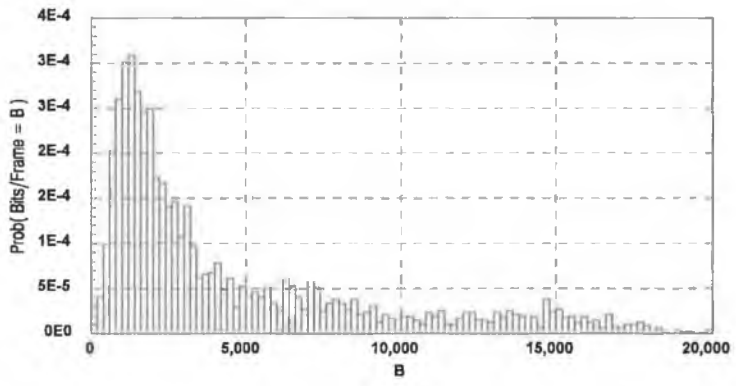


Fig. 4.18 *Shifted histogram of the listen data in Fig. 4.17 with a shift = -1322 bits/frame.*

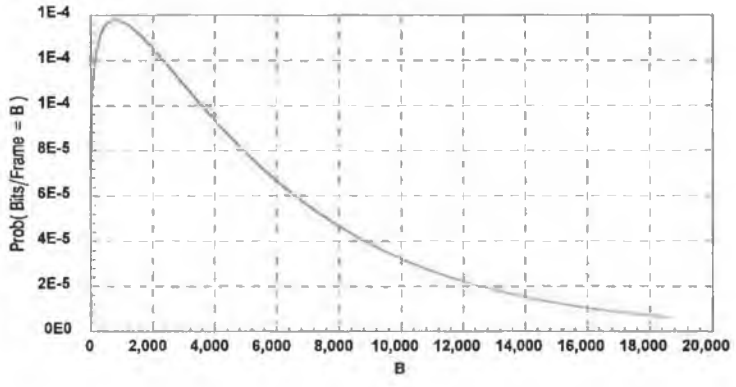


Fig. 4.19 *Gamma Probability Density Function to approximate Fig. 4.18*

acceptability is a degree of mis-match in the range $0 \leq error \leq 0.3$ - 0.4 between two compared probability densities. Therefore, since the variation between Figs. 4.18 and 4.19 is only 0.47, the Gamma probability density function is a slightly weak approximation for the shifted listening duration data histogram. The shifted histogram in Fig. 4.18 though exhibits a strong Gamma probability density function characteristic, hence the approximation will remain.

With a Gamma probability density function to weakly approximate the shifted distribution of listening data in videophone sequences, the data correlation between listening frames must now be considered. Fig. 4.20 is a sample autocorrelation of listening data over the mean listening duration, μ_{LD}, and Fig. 4.21 is the sample

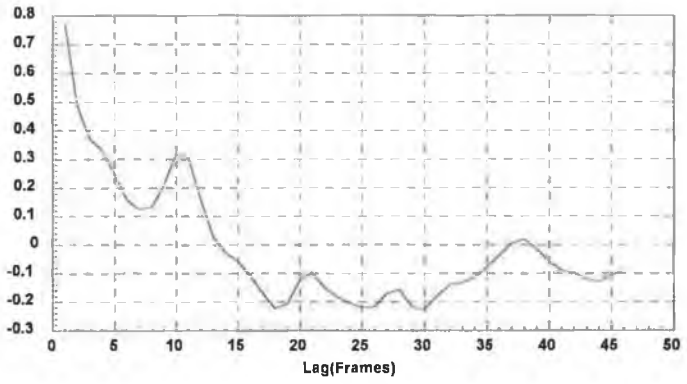


Fig. 4.20 *The Sample Autocorrelation of Listening data over the period* μ_{LD}

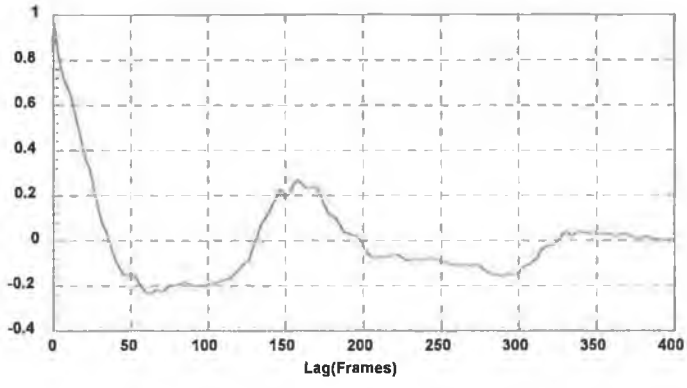


Fig. 4.21 *The Sample Autocorrelation of Listening data over a period* > μ_{LD}

autocorrelation over a period > μ_{LD}. Similar to the talking duration data autocorrelation analysis in {4.2.3.1.2}, the latter sample autocorrelation(> μ_{LD}) is primarily investigative to determine whether data in listening durations is independent. The sample autocorrelation of Fig. 4.20 exhibits primarily a First Order Autoregressive process characteristic within a *lag* = 38. A First Order Autoregressive characteristic can also be detected within a *lag* = 10 but the overall autocorrelation characteristic has more significance because it is present virtually within the full

mean listening duration of a videophone sequence. The sample autocorrelations of Figs. 4.20 and 4.21 unfortunately do not exhibit any similarity. The sample autocorrelation in Fig. 4.21 though detects a level of correlation around a *lag* = 150 frames but similar to the talking data case, the correlation analysis has detected a period of similar motion within listening periods. This effect is again caused by the video coding process of the H.261 Video Codec where similar levels of motion during listening periods in a video sequence result in similar levels of output data. Also, this correlation effect occurs at similar instants in both talking and listening periods(see Figs. 4.12 and 4.21, *lag* = 150) hence the cause is spread across both talking and listening periods within the video sequence. These correlations thus can only be related to periods of similar motion. A strong autocorrelation is present within a *lag* = 0 - 50($< \mu_{LD}$). Therefore, from this analysis listening duration data is primarily independent from data of other listening periods though a certain amount of periodic movement seems to exist during listening periods.

4.2.3.3 Synopsis

The Talk-Listen Strategy, defined in {4.2.3}, originated by adopting a Video Sequence Pre-Video Coding analysis approach to characterise videophone sources. It was assumed that the output data statistical characteristics for talking and listening durations from a 1-Layer VBR H.261 Video Codec would differ significantly. To help verify this assumption a videophone sequence was analysed.

Several approximations were provided by the analysis of videophone sequences under the Talk-Listen strategy. In particular, the talking duration and listening duration distributions were approximated by Exponential and Gamma probability density functions, respectively. Male or female videophone subjects were also found to have similar duration probability distributions. Output traffic distributions from a 1-Layer VBR H.261 Video Codec during talking and listening periods were also approximated. Talking and listening data distributions were satisfactorily approximated by Gamma probability density functions determined by the Norm Variation function(4.2). The approximation was slightly weaker for the listening duration data. This analysis agrees with that of [34] where the distribution of the number of cells(see {2.2.3.1}) per frame was found to be Gamma distributed. The analysis in [34] was carried out on a 30min. videoconference type sequence where different scene types(i.e. talking and listening) were not distinguished. Therefore, one would assume a superposition of the Gamma distributed talking and listening videophone data from the Talk-Listen strategy analysis would result in an overall Gamma distributed data stream. This is a fair deduction because $\mu_{td} \approx \mu_{ld}$ and $\sigma_{td} \approx \sigma_{ld}$ but a longer videophone sequence will be required to prove this assumption(see Fig. 4.39, the histogram of data in Fig.4.3).

The objective of the analysis though was to prove that talking data would have a homogeneously varying characteristic around a particular mean value, and likewise listening data but on a different mean level. The analysis revealed $\mu_{td} > \mu_{ld}$,but not significantly. In comparing Figs. 4.8(concatenated talking duration data) and 4.17(concatenated listening duration data), there is a marginal difference between both data streams which exhibit similar variation in output data($\sigma_{td} \approx \sigma_{ld}$) i.e. the output data exhibits a "spiky" or "bursty" characteristic. Therefore, the output data varies non-homogeneously about their respective mean values in both talking and listening data cases, clearly revealed by the Gamma distribution of both data sets. In relation to correlations within talking and listening periods alone, the sample autocorrelations in Figs. 4.11 and 4.20, respectively, portray a strong indication of output data varying homogeneously about their respective mean values(i.e. the data varies homogeneously about a mean level within a single talking or listening period).

The Talk-Listen Strategy analysis of 1-Layer VBR H.261 Video Codec output data for videophone sequences has shown :

1. the talking duration is Exponentially distributed and the listening duration is Gamma distributed, after imposing histogram shifts;
2. the talking duration data and listening duration data are both Gamma distributed after imposing histogram shifts;
3. data within talking and listening durations is independent;
4. the data within talking and listening durations mainly exhibit a First Order Autoregressive autocorrelation characteristic.

The shifts imposed on each corresponding distribution in this analysis must be re-introduced during any simulation.

It must be remembered also that only one sequence was analysed to make these decisions. Further sequences must be analysed to base these decisions on more concrete foundations. As a further note, in the analysis of talking and listening durations, the process of determining the boundary between talking and listening periods was manual and hence is very subjective.

4.2.4 The Motion Level Strategy

This strategy, as opposed to the Talk-Listen Strategy, is not solely related to a Video Sequence Pre-Video Coding analysis approach of videophone sequences but is more related to a combination of the video codec coding process and Video Sequence Pre-Video Coding analysis. The video coding process is applied at the macroblock level(see {3.2.2}) and the decision to code the macroblock in either Intraframe or Hybrid mode relies on the decision process implemented in the RM8 video coding model(see {3.2.1.7})(this decision technique was implemented in the coding process

applied to the Robbie sequence). Therefore, high output data per frame from the 1-Layer VBR H.261 Video Codec is of consequence due to periods of high motion which subsequently contain a mixture of Intraframe and Hybrid coded macroblock data. The larger majority of macroblocks would be Intraframe coded in this case as a result of the macroblock contents varying(i.e. the luminance and chrominance) under the influence of motion within the video sequence. During periods of low motion in a video sequence, the majority of macroblocks will be Hybrid coded because the luminance and chrominance data of the macroblocks will have remained virtually unchanged, hence the output data per frame will be low. Fig. 4.22 shows a section of output data from Fig. 4.3 which illustrates the transitions from high motion and low motion in the sample video sequence.

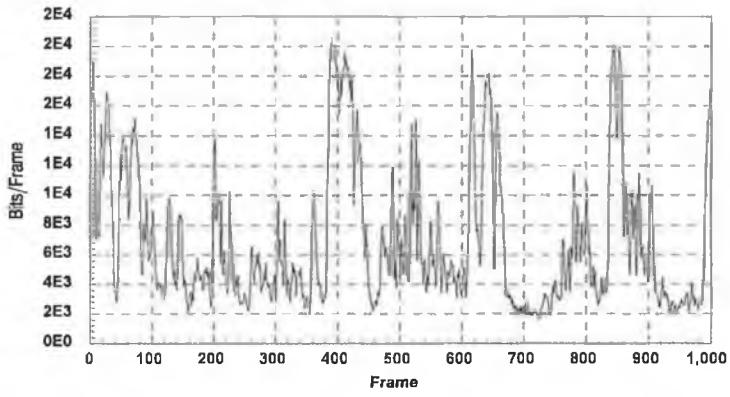


Fig. 4.22 *The High and Low Motion transitional characteristic in Fig. 4.3*

The objective of this strategy is to separate the high and low motion data and analyse this data to determine approximations to model the output data of a 1-Layer VBR H.261 Video Codec for videophone type video sequences. The separation technique employs a threshold, T, and an interval, Δ. The output data mean within the

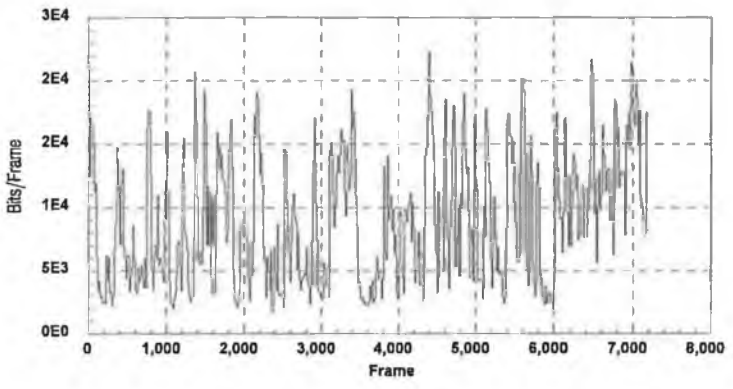


Fig. 4.23 *The output data of Fig. 4.3 smoothed by* $\Delta = 15$ *frames.*

interval Δ, μ_Δ, is calculated and if $\mu_\Delta > T$, the output data corresponds to high motion. Alternatively if $\mu_\Delta \leq T$, the output data corresponds to low motion. A distribution of the high and low motion durations must also be obtained to estimate the duration the output data remains in these periods. Unfortunately, gender comparison cannot be made implementing this strategy due to the availability of only one videophone sequence which is male.

For the following analysis, T = 10 kbits/frame was chosen from Fig. 4.22 to discriminate between high motion and low motion data. The separation process utilised $\Delta = 15$ frames which smoothes the output data of Fig. 4.3 to that in Fig. 4.23.

4.2.4.1 High Motion Sequence Analysis

Analysis of high motion periods in a videophone sequence involves determining probability distributions for the duration present in high motion and the output data associated with those periods. Correlation analysis is also necessary to statistically characterise high motion output data from a 1-Layer VBR H.261 Video Codec.

4.2.4.1.1 High Motion Duration

During the isolation of high motion output data from the data stream in Fig. 4.3, it was also necessary to gather statistics on the duration the output data was in a high motion state. Fig. 4.24 is a shifted histogram of the duration spent in high motion by the output data in Fig. 4.3. The mean(μ_{HMD}) and standard deviation(σ_{HMD}) of the shifted histogram are :

$$\mu_{HMD} = 56 \text{ frames},$$
$$\sigma_{HMD} = 43.69 \text{ frames}.$$

The histogram was shifted -15 frames to accommodate an Exponential probability density function(4.1) approximation with parameter $\lambda = 0.0179$ ($= 1/\mu_{HMD}$). During the simulation of high motion periods based on this analysis, the shift imposed on the high motion duration histogram must be re-introduced to give the proper length of time a videophone source is in high motion. Fig. 4.25 shows the approximation which has a variation of 0.3, determined by (4.2), from the histogram in Fig. 4.24. The variation indicates that a fairly reasonable approximation to the shifted duration histogram a videophone sequence is in high motion has been achieved (during the simulation of a high motion duration based on this analysis, the shift imposed on the histogram in Fig. 4.24 must be re-introduced). By comparing both Figs. 4.24 and 4.25, they fail to approximate each other for High Motion Durations $\geq$ 75 frames. Due to the small size of the sample, an Exponential Probability Density Function is assumed to model the Shifted Histogram in Fig. 4.24 satisfactorily and the error analysis also concurs.

Even though the high motion duration shifted histogram in Fig. 4.24 is reasonably approximated by an Exponential probability density function in Fig. 4.25 the histogram might be better approximated by alternative functions. In [37], scene

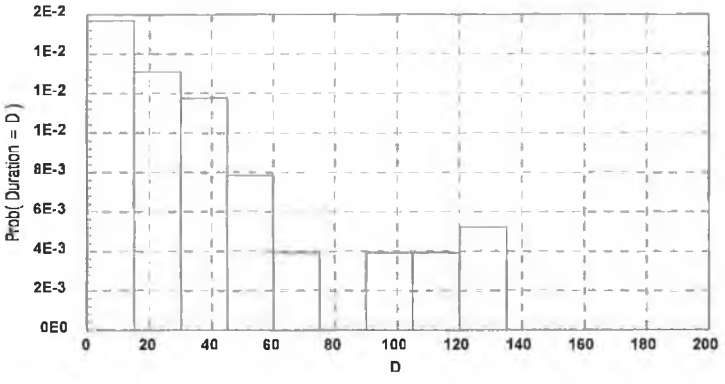


Fig. 4.24 *Shifted Histogram of the duration(in frames) Fig. 4.3 is in High Motion with a shift = -15 frames.*

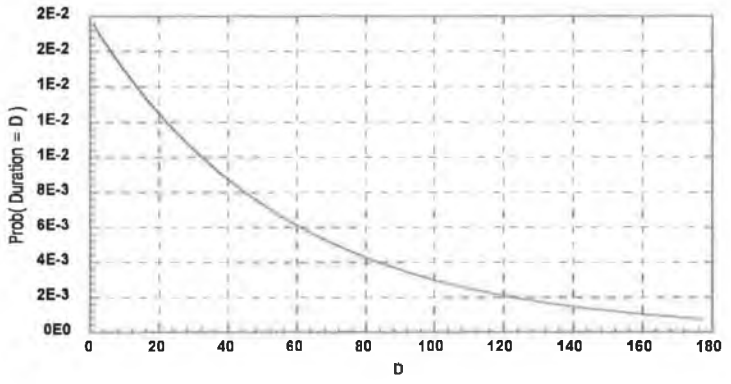


Fig. 4.25 *Exponential Probability Density Function approximation to Fig. 4.24*

lengths(high and low motion periods can be generally regarded as scenes) are characterised by the upper positive section of a Cauchy type distribution which exhibits a constant level distribution at the tail. A Cauchy type distribution could approximate the histogram in Fig. 4.24 more closely because the histogram does not fully decay exponentially. The Cauchy type distribution density function is as follows :

$$f(x) = \frac{a}{x^n + b^2},$$

where a and b are parameters of the function and n has a value close to 2. A further complex and even better approximation to the shifted high motion duration histogram in Fig. 4.24 might be achieved by implementing a combination of two Gaussian probability density functions.

4.2.4.1.2 The Bits per High Motion Frame

The output data corresponding to the high motion periods in Fig. 4.3 was isolated and is presented in Fig. 4.26. A shifted histogram of this data is illustrated in Fig.4.27 which has a mean(μ_{hmd}) and standard deviation(σ_{hmd}) of :

$$\mu_{hmd} = 10141.69 \text{ bits/frame},$$
$$\sigma_{hmd} = 4962.69 \text{ bits/frame}.$$

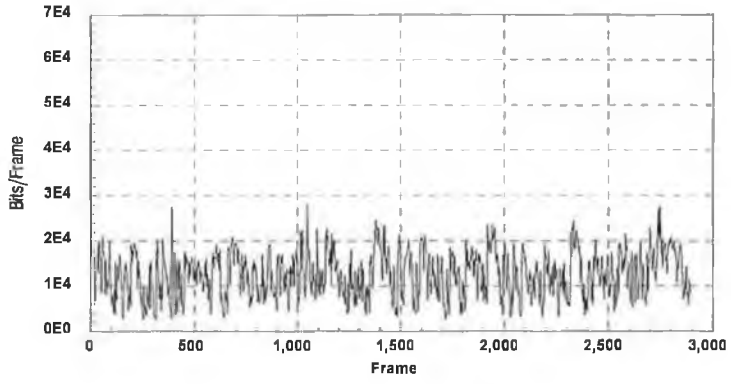


Fig. 4.26 *High Motion data isolated from Fig. 4.3(T = 10 kbits/frame, Δ = 15 frames).*

The histogram was shifted -2549 bits/frame to accommodate a Gaussian probability density function approximation in Fig. 4.28. A Gaussian probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-((x-\mu)/\sigma)^2/2}, \; -\infty < x < +\infty, \tag{4.4}$$

where the parameters μ and σ are the mean and standard deviation of the function, respectively. For the approximation, $\mu = \mu_{hmd}$ and $\sigma = \sigma_{hmd}$, from which there is an approximation variation of 0.218, determined by (4.2). The variation indicates that the Gaussian probability density function of Fig. 4.28 is a suitable match for the shifted distribution of high motion output data from a 1-Layer VBR H.261 Video Codec coded videophone sequence(during the simulation of high motion data based on this analysis, the shift imposed on the high motion data histogram must be re-introduced).

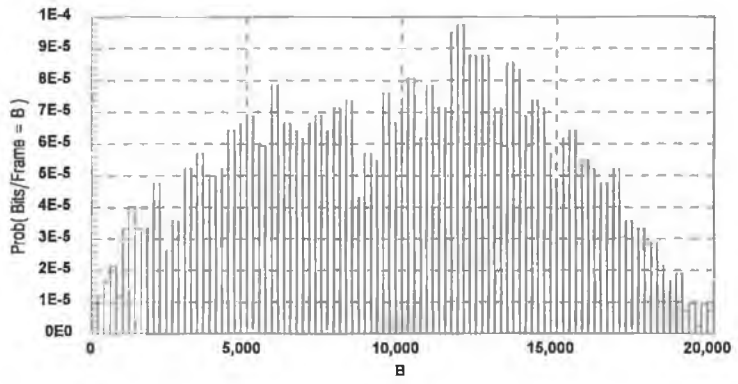


Fig. 4.27 *Shifted Histogram of the data in Fig. 4.26 with shift = -2549 bits/frame.*

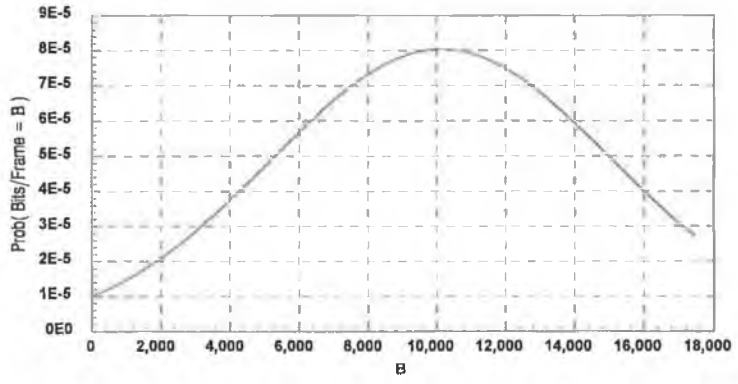


Fig. 4.28 *The Gaussian Probability Density Function approximation to the histogram in Fig. 4.27*

From the statistical analysis on high motion data, the data distribution was found to be Gaussian distributed. In the analysis for correlations in the data of Fig. 4.26, sample autocorrelations were taken over the period μ_{HMD} and over a period > μ_{HMD} which are shown in Figs. 4.29 and 4.30, respectively. The latter sample autocorrelation is primarily investigative to determine whether high motion period data within a single high motion scene is independent. Due to the overall high motion data having a Gaussian distribution, it might be assumed the correlation characteristic would be that of a First Order Autoregressive process. In Fig. 4.30 the output data between different high motion periods is mainly uncorrelated. Certain levels of autocorrelation are present in the data stream of Fig. 4.26 but the overall

autocorrelation characteristic is complex and suggests strongly the data produced within a single high motion period is independent from data of other high motion periods. The short-term sample autocorrelation over the period μ_{HMD} indicates primarily a First Order Autoregressive characteristic as was suggested.

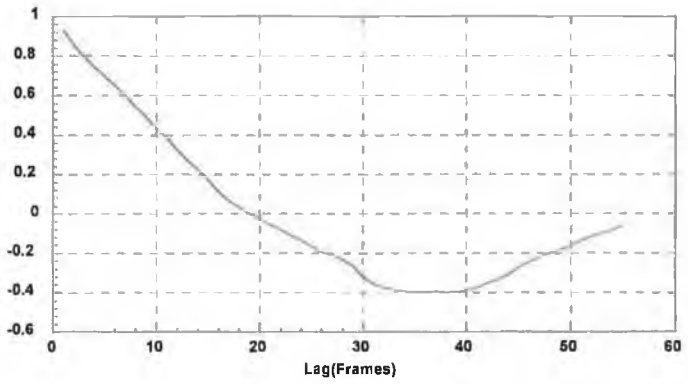


Fig. 4.29 *The Sample Autocorrelation of Fig. 4.26 over a period* μ_{HMD}

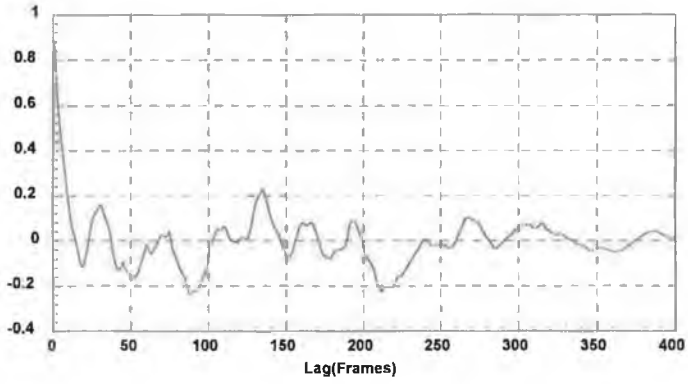


Fig. 4.30 *The Sample Autocorrelation of Fig.4.26 over a period* > μ_{HMD}

4.2.4.2 Low Motion Sequence Analysis

In the low motion sequence analysis, a low motion duration distribution is determined and approximated. The corresponding output data from low motion periods within a 1-Layer VBR H.261 Video Codec, is also analysed for distribution and correlations.

4.2.4.2.1 Low Motion Duration

During the isolation of low motion output data from Fig. 4.3, low motion duration statistics were accumulated. A histogram was compiled of the low motion durations to determine a suitable distribution. The low motion shifted histogram is shown in Fig. 4.31 and was determined to have a mean(μ_{LMD}) and standard deviation(σ_{LMD}) of :

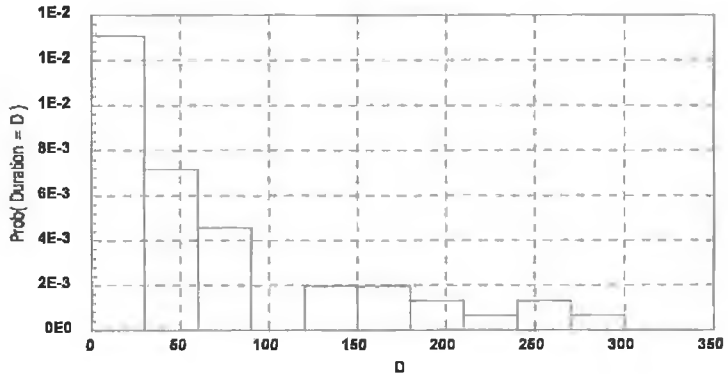


Fig. 4.31 *Shifted Histogram of the Low Motion Duration(in frames) in videophone sequences with a shift = -15 frames.*

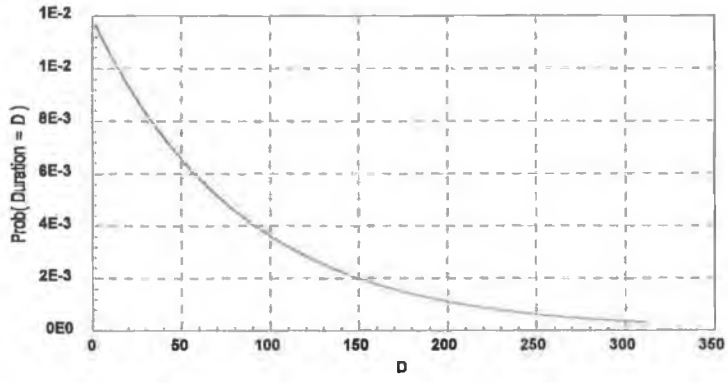


Fig. 4.32 *The Exponential Probability Density Function approximation to Fig. 4.31*

μ_{LMD} = 84 frames,

σ_{LMD} = 83.53 frames.

The histogram was shifted by -15 frames to accommodate an Exponential probability density function(4.1) approximation, with $\lambda = 0.011905$ ($= 1/\mu_{LMD}$)(Fig. 4.32). The goodness of fit test(4.2) produced a variation of 0.293 between Figs. 4.31 and 4.32 which indicates a reasonable fit has been achieved. When simulating a low motion duration based on this analysis, the imposed shift on the low motion duration histogram must be re-introduced.

4.2.4.2.2 The Bits per Low Motion Frame

Fig. 4.33 is the output data isolated from Fig. 4.3 which corresponds to the low motion periods identified from the analysis in {4.2.4.2.1}. To statistically characterise the data, it was analysed to determine its mean, standard deviation, probability distribution and correlation. Fig. 4.34 is a shifted histogram of the low

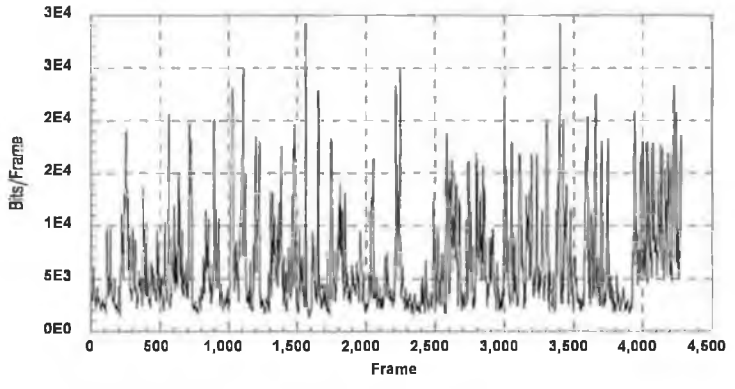


Fig. 4.33 *The Low Motion data isolated from Fig. 4.3*

motion data in Fig. 4.33 with a shift = -1322 bits/frame to accommodate a Gamma probability density function(4.3) approximation. The mean(μ_{lmd}) and standard deviation(σ_{lmd}) of the shifted histogram were found to be :

$$\mu_{lmd} = 5042.89 \text{ bits/frame},$$
$$\sigma_{lmd} = 4584.9 \text{ bits/frame}.$$

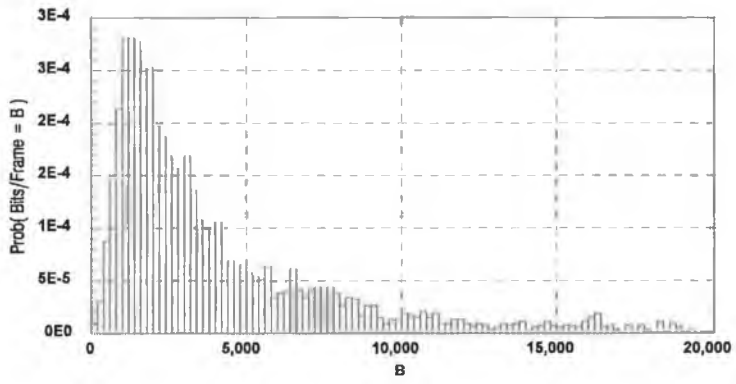


Fig. 4.34 *Shifted Histogram of the data in Fig. 4.33 with shift = -1322 bits/frame.*

Fig. 4.35 shows the Gamma probability density function approximation to the shifted histogram in Fig. 4.34. The goodness of fit test(4.2) gives a variation of 0.396 for Gamma probability density function parameters :

$$\lambda = 0.00024,$$
$$s = 1.21.$$

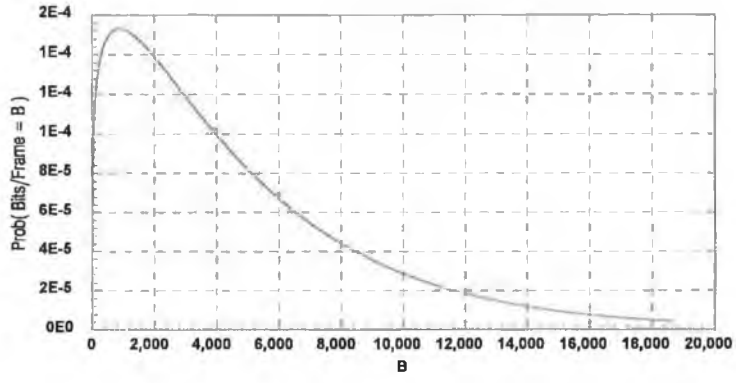


Fig. 4.35 *The Gamma Probability Density Function approximation to Fig. 4.34*

During the simulation of low motion data based on this analysis, the shift imposed on the low motion data histogram must be re-introduced.

With the low motion output data distribution from a 1-Layer VBR H.261 Video Codec for videophone sequences approximated, attention turns to correlation

analysis to determine a process which will characterise the data of low motion periods in a videophone source. Sample autocorrelations were taken over a period μ_{LMD} and over a period > μ_{LMD} and are shown in Figs. 4.36 and 4.37, respectively. Again, the latter autocorrelation is mainly investigative to determine whether low motion data of a single low motion scene is independent. The autocorrelation characteristic of Fig. 4.37 does not bear any resemblance to Fig. 4.36 but it does exhibit certain elements of periodicity within the data of low motion sections of the analysed video sequence. These periodic events also decay to a stationary level but may have remained at a constant stationary level had the autocorrelated data sequence been longer. Overall,

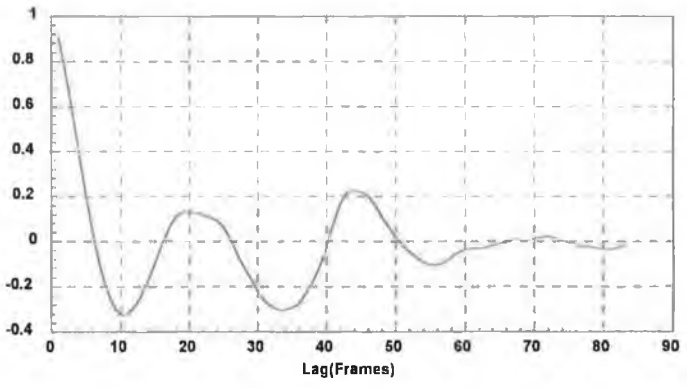


Fig. 4.36 *The Sample Autocorrelation of Low Motion data over the period* μ_{LMD}

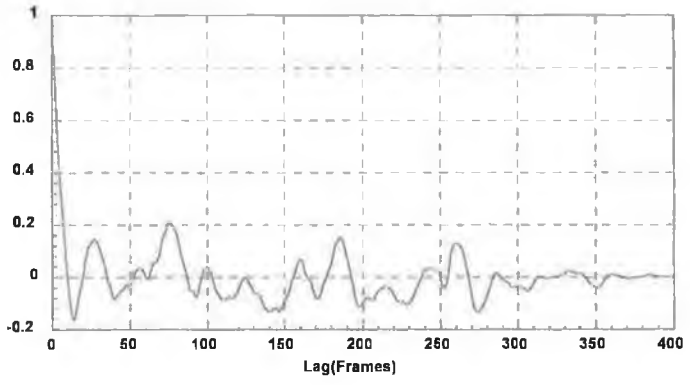


Fig. 4.37 *The Sample Autocorrelation of Low Motion data over a period* > μ_{LMD}

the autocorrelation characteristic is complex and difficult to characterise, therefore low motion period data is assumed to be independent from the data of other low motion periods. The autocorrelation characteristic of Fig. 4.36 on the other hand is more interesting. Fig. 4.36 exhibits a decreasing short-term correlation during a low motion interval which apparently exhibits a Second Order Autoregressive characteristic. Alternatively, the characteristic exhibits a First Order Autoregressive characteristic within a $lag = 20$.

4.2.4.3 Synopsis

The Motion Level strategy combines the knowledge of Video Sequence Pre-Video Coding analysis and the coding process of the 1-Layer VBR H.261 Video Codec where the majority of macroblocks are Intraframe coded during periods of high motion and Hybrid(Intraframe and Interframe) coded during periods of low motion. Hence, the information in each frame of a video sequence will have a mixture of Intraframe and Hybrid coded macroblocks. In the motion level strategy analysis, the high and low motion periods are identified and the corresponding data for each motion type was isolated.

In the analysis of both high and low motion durations, Exponential probability density functions were found to approximate both duration distributions. The variations between shifted histograms and probability density functions determined by the Norm Variation function(4.2) gave clear indication of the strong approximations achieved. The high motion duration shifted histogram in Fig. 4.24 seems to require alternative approximating density functions i.e. Cauchy type or more complex functions, but the applied Exponential probability density function approximates the shifted histogram quite satisfactorily. Also on visually analysing both shifted histograms in Figs. 4.24 and 4.31, both histograms exhibit exponentially decaying characteristics. Therefore, both high and low motion durations can be characterised by an exponentially distributed random variable.

The output data analysis of each motion type also revealed interesting results. The high motion output data shifted histogram was approximated by a Gaussian probability density function and on observing this data in Fig. 4.26 it clearly exhibits a homogeneous varying characteristic. This homogeneous characteristic could possibly be related to the high majority of macroblocks which are Intraframe coded where prediction is not applied during the video coding process. Consequently during periods of high motion the frame information in a video sequence would seem to vary homogeneously about a mean level. The low motion output data in Fig. 4.33 displays larger output data deviations with an obvious floor for minimum output data. By analysing the low motion output data in Fig. 4.33, the data distribution is correctly approximated by a Gamma probability density function. In both data distribution cases the random variables must be positively shifted due to the imposed

shifts applied to their respective histograms to allow approximations by known probability density functions. Also from the analysis $\mu_{hmd} \approx 2\mu_{lmd}$, creating a clear divide between high and low motion states, confirming the method of analysis, but the data from both states exhibit similar variation from the mean level i.e. $\sigma_{hmd} \approx \sigma_{lmd}$.

The autocorrelation characteristics within high and low motion periods exhibit First and Second Order Autoregressive characteristics as shown in Figs. 4.29 and 4.36 respectively. Alternatively, a First Order Autoregressive process may adequately model low motion data instead of implementing a Second Order Autoregressive process. This will be simulated and verified in Chapter 6. The autocorrelation characteristics in Figs. 4.30 and 4.37 do not fully confirm the independence of data within single high and low motion periods but in any case their characteristics are highly complex to model. Therefore the data within each high and low motion state may be assumed independent. The periods of correlation detected over the longer autocorrelated data is assumed to originate from periods of similar motion.

Throughout the motion level strategy analysis approach, the threshold, T, remained constant at 10 kbits/frame and the analysis interval, Δ, remained constant at 15 frames. Further analyses were carried out for Δ = 5, 10, 20 and 25 frames. In all cases, the results were approximate to those obtained for Δ = 15 frames.

Therefore, the Motion Level Strategy analysis of 1-Layer VBR H.261 Video Codec output data for videophone sequences has shown :

1. high motion duration is roughly Exponentially distributed and low motion duration is Exponentially distributed, after imposing histogram shifts;
2. high motion duration data is Gaussian distributed and low motion duration data is Gamma distributed after imposing histogram shifts;
3. data within high and low motion durations is roughly independent;
4. data within high motion durations exhibits a First Order Autoregressive autocorrelation characteristic and data within low motion durations exhibits mainly a Second Order Autoregressive autocorrelation characteristic.

The shifts imposed on each corresponding histogram during this analysis must be re-introduced during any simulation.

4.3 The Long-term Autocorrelation of 1-Layer H.261 Video Codec VBR Data

An interesting and important feature of compressed video sources output traffic is they exhibit different behaviour over short-term and long-term autocorrelations as was discovered(see {4.2.3.1.2},{4.2.3.2.2}, {4.2.4.1.2} and {4.2.4.2.2}). As noted in [38], for some types of video traffic, such as broadcast television, videoconferencing and longer videophone sequences (showing persons talking and listening), one can see : a relatively fast-decaying *short-term correlation* corresponding to a uniform activity level without scene changing and a slow-decaying *long term correlation* corresponding to sudden changes in the gross activity level of the scene (e.g., scene changing in broadcast TV or changes between listener and talker modes in a videophone conversation).

To highlight the existence of long term correlations in videophone sequences, the sample autocorrelation of a section of data from Fig. 4.3 was obtained(Fig. 4.38). The long-term autocorrelation is noticeably non-stationary, i.e. the characteristic does not converge to a stationary state. The presumption was that this characteristic originated from the Talk-Listen process of videophone sequences but could equally well result from the motion in videophone sequences. The Motion Level strategy analysis attempted to describe the origin of the long-term autocorrelation by this

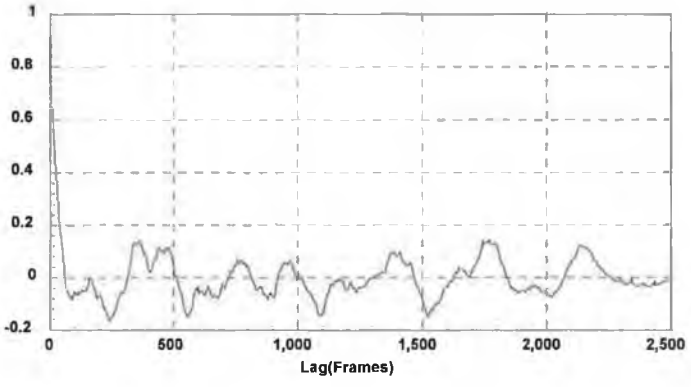


Fig. 4.38 *The Sample Autocorrelation of a section of data from Fig. 4.3*

approach and was equally plausible, if not better. The long-term autocorrelation component may also be the result of periodic movement during the videophone sequence e.g. mouth region motion, head movements, hand movements, etc. Therefore, long-term correlations must be analysed further to determine the reason for their existence to ensure proper characterisation.

4.4 On the parameters to describe H.261 Video Codec Data

The parameters recommended to characterise the output data of a video source[3] are :

- mean bit rate;
- variance;
- peak bit rate;
- burstiness (peak bit rate/mean bit rate);
- quality.

In the analysis to characterise the output data in Fig. 4.3 according to these recommendations, the statistical results in Table 4.2 were obtained. To determine the validity of these statistical parameters, histograms for the output data in Figs. 4.3 and 4.1 were obtained to determine the distribution of 1-Layer VBR H.261 Video Codec data and 2-Layer VBR H.261 Video Codec data. Both data histograms are shown in Figs. 4.39 and 4.40, respectively. From these histograms it is quite difficult to characterise the output data distributions. Fig. 4.39 would require quite a complex probability density function approximation and does not exhibit a Gamma distribution

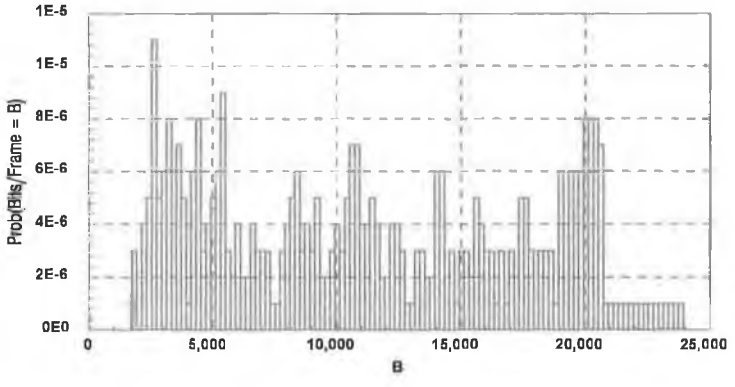


Fig. 4.39 *Histogram of the data in Fig. 4.3(B = Bits/Frame)*

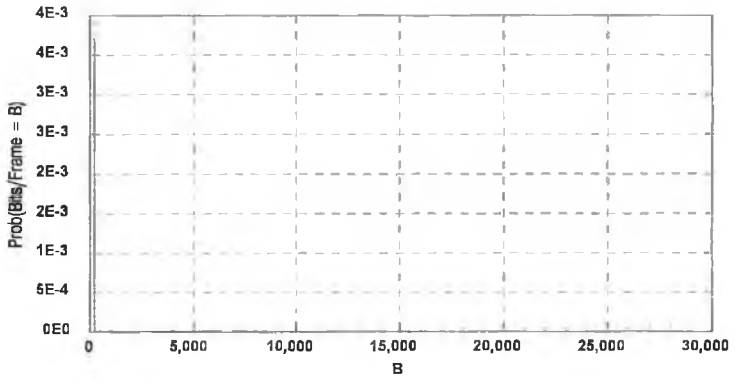


Fig. 4.40 *Histogram of the data in Fig. 4.1*

as was assumed in {4.2.3.3} were the data of both talking and listening durations exhibited Gamma distributions. The histogram in Fig. 4.40 of the data on the second channel of a 2-Layer H.261 Video Codec exhibits largely a Uniform distribution(the large probability of low bits/frame near to B = 0 forces a large vertical scale subsequently causing probabilities for B > 0 to go unnoticed). Therefore, from this analysis it does not help just to determine statistical parameters to characterise the output data from a 1-Layer VBR or a 2-Layer H.261 Video Codec, more information is required to characterise the data as illustrated with this example.

In comparing the recommended analysis approach to both analysis strategies (Talk-Listen and Motion Level) presented in this chapter, both analysis techniques give far superior performances over the recommended technique. This may be deduced from the reasonably successful approximations that were made, being better for the Motion Level strategy than the Talk-Listen strategy but this will not be fully confirmed until the model simulations are discussed in Chapter 6. The statistical parameters derived during the analysis also supported the shifted histogram approximations. Therefore, the analysis strategies presented in this chapter are more accurate in their characterisation of H.261 Video Codec output data. The data analysis approaches adopted have also influenced the modelling approaches presented in the following chapter.

5. VBR Video Source Models

5.1 Origin of Short-term and Long-term Correlations in 1-Layer H.261 Video Codec VBR Data

The short-term and long-term correlations must be defined more precisely. These notions are related to different time resolutions as shown in Fig. 5.1. The *short-term* correlation is the sample autocorrelation function measured within the same scene (it is assumed there is only moderate motion in the scene) where a scene refers to a generalised scene, not a scene in the movie sense. So the short-term correlations describe the behaviour of the sample autocorrelation function only on the burst layer. The mean duration of a scene and the distribution of its duration depends on the video source, but it is usually in the order of 2-5 sec. or more[16]. During this time the codec produces 50-125 frames (for PAL standard video). The sample autocorrelations within the same scene rapidly decrease. During a time less than the mean duration they become negligible. All models without any hierarchical structure{5.2} appropriately describe the codec output traffic within the same scene. In particular, they show rapidly decreasing autocorrelations(see e.g. [8, 20]).

The long-term correlations are related to another time resolution which depends on the scene changing process. They become negligible only after 10-20 scenes or more. The long-term correlations describe the behaviour of the sample autocorrelation function on both burst and scene layers.

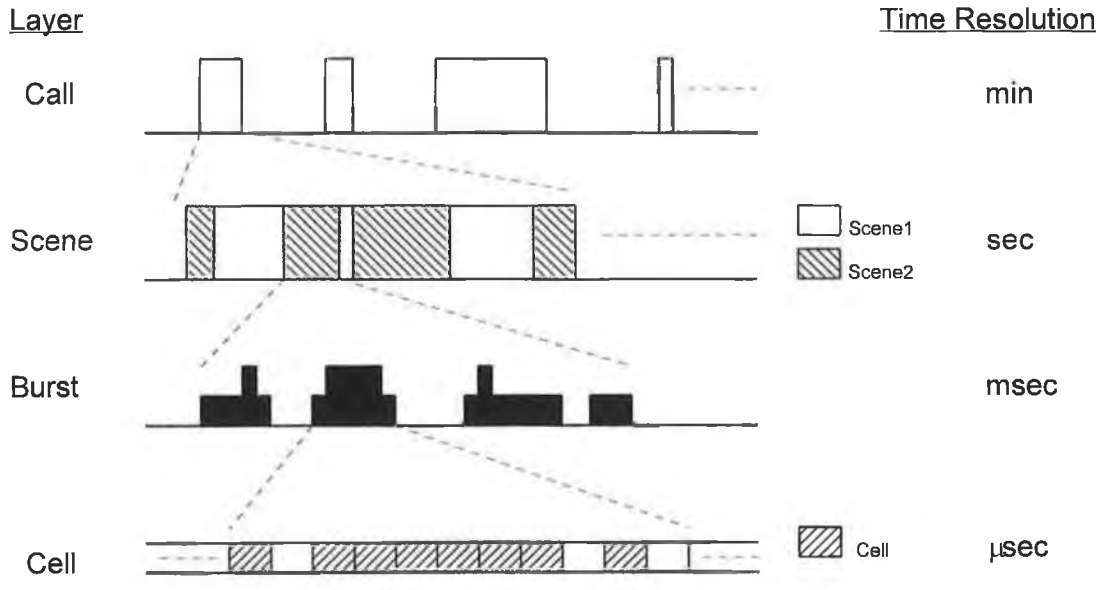


Fig. 5.1 *The Layered and Time Resolution Structure of Videophone Traffic*

Such behaviour of the sample correlation function can be explained heuristically as follows. Suppose for simplicity that there are only two scene types with different statistical characteristics (e.g. stationary distribution, stationary arrival bit rate or whatever). Also suppose a scene of the first type changes to a scene of the second type and vice versa. Let the codec operate on the frame level and begin with a scene of the first type at initial time $t_0 = 0$ (time is discrete with an interframe time interval 0.040 sec). During some time Δt_0, in the order of 1-2 sec., the statistical characteristics of the codec approach the stationary characteristics of the first scene type. The (short-term) autocorrelations (more precisely, the sample autocorrelation function) measured during this period will depend *only* on the stationary characteristics of this scene type. Next, the scene of the first type changes to a scene of the second type at (random) time t_1. During some time Δt_1, in the order of 1-2 secs., the statistical characteristics of the codec again approach the stationary characteristics but of the second scene type. Then again this scene type changes to a scene of the first type and so on.

It is clear that the statistical characteristics of the whole output traffic will approach the stationary distribution after a sufficiently large time t, since, for example, the probability to be a scene of the first or second type at time t, must also be close to their stationary probabilities. Clearly, these probabilities can approach their stationary probabilities only after a sufficiently large number of scene changes. As the mean scene duration is in the order of seconds, then time t, in particular, has to be in the order of a multiple of ten seconds or in minutes. We note that the autocorrelation vanishing speed of a stochastic process(i.e. the autocorrelation characteristic of a stochastic process decays to 0) and its speed of approaching a stationary regime are closely related. Thus, a process, slowly approaching its stationary regime, as a rule, exhibits slow decaying autocorrelations and vice versa. So the sample autocorrelation function (or long-term correlations) measured during this period will exhibit slow decaying autocorrelations[46].

To model these characteristics, a layered stochastic model is required to simulate the layered hierarchy of a videophone sequence. In the following sections, stochastic models are reviewed and new models are introduced based on the analysis in chapter 4 and on the above definitions for short-term and long-term correlations in video sources.

5.2 A VBR Video Source Model Review

5.2.1 General

In the process to characterise VBR video sources(videophone sources in this case) for ATM B-ISDN simulation, the literature was searched to obtain a model which would satisfactorily simulate the characteristics observed in videophone data.

All models found were of the Autoregressive and Markov Chain type and in the majority of cases neglected the long-term autocorrelation component. This was due to the analysis of short duration video sequences and also could possibly be put down to alternative coding algorithms. In [38] a Markov Chain model is presented which attempts to capture the long-term autocorrelation component. In [34] the long-term autocorrelations have been neglected completely because the effects of scene changing in a video sequence were disregarded.

5.2.2 Autoregressive Models

The short-term autocorrelation analysis of videophone/videoconference sources revealed an exponentially decaying autocorrelation[8, 20, 35]. The data in the short-term was also found to be Gaussian distributed[8, 20, 35, 9]. Therefore, with these statistical characteristics combined, they point to the classic First Order Autoregressive process.

5.2.2.1 First Order Autoregressive Model

Let $B(t)$ denote the number of cells in the t-th frame, where $t = 0,1,2,.....$ Then in this model

$$B(t) = \max\{\mu + \alpha B(t\text{-}1) + \beta\omega(t), 0\}$$

where $\omega(t)$ is an independent Gaussian distributed random variable with mean 0 and variance $\sigma^2 = 1$. The constants $0 < \alpha < 1$, $\beta > 0$, $\mu > 0$ are the parameters of the model. This model is often referred to as an AR(1) model[8, 20, 35].

5.2.2.2 Superposition of two First Order Autoregressive Models

This autoregressive model attempts to capture the longer term autocorrelation component inherent in videophone/videoconference sources where the First Order Autoregressive process failed. The model combines a fast decaying First Order Autoregressive process(for the short-term autocorrelations) with a slower decaying First Order Autoregressive process(for the long-term autocorrelations).

Let $B(t)$ denote the number of cells in the t-th frame, where $t = 0,1,2,.....$ Then in this model

$$B(t) = B_1(t) + B_2(t),$$

where $B_1(t)$ and $B_2(t)$ are defined by

$$B_i(t) = \max\{\alpha_i B_i(t\text{-}1) + \beta_i \omega_i(t), 0\}, \quad i = 1,2.$$

$\omega_i(t)$ are independent Gaussian distributed random variables with mean μ_i and variance, $\sigma_i^2 = 1$. The constants α_1, α_2, β_1, β_2 are the parameters of the model[39].

5.2.3 Markov Chain Models

The previous Autoregressive models would be regarded as continuous state models in discrete time. For queueing analysis discrete state continuous time models, such as Markov Chains[49], are more convenient and are also capable of capturing the autocorrelation characteristics of autoregressive processes. Even though queueing analysis requires continuous time Markov chains, they can also be characterised in discrete time. Therefore, the main benefit of Markov Chain models is their flexible implementation in either continuous or discrete time and their simplicity.

5.2.3.1 Birth-Death Continuous Time Video Model (Linear Case)

In this model the continuous time Markov chain has states $\{0,1,....,M\}$(Fig. 5.2). The transition intensities are $\lambda_{i,i+1} = (M-i)\lambda$; $\lambda_{i,i-1} = i\mu$, if $i = 0,1,...,M$ and $\lambda_{ij} = 0$, otherwise. Bit rate in state i is equal to iA. The constants M, λ, μ, A are the parameters of the model.

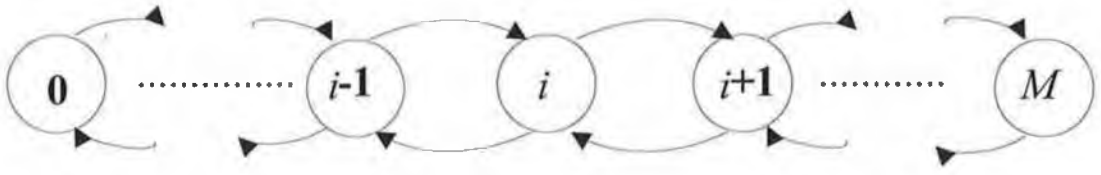


Fig. 5.2 *Birth-Death Continuous Time Video Model with M states*

In fact, this model is a superposition of M identical independent continuous time Markov chains with states $\{0,1\}$ and transition intensities $\lambda_{01} = \lambda$, $\lambda_{10} = \mu$, $\lambda_{00} = \lambda_{11} = 0$(Fig. 5.3). Bit rate in state 1 is equal to A and in state 0 is equal to 0[8].

Heuristic justification of the model. Suppose a frame is divided into M equal blocks which are coded independently(Fig. 5.4). This assumption is applicable for a DCT type codec due to the coding being of a block nature. The bit rate from each block corresponding to a single scene is approximated in the following way. If this bit rate is

Fig. 5.3 *Two State Continuous Time Markov Chain*

greater than the level $\overline{A}$ ($\overline{A}$ is a constant depending on A), then it is approximated by bit rate equal to A and by 0, otherwise. Due to a sufficiently large M such an approximation can be reasonable. Next assume the duration of the *active* period (with bit rate equal A) is exponentially distributed with mean μ and the duration of the *nonactive* period (with bit rate equal 0) is exponentially distributed with mean λ. The crucial assumption in the model is that all blocks change active and nonactive states *independently*.

Due to the blocks in a frame being independent, a sequence of frames will thus produce temporal parallel processes within each block. The Markov Chain within each independent block is a continuous time process therefore to obtain the output data per frame in a video sequence the independent block continuous time Markov Chain must be sampled at discrete times corresponding to frame transmission instances, i.e. every 40msec. in the PAL standard case. At these instants, the states of all independent blocks within the frame are superpositioned to give the output data per frame from the simulated video sequence. This modelling description clearly corresponds to a Birth-Death Continuous Time model which is a superposition of M independent mini-processes.

Alternatively, to emulate the transmission of bufferless real video data the independent blocks within a frame must be transmitted sequentially at discrete times within the duration of a frame. Again the data generating process within each independent block is a continuous time Markov Chain which must be sampled at its transmission instant. The state of the independent block model at this instant is the data transmitted. In comparison to the previous Birth-Death continuous time modelling approach, this method of modelling requires a higher degree of description because sequential block transmission modelling is analytically more demanding. The state of each independent block must be remembered.

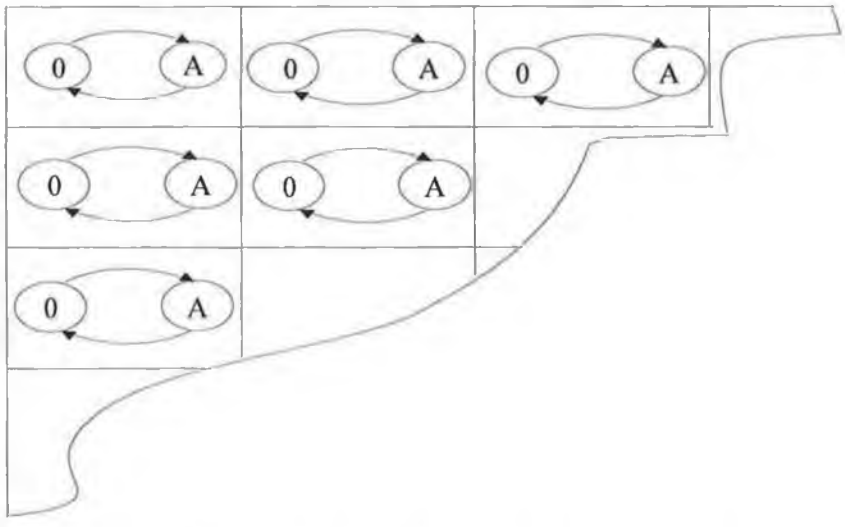


Fig. 5.4 *Section of a Video Frame divided into blocks*

The application of a Birth-Death Continuous Time model to the description of a single scene frame sequence above simplifies the contents of a video frame and is largely unrealistic. The modelling process approximates the contents of a frame by independent blocks but in real video sequences there is dependence between some blocks. For example, if one has a *head + shoulders* picture, all blocks corresponding to the shoulders area or the facial area are dependent. Another example would be the dependence of lower arm movement on upper arm movement and vice versa(if the arms were visible also in the frame). Within other regions of a head + shoulders picture though block independence will hold. For example, background regions far removed from the main subject in a head + shoulders picture are totally independent. Therefore, in order to realise this model realistically one needs to add an interaction between different blocks within regions where these blocks are dependent on each other.

5.2.3.2 Two Dimensional "Birth-Death" Continuous Time Video Model (Linear Case)

This model is a more complex variant of the preceding model and can be described as follows. Consider two Markov chains with states $\{0_1,1_1\}$, $\{0_2,1_2\}$, respectively, and transition intensities $\lambda_{01}^{(n)} = \lambda_n$, $\lambda_{10}^{(n)} = \mu_n$, $\lambda_{00}^{(n)} = \lambda_{11}^{(n)} = 0$, where $n = 1,2$. The bit rate in state 1_n is equal to A_n and in state 0_n is equal to 0.

Let $M = M_1 + M_2$ be the number of blocks. Then the model is a *superposition* of M_1 mutually independent two-state Markov chains with the parameters λ_1, μ_1 and

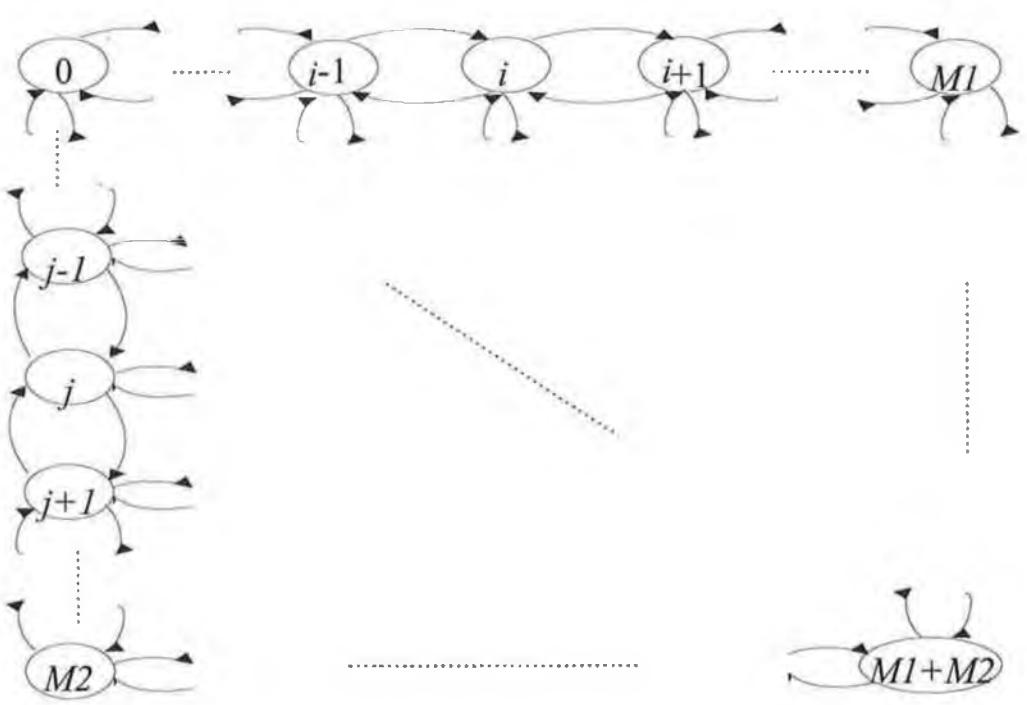


Fig. 5.5 *Two Dimensional Birth-Death Continuous Time Video Model*

M_2 mutually independent two-state Markov chains with the parameters λ_2, μ_2(Fig. 5.5). Both groups of Markov chains are assumed to be independent. Thus, the constants M_1, M_2, λ_1, λ_2, μ_1, μ_2, A_1, A_2 are parameters of the model[38].

Heuristic justification of the model. The difference between this model and the previous one is here M blocks are divided into two groups of blocks with M_1 and M_2 blocks, $M = M_1 + M_2$, respectively. The two groups correspond to a video sequence containing two separate scene types. Scene types are based on different levels of motion and not on movie scenes as such, but more than often a change in motion scene corresponds to a change in movie scene. The two scene types referred to here are high and low motion scenes. Hence this model is also applicable to DCT type video coding. The bit rate from each block of the first group of blocks is approximated by A_1 if the real bit rate is greater than the level $\bar{A}_1$($\bar{A}_1$ is a constant depending on A_1) and by 0, otherwise. The bit rate from each block of the second group of blocks on the other hand is approximated by A_2 if the real bit rate is greater than the level $\bar{A}_2$($\bar{A}_2$ is a constant depending on A_2) and by 0, otherwise(Fig. 5.6).

Next assume that the duration of the *active* period of each block from the first group of blocks (with bit rate equal A_1) is exponentially distributed with mean μ_1 and the duration of the *nonactive* period (with bit rate equal 0) is exponentially distributed with mean λ_1. But the duration of the *active* period of each block from the second group of blocks (with bit rate equal A_2) is assumed to be exponentially distributed with mean μ_2 and the duration of the *nonactive* period (with bit rate equal 0) exponentially distributed with mean λ_2.

The assumptions here are the same as in the previous model : all blocks change active and nonactive states *independently* and, moreover, each block has to remain in its group.

Again, due to the blocks in a frame being independent, a sequence of frames will thus produce temporal parallel processes within each block. The Markov Chain within each independent block is a continuous time process therefore to obtain the output data per frame in a video sequence the independent block continuous time Markov Chain must be sampled at discrete times corresponding to frame transmission instances, i.e. every 40msec. in the PAL standard case. At these instants, the states of all independent blocks within the frame are superpositioned to give the output data per frame from the simulated video sequence. This modelling description clearly corresponds to the Two Dimensional Birth-Death Continuous Time model which is a superposition of $M = M1+M2$ independent mini-processes.

Alternatively, to emulate the transmission of bufferless real video data the independent blocks within a frame must be transmitted sequentially at discrete times within the duration of a frame. Again the data generating process within each independent block is a continuous time Markov Chain which must be sampled at its

transmission instant. The state of the independent block model at this instant is the data transmitted. In comparison to the superposition of independent block models to obtain the output data per frame instant, this method of modelling requires a higher degree of description because sequential block transmission modelling is analytically more demanding. The state of each independent block must also be remembered.

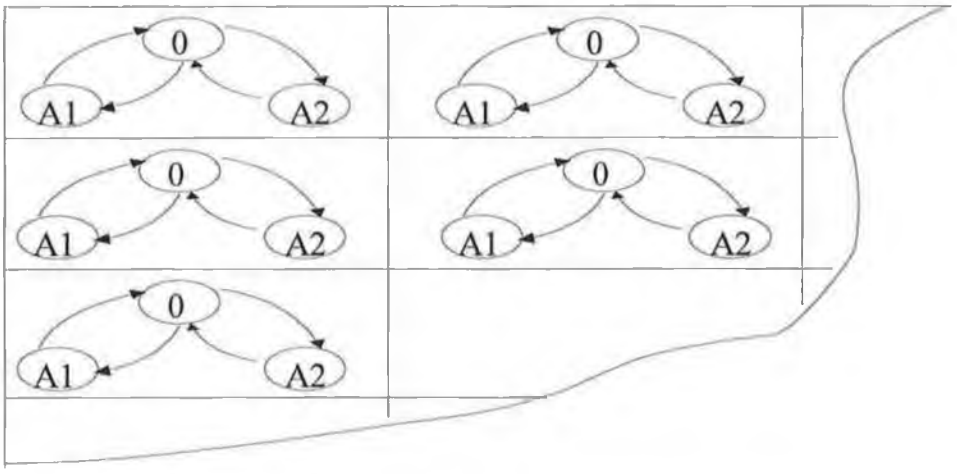


Fig. 5.6 *Section of a Video Frame divided into blocks*

The application of a Two Dimensional Birth-Death Continuous Time model to the description of a two scene frame sequence above simplifies the contents of a video frame and is largely unrealistic. The modelling process approximates the contents of a frame by independent blocks but in real video sequences there is dependence between some blocks as described in the Birth-Death Continuous Time Video Model. Therefore, in order to realise this model realistically also, one needs to add an interaction between different blocks within regions where these blocks are dependent on each other.

5.2.3.3 Birth-Death Discrete Time Video Model (Non-linear Case)

In this model the discrete time Markov chain has states $\{0,1,\ldots, M\}$(Fig. 5.7) where the discrete instants are frame instants. The states of the model contain the output data per frame. Transition probabilities are $p_{i,i+1} = c_1(M\text{-}i)^{\alpha}$; $p_{i,i-1} = c_2 i^{\alpha}$, if $i = 0,1,\ldots, M$ and

$$p_{ii} = 1 - p_{i,i+1} - p_{i,i-1}, \quad i = 1,\ldots, M\text{-}1,$$

$$p_{00} = 1 - p_{01}; \quad p_{MM} = 1 - p_{M,M\text{-}1},$$

and $p_{ij} = 0$, otherwise. Bit rate in state i is equal to iA. The constants α, M, c_1, c_2, A are the parameters of the model. The parameters $c_1 > 0$, $c_2 > 0$ must be small enough

to satisfy the condition : $p_{ii} \geq 0$ for $i = 0,1,....,M$. In [35, 40] the parameter α is chosen to equal 2.

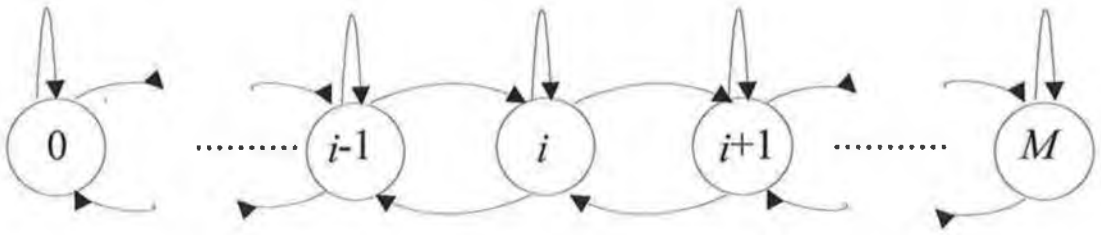


Fig. 5.7 *Birth-Death Discrete Time Video Model*

5.2.3.4 Discrete Autoregressive Video Model : DAR(1) model

A finite-state Markov chain model is used to generate the number of cells associated with each video frame. Jumps between states occur at discrete instants corresponding to discrete frame instants. The transition matrix of the model is given by

$$P = \rho I + (1 - \rho) R,$$

where ρ is the autocorrelation coefficient and I is the identity matrix[34]. In practice, it has been observed[34] that the negative binomial probabilities $(f_0, f_1, ..., f_K, f_K^c)$ defined by

$$f_k = \binom{k+r-1}{k} p^r (1-p)^k = \binom{-r}{k} p^r (p-1)^k, \quad i = 0,1,....,K$$

and

$$f_K^c = 1 - f_0 - f_1 - ..... - f_K,$$

is a good choice for the rows of the matrix R, where the parameter K represents the maximum number of cells that can be generated by the codec for a single video frame, r and ρ are derived from the mean rate m of the video sequence and its variance σ^2, via the relations :

$$m = \frac{r(1-p)}{p},$$
$$\sigma^2 = \frac{r(1-p)}{p^2}.$$

5.2.4 Discussion

The above reviewed models were necessary step by step approaches to VBR videophone modelling. The First Order Autoregressive model in {5.2.2.1} is a traditional and accepted engineering model. Therefore, on discovering an exponentially decaying autocorrelation in videophone type sources[8, 20, 35], the obvious model to apply was a First Order Autoregressive model. Subsequent modelling approaches branched from this fundamental model e.g. Birth-Death Continuous Time Video Model, and were tuned to exhibit a similar autocorrelation to that of a First Order Autoregressive process.

In most cases though the coding processes utilised were of a non-H.261 Video Coding Standard, the analysed video sequences were short in duration(10-20secs.) and scene changing was not an issue. Consequently, the full characteristics of video sources were neglected. In order to capture these neglected characteristics that were uncovered in the video source analysis carried out in chapter four, hierarchical models are introduced below. Hierarchical models may utilise these reviewed models, because they are effectively models of data processes associated with particular scene types. The models may also be of benefit in characterising the higher layers in hierarchical models.

5.3 A VBR Video Source Hierarchical Model

5.3.1 The Natural Hierarchy of a Video Source

As already mentioned, a layered description of ATM cell traffic depending on the time resolution is very natural(see Fig. 5.1) [41, 42]. The model of an ATM source can be observed on call layer, on burst layer or on cell layer. According to this, the time resolution differs in orders of magnitude, that is, in minutes, milliseconds or microseconds. Additionally, VBR video sources provide another natural layer - scene layer with the time resolution in the order of seconds. A scene in this case is not a scene in the strictest sense of the word i.e. a television scene, theatre scene, etc. The scene layer generally refers to periods of varying activity within a video sequence which is reflected in the video codec output data. In most cases though, a change in activity corresponds to a change of scene in the video sequence i.e. this can be observed in television sequences in particular.

One can see that in compressed video sources there are two processes of different origin. The first process describes scene changing and scene duration. Basically, it is related to exterior activity with respect to the video codec (e.g., human activity - scene changing by camera-man, movie editing and so on). The second process is output traffic from the compressed video source which describes the

encoding algorithm activity during each scene. Due to these reasons, it is natural to separate these processes.

Thus, in order to construct an appropriate model for a compressed video source the following must be defined :

(i) a collection of similar and different scene types(generalised scenes that is);
(ii) a scene changing process between the different scene types;
(iii) a process to describe scene length for each scene type;
(iv) a process to describe traffic output for each scene type;
(v) a procedure for choosing an initial scene type and an initial state of output traffic for each scene type;

(i). The scenes are divided into different types. Each type is characterised by a traffic output process. For example, the different scene types can be defined by statistical characteristics, (e.g. the mean bit rate during a scene[37]). Scenes with the same statistical characteristics are then off the same scene type.

(ii). Generally, this process is a stochastic one. But in some special situations it can be deterministic. For example, for a videoconferencing source the process of changing between different scene types (e.g. between a person and a graphical information display, both containing moderate motion) is very simple: scenes of the first and second type alternate between each other.

(iii). Each given scene type has a random length which can be defined by a proper distribution function. The distribution function depends on the kind of video model (videophone, movie, videoconferencing), variety of scene types, etc.

(iv). For each scene type the output traffic must be specified. In fact, most papers (see e.g. [38, 8, 20, 43, 9]) were devoted to modelling the output traffic for only one scene type. There is a review of such models in {5.2}.

(v). The procedure to choose an initial scene type must be described. This choice can be deterministic or stochastic and can be modelled by a proper distribution function defined on the set of scene types.

In order to obtain a more accurate model, another important feature of described video sources must be accounted for. At the beginning of a scene there is a sufficiently large jump in the bit rate(see Fig. 4.3). This effect is related to the fact that the first frames of each scene are totally encoded using an *intraframe* technique

while other frames are encoded using a *hybrid*(interframe and intraframe) technique. So each scene has to be divided into two stages such that during the first and second stages of each scene the codec implements intraframe and hybrid encoding, respectively.

5.3.2 A Formal Description of the Natural Hierarchy

Formally, the compressed video source hierarchical model can be described as follows. Assume there is a set of different scene types $\{S^{\theta}, \theta \in \Im\}$, where $\Im$ is an index set. In general, the set $\Im$ can be finite, countable or some domain in $\mathbf{R}^d$, where d is the number of parameters which determine the output traffic process $B_{\theta}(t)$, corresponding to scene type θ in a family of processes. For example, if $B_{\theta}(t)$ is the Poisson process, then $d = 1$ and $\theta = \lambda$ is its intensity. The choice of initial scene type is defined by a distribution function $F^{\Im}$ on the set $\Im$.

The scene types alternate between each other in some random or deterministic order which is modelled by a stochastic process $\hat{T}(n)$, $n = 0,1,....$ defined on $\Im$ and denoting the type of the n-th scene. $T(t)$ denotes the type of scene at time t, where time t is assumed to be continuous or discrete. For each $\theta \in \Im$, two random variables $L_{\theta,1}$, $L_{\theta,2}$ describe the durations of the first and second stages of the scene of type S^{θ}. The first stage of scene type S^{θ} corresponds to totally intraframe coded video data due to a change of scene in the video sequence while the second stage of S^{θ} is hybrid coded. To each scene type S^{θ} there corresponds stochastic processes $B_{\theta,1}(t)$, $B_{\theta,2}(t)$ which describe the output traffic during the first and second stages of the scene S^{θ}, respectively. Distribution functions $F^{in}_{\theta,1}$, $F^{in}_{\theta,2}$ determine the random choice of an initial state of output traffic for the first and second stages of scene S^{θ}, respectively. The process $B(t)$, combining $B_{\theta,1}(t)$ and $B_{\theta,2}(t)$, describes the output bit rate at time t or the number of cells in the t-th frame.

5.3.3 Method of Hierarchical Model Simulation

At initial time t_0 one chooses at random (with the distribution function $F^{\Im}$) a parameter $\theta_0 = \hat{T}(0) \in \Im$. Then during a random time interval $s_0 = L_{\theta,1}$ from t_0 to t_0+s_0 the process $B(t) = B_{\theta,1}(t)$ produces traffic beginning from some initial state $B_{\theta,1}(t_0) = b_{\theta,1}$ chosen with the distribution function $F^{in}_{\theta,1}$. During a random time interval $r_0 = L_{\theta,2}$ from t_0+s_0 to $t_1 = t_0+s_0+r_0$ the process $B(t) = B_{\theta,2}(t)$ produces traffic beginning from some initial state $B_{\theta,2}(t_0+s_0) = b_{\theta,2}$ chosen by the distribution function $F^{in}_{\theta,2}$. At time t_1 again one chooses a parameter $\theta_1 = \hat{T}(1)$ and random time intervals $s_1 = L_{\theta,1}$, $r_1 = L_{\theta,2}$ and so on.

5.3.4 Hierarchical Video Source Models

5.3.4.1 General

Here some examples of video source models are considered. To describe the process $B(t)$, the model must be specified completely. In all what follows $L_{\theta,1} \equiv 0$ is assumed, i.e. for simplicity the first stage of each scene is omitted. Therefore $L_\theta = L_{\theta,2}$, $B_\theta = B_{\theta,2}$, $F_\theta^{in} = F_{\theta,2}^{in}$.

5.3.4.2 Continuous Time Models

Here continuous time video source models are described which are specified on both burst and scene layers.

(i) *Markov Video source model* (*CMV model*).

The process $T(t)$ is a homogeneous continuous time Markov chain with states $\mathfrak{I} = \{1,....,k\}$ (k is finite) and transition intensity matrix $Q = \{\lambda_{\theta\theta'}\}_{\theta,\theta'=1,....,k}$. Initial choice of $T(0)$ is defined by a vector $(p_1,....,p_k)$,

$$p_\theta \geq 0, \quad p_1 + ... + p_k = 1,$$

where p_θ is the probability to choose scene type S^θ, $\theta \in \mathfrak{I}$. For each $\theta \in \mathfrak{I}$, the process $B_\theta(t)$ is a homogeneous continuous time Markov chain with states $\{1,...., n_\theta\}$ and transition intensity matrix $Q_\theta = \{\gamma^\theta_{mm'}\}_{m,m'=1,...., n_\theta}$ which produces a bit rate $b^\theta(m) \geq 0$ in state $m \in \{1,...., n_\theta\}$, where $b^\theta(\cdot)$ is a (non random) bit rate function. The distribution function F_θ^{in} describes the random choice of initial bit rate for scene type S^θ. $F_\theta^{in}(m)$ is the probability to choose the state $m \in \{1,...., n_\theta\}$. Note that the random variable L_θ is implicitly described. It is exponentially distributed with parameter

$$\sum_{\substack{\theta'=1 \\ \theta' \neq \theta}}^{k} \lambda_{\theta\theta'}.$$

(ii) *Alternating Markov video source model* (*or CAMV model*).

In this case k = 2(Fig. 5.8) for the previous model and the process

$$\hat{T}(n+1) = \begin{cases} 1, & \text{if } \hat{T}(n) = 2 \\ 2, & \text{if } \hat{T}(n) = 1 \end{cases}$$

for $n = 0,1,....$ Initial choice of $\hat{T}(0)$ is defined by a vector (p_1, p_2),

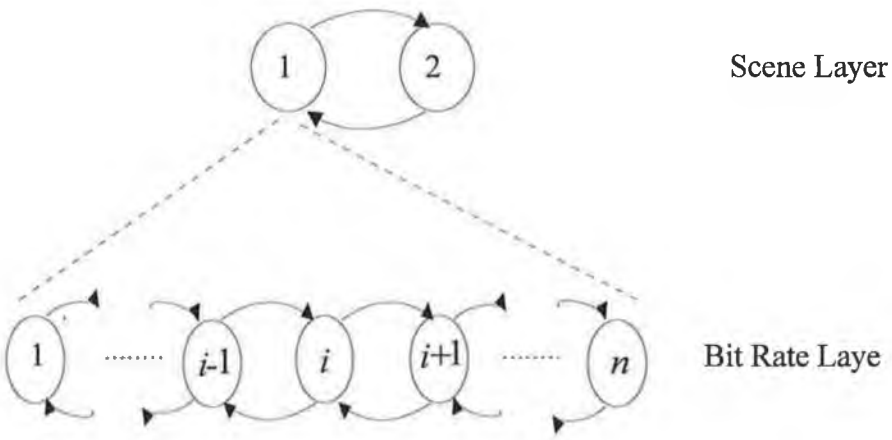


Fig. 5.8 *Continuous Time Markov Video Source Model with k = 2*

$$p_1 \geq 0, \quad p_2 \geq 0, \quad p_1 + p_2 = 1,$$

where p_θ is the probability to choose scene type S^θ, $\theta = 1,2$.

(iii) *Semi-Markov Video source model* (*or CSMV model*).

The process $\hat{T}(n)$ is a homogeneous discrete time Markov chain with states $\Im = \{1,....,k\}$(k is finite) and with transition matrix $P = \{ p_{\theta\theta'}\}_{\theta,\theta' = 1,....,k}$. Initial choice of $\hat{T}(0)$ is defined by a vector $(p_1,....,p_k)$,

$$p_\theta \geq 0, \quad p_1 + ... + p_k = 1,$$

where p_θ is the probability to choose scene type S^θ, $\theta \in \Im$. Random variable L_θ has a positive distribution function G_θ, $\theta \in \Im$. For each scene type S^θ the process $B_\theta(t)$ is a homogeneous continuous time Markov chain with states $\{1,...., n_\theta\}$ and transition intensity matrix $Q_\theta = \{\gamma^\theta_{mm'}\}_{m,m' = 1,...., n_\theta}$ which produces a bit rate $b^\theta(m) \geq 0$ in state $m \in \{1,...., n_\theta\}$. The distribution function F^{in}_θ describes the random choice of .

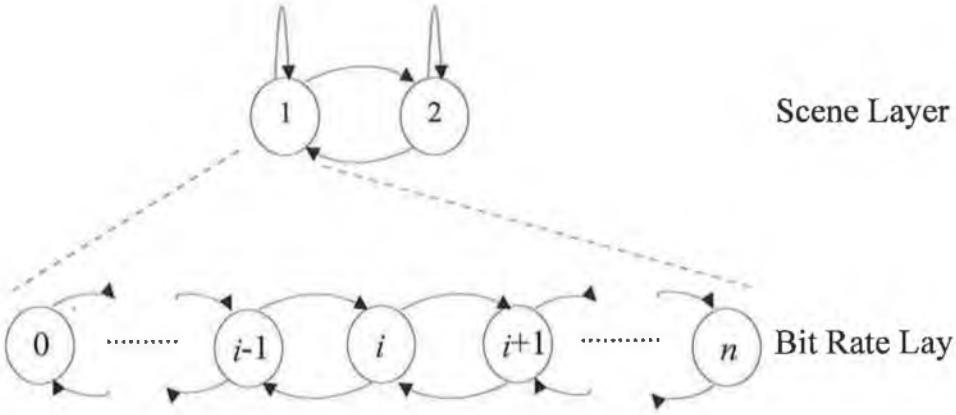


Fig. 5.9 *Continuous Time Semi-Markov Video Source Model with k = 2*

initial bit rate for scene type S^{θ}. $F_{\theta}^{in}(m)$ is the probability to choose the state $m \in \{1,...., n_{\theta}\}$

(iv) *Alternating Semi-Markov Video source model (or CASMV model).*

In this case k = 2(Fig. 5.9) for the previous model and the process

$$\hat{T}(n+1) = \begin{cases} 1, & \text{if } \hat{T}(n) = 2 \\ 2, & \text{if } \hat{T}(n) = 1 \end{cases}$$

for $n = 0,1,.....$ Initial choice of $\hat{T}(0)$ is defined by a vector (p_1, p_2),

$$p_1 \geq 0, \quad p_2 \geq 0, \quad p_1 + p_2 = 1,$$

where p_{θ}, is the probability to choose scene type S^{θ}, $\theta = 1,2$.

(v) *Non-Markov Video source model (or CNMV model).*

Let $\Im$ be a domain in $\mathbf{R}^d$ and let $F^{\Im}$ be a probabilistic measure on $\Im$. The type of each new scene, including initial choice of scene type, is chosen independently of the previous scene types and randomly with the measure $F^{\Im}$. Random variable L_{θ} has a distribution function G_{θ}, $\theta \in \Im$. For each scene type S^{θ} the process $B_{\theta}(t)$ is a homogeneous continuous time Markov chain with states $\{1,...., n_{\theta}\}$ and transition intensity matrix $Q_{\theta} = \{\gamma^{\theta}_{mm'}\}_{m,m' = 1,...., n_{\theta}}$ which produces a bit rate $b^{\theta}(m) \geq 0$ in state $m \in \{1,...., n_{\theta}\}$. The distribution function F_{θ}^{in} describes the random choice of initial bit rate for scene type S^{θ}. $F_{\theta}^{in}(m)$ is the probability to choose the state $m \in \{1,...., n_{\theta}\}$(Fig. 5.10).

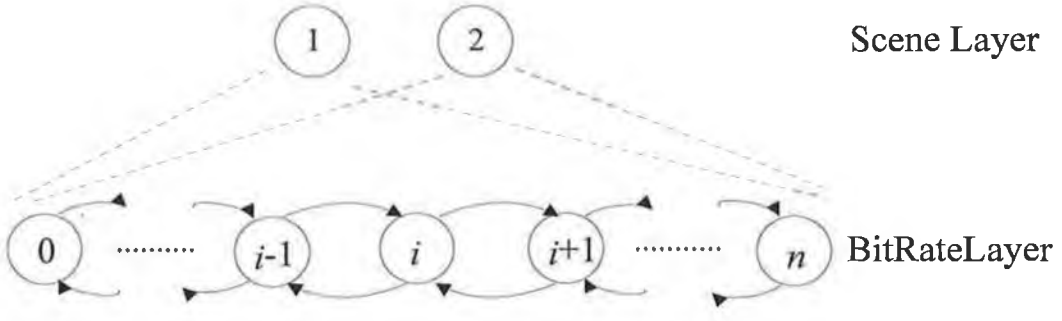


Fig. 5.10 *Continuous Time Non-Markov Video Source Model*

5.3.4.3 Discrete Time Models

Here discrete time video source models are described which are also specified on both burst and scene layers.

(i) *Markov Video source model* (*or DMV model*).

The process $T(t)$ is a homogeneous discrete time Markov chain with states $\mathfrak{I} = \{1,...., k\}$($k$ is finite) and transition matrix

$$P = \{p_{\theta\theta'}\}_{\theta,\theta' = 1,...., k}.$$

For each $\theta \in \mathfrak{I}$, the process $B_\theta(t)$ is a homogeneous discrete time Markov chain with states $\{1,...., n_\theta\}$ and transition matrix

$$P_\theta = \{p^{\theta}_{mm'}\}_{m,m' = 1,...., n_\theta}$$

which produces a bit rate $b_\theta(m) \geq 0$ in state $m \in \{1,...., n_\theta\}$, where $b_\theta(\cdot)$ is a bit rate (non random) function. The distribution function F^{in}_θ describes the random choice of initial bit rate for scene type S^θ. $F^{in}_\theta(m)$ is the probability to choose the state $m \in \{1,...., n_\theta\}$.

(ii) *Semi-Markov video source model* (*or DSMV model*).

The process $\hat{T}(n)$, $n = 1,2,....$ is a homogeneous discrete time Markov chain with states $\{1,...., k\}$ and transition matrix $P = \{p_{\theta\theta'}\}$. Initial choice of $\hat{T}(0)$ is defined by a vector $(p_1,...., p_k)$,

$$p_\theta \geq 0, \quad p_1 + ... + p_k = 1,$$

where p_θ is the probability to choose scene type S^θ, $\theta \in \mathfrak{I}$. Random variable L_θ has a distribution function G_θ, $\theta \in \mathfrak{I}$. For each $\theta \in \mathfrak{I}$, the process $B_\theta(t)$ is a homogeneous discrete time Markov chain with states $\{1,...., n_\theta\}$ and transition matrix $P_\theta = \{p^{\theta}_{lm}\}$.

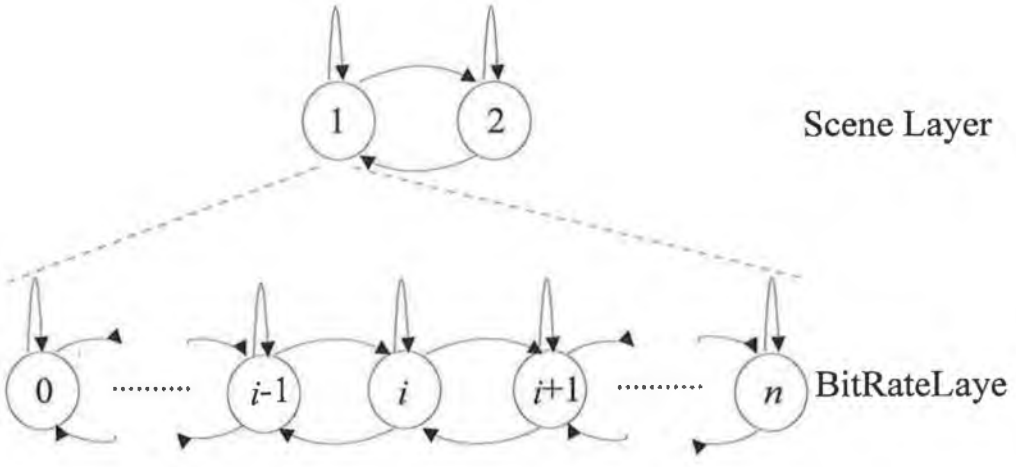


Fig. 5.11 *Discrete Time Markov/Semi-Markov Video Source Model with k = 2.*

which produces a bit rate $b_\theta(m) \geq 0$ in state $m \in \{1,...., n_\theta\}$, where $b_\theta(\cdot)$ is a bit rate (non random) function. The distribution function F_θ^{in} describes the random choice of initial bit rate for scene type S^θ. $F_\theta^{in}(m)$ is the probability to choose the state $m \in \{1,...., n_\theta\}$

(iii) *Alternating Semi-Markov Video source model* (*or DASMV model*).

In this case k = 2(Fig. 5.11) for the previous model and the process

$$\hat{T}(n+1) = \begin{cases} 1, & \text{if } \hat{T}(n) = 2 \\ 2, & \text{if } \hat{T}(n) = 1 \end{cases}$$

for $n = 0,1,.....$ Initial choice of $\hat{T}(0)$ is defined by a vector (p_1, p_2),

$$p_1 \geq 0, \quad p_2 \geq 0, \quad p_1 + p_2 = 1,$$

where p_θ is the probability to choose scene type S^θ, $\theta = 1,2$.

(iv) *Autoregressive video source models* (*or DAV model*).

The process $T(t)$ is a homogeneous discrete time Markov chain with states $\Im = \{1,...., k\}$ and transition matrix $P = \{p_{\theta\theta'}\}$(Fig. 5.12). For each $\theta \in \Im$, $B_\theta(t)$ is an autoregressive process of the first or second order, $B_\theta(-1) = B_\theta(-2) = 0$.

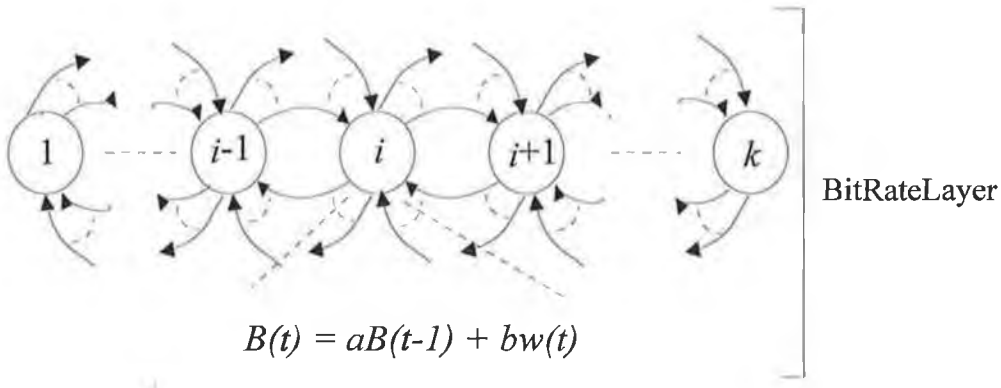


Fig. 5.12 *Discrete Time Autoregressive Video Source Model.*

5.3.4.4 Structurally Stable Fluid Video Models

Here *structurally stable fluid models* which are specified only on the scene layer are described. They are crude models but beneficial to queueing analysis.

(i) *Fluid Markov Video model* (*or FMV model*).

Time is continuous or discrete. The process $T(t)$ is a homogeneous continuous (discrete) time Markov chain with states $\Im = \{1,...., k\}$(k is finite) and transition intensity matrix $Q = \{\lambda_{\theta\theta'}\}_{\theta,\theta'=1,....,k}$ (transition matrix $P = \{p_{\theta\theta'}\}_{\theta,\theta'=1,....,k}$). Initial choice of $T(0)$ is defined by a vector $(p_1,...., p_k)$,

$$p_1 \geq 0, \quad p_2 \geq 0, \quad p_1 + p_2 = 1,$$

where p_θ is the probability to choose scene type S^θ, $\theta \in \Im$. For each $\theta \in \Im$, $B_\theta(t) \equiv b_\theta$ is a constant bit rate.

(ii) *Fluid Semi-Markov Video source model* (*or FSMV model*).

Time is continuous or discrete. The process $\hat{T}(n)$ is a homogeneous discrete time Markov chain with states $\Im = \{1,...., k\}$(k is finite) and with transition matrix $P = \{p_{\theta\theta'}\}_{\theta,\theta'=1,....,k}$. Initial choice of $\hat{T}(0)$ is defined by a vector $(p_1,...., p_k)$,

$$p_\theta \geq 0, \quad p_1 + ... + p_k = 1,$$

where p_θ is the probability to choose scene type S^θ, $\theta \in \Im$. Random variable L_θ has a distribution function G_θ, $\theta \in \Im$. For each $\theta \in \Im$, $B_\theta(t) \equiv b_\theta$ is a constant bit rate.

(iii) *Fluid Alternating Semi-Markov Video source model* (*or FASMV model*).

In this case k = 2(Fig. 5.13) for the previous model and the process

$$\hat{T}(n+1) = \begin{cases} 1, & \text{if } \hat{T}(n) = 2 \\ 2, & \text{if } \hat{T}(n) = 1 \end{cases}$$

for $n = 0,1,.....$ Initial choice of $\hat{T}(0)$ is defined by a vector (p_1, p_2),

$$p_1 \geq 0, \quad p_2 \geq 0, \quad p_1 + p_2 = 1,$$

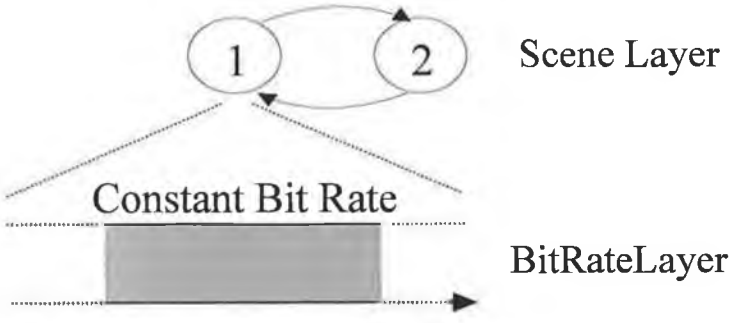


Fig. 5.13 *Stable Fluid Markov/Semi-Markov Video Source Model with k = 2*

where p_θ, is the probability to choose scene type S^θ, $\theta = 1,2$.

(iv) *Fluid Non-Markov Video source model* (*or FNMV model*).

Let $\Im$ be a domain in $\mathbf{R}^d$ and let $F^\Im$ be a probabilistic measure on $\Im$. The type of each new scene, including initial choice of scene type, is chosen independently of the previous scene types and randomly with the measure $F^\Im$. Random variable L_θ has a distribution function G_θ, $\theta \in \Im$. For each $\theta \in \Im$, $B_\theta(t) \equiv b_\theta$ is a constant bit rate.

5.3.5 Discussion of Models

Some remarks are given here about the described models.

- The difference between the continuous time models and the corresponding discrete time models is not essential. This is only a question of convenience.

- First stage of each scene, say of type S^θ, $\theta \in \Im$, in a real video source can also be modelled as a supplementary scene type, or by choosing a special starting set of initial states (and defining the distribution function F_θ^{in} on this set) which produces a larger amount of data on commencement of the scene or be ignored [37].

- All alternating models (CASMV, DASMV, FAV models) are appropriate for video sources which have only two different scene types. They are also convenient to use for simulations.

- All the markovian models (CMV, CAMV, DMV, DAMV, FMV models) are strong simplifications of the model and simpler to use in theoretical studies. The reason for this is that the probability density function of scene duration behaves as an inverse polynomial. For example, in [37] it was approximated by

$$f(x) = \frac{a}{x^\alpha + b},$$

where $a > 0$, $b > 0$, $\alpha > 0$ are some parameters and α has a value close to 2 (for $\alpha = 2$, it is simply the probability density function for the Cauchy distribution).

- The long-term correlated VBR sources are of interest due to their strong impact on the behaviour of an ATM network. The next step under consideration is to study the multiplexing gain of such sources. The relation between the buffer size of a multiplexer and buffer congestion consists of two distinct parts (see e.g. [44, 45]), describing cell and burst congestions, respectively. VBR video sources, due to their slow decaying autocorrelation functions, can contribute to *scene congestion.*

- Two groups of models are presented which give descriptions on different layers. The models from {5.3.4.2} and {5.3.4.3} are defined on both burst and scene layers while the models from {5.3.4.4} are only defined on the scene layer. Thus, for example, the structurally stable fluid semi-Markov video model (or FSMV model) can describe scene congestion while semi-Markov video models (CSMV and DSMV models) can describe burst and scene congestion.

5.4 On the Modelling of a 2-Layer Video Source

A procedure is described here to obtain a 2-layer video source from a 1-layer VBR video source. Observations are based on comparisons of the output traffic from two codecs. Both the 1-layer and 2-layer video codecs are H.261 based(see Figs. 4.2 and 4.3).

The traffic output from the 2-layer codec is divided into two layers which have different priorities. The first layer has a higher priority than the second layer. In the 2-layer codec the following strategy (or *CBR + VBR strategy*) is realised. The first layer attempts to utilise the full bandwidth, B_{max}, put at its disposal. Experiments show this bandwidth is practically fully utilised. Thus it can be assumed in the model for a 2-layer video source that the first layer has a constant bit rate equal to the bandwidth put at its disposal. The experiments also show that the second layer of the 2-layer codec produces output traffic only when the 1-layer codec output traffic exceeds the bandwidth B_{max}, which is allocated for the first layer of the 2-layer codec. Moreover, the second layer output traffic at that moment is exactly equal to the difference between the output traffic from the 1-layer VBR codec and the output traffic from the first layer of the 2-layer codec for the same video sequence (allowing for header traffic on the second layer from the 2-layer codec, which is quite small).

Therefore the 2-layer video source model can be described as follows. Denote the bit rates from the first and second layers at time t by $B^{(1)}(t)$ and $B^{(2)}(t)$, respectively. Assume a 1-layer VBR video source model and let $B(t)$ be an amount of output traffic at time t. Then

$$B^{(1)}(t) = B_{max}, \quad B^{(2)}(t) = \max(0, B(t) - B_{max}).$$

Another strategy (or *VBR + VBR strategy*) of operation of the 2-layer codec can also be realised from this analysis. In this case the output traffic $B(t)$ from the 1-layer VBR codec can be divided between two layers of the 2-layer codec as follows. If $B(t)$ is less than B_{max} then this traffic is transmitted on the first layer of the 2-layer codec. But, if $B(t)$ exceeds B_{max}, then the first layer of the 2-layer codec utilises the bandwidth B_{max} fully at this moment, while the second layer, at the same moment, transmits the extra traffic $B(t) - B_{max}$.

Thus in this case, let $B(t)$ be an amount of output traffic from the 1-layer VBR video source model at time t, then

$$B^{(1)}(t) = \min(B(t), B_{max}), \quad B^{(2)}(t) = \max(0, B(t) - B_{max}).$$

Clearly, for this strategy the total traffic $B^{(1)}(t) + B^{(2)}(t)$ from the 2-layer video source model is equal to the traffic from the 1-layer VBR video source model at time t. Hence, a class of video source models for 2-layer video codecs have been obtained.

6. VBR Video Model Simulation

6.1 Simulation Requirement

To determine the impact video sources will have on ATM B-ISDN resources e.g. bandwidth, buffer dimensions, number of node traversals permitted in a link, etc., network simulations are required. For this purpose, video source models which exhibit the characteristics revealed by analysis in chapter 4 must be obtained. The various hierarchical models presented in chapter 5 are expected to do so and one will be simulated to determine it's effectiveness.

In this chapter, videophone/videoconference source models are simulated by stochastic processes in discrete time and are then analysed and compared to the characteristics obtained by analysis in chapter 4. The short-term autocorrelations will be modelled by First Order Autoregressive Processes but it remains to be seen whether the longer term autocorrelations will be captured by the hierarchical modelling approach inherent in all the models presented in Chapter 5.

Ultimately, this video source model will be applied to narrowband and broadband network simulations to determine the impact on ATM B-ISDN networks to maintain a uniform level of QOS(Quality of Service) required by a videophone/videoconference source.

6.2 Simulation Software Description and Design

The 1-Layer VBR H.261 Video Codec traffic models were written in C code and the software structure of each model is of the general design in Fig. 6.1. The software is designed for two layer models with only two states in the higher layer. For various model simulations, the *Generate Random Duration D* and *Generate Data* modules are the only sections of the software that will require alteration. The Gaussian and Gamma distributed random variables that are implemented in these sections were obtained from [47]. The Gaussian random variable routines in [47] returns a random variable with mean(μ) = 0 and standard deviation(σ) = 1, $RV_{(\mu=0,\sigma=1)}$. If the Gaussian random variable is required to have a mean = μ_w and standard deviation = 1, μ_w is simply added to $RV_{(\mu=0,\sigma=1)}$. A Gamma probability density function is controlled by two parameters, λ and s but the Gamma random variable routine in [47] only requires one parameter, s i.e.

$$P[t < \xi < t + dt] = \frac{t^{s-1}e^{-t}}{\Gamma(s)}dt$$

where $\Gamma(s)$ is the gamma function defined as

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t}\, dt .$$

The function may be altered by letting $\xi = \frac{\hat{\xi}}{\lambda}$ to incorporate the parameter λ into the function :

$$P\left[t < \frac{\hat{\xi}}{\lambda} < t + dt \right] = P\left[\lambda t < \hat{\xi} < \lambda t + \lambda dt \right] = \frac{\lambda (\lambda t)^{s-1} e^{-\lambda t}}{\Gamma(s)} dt .$$

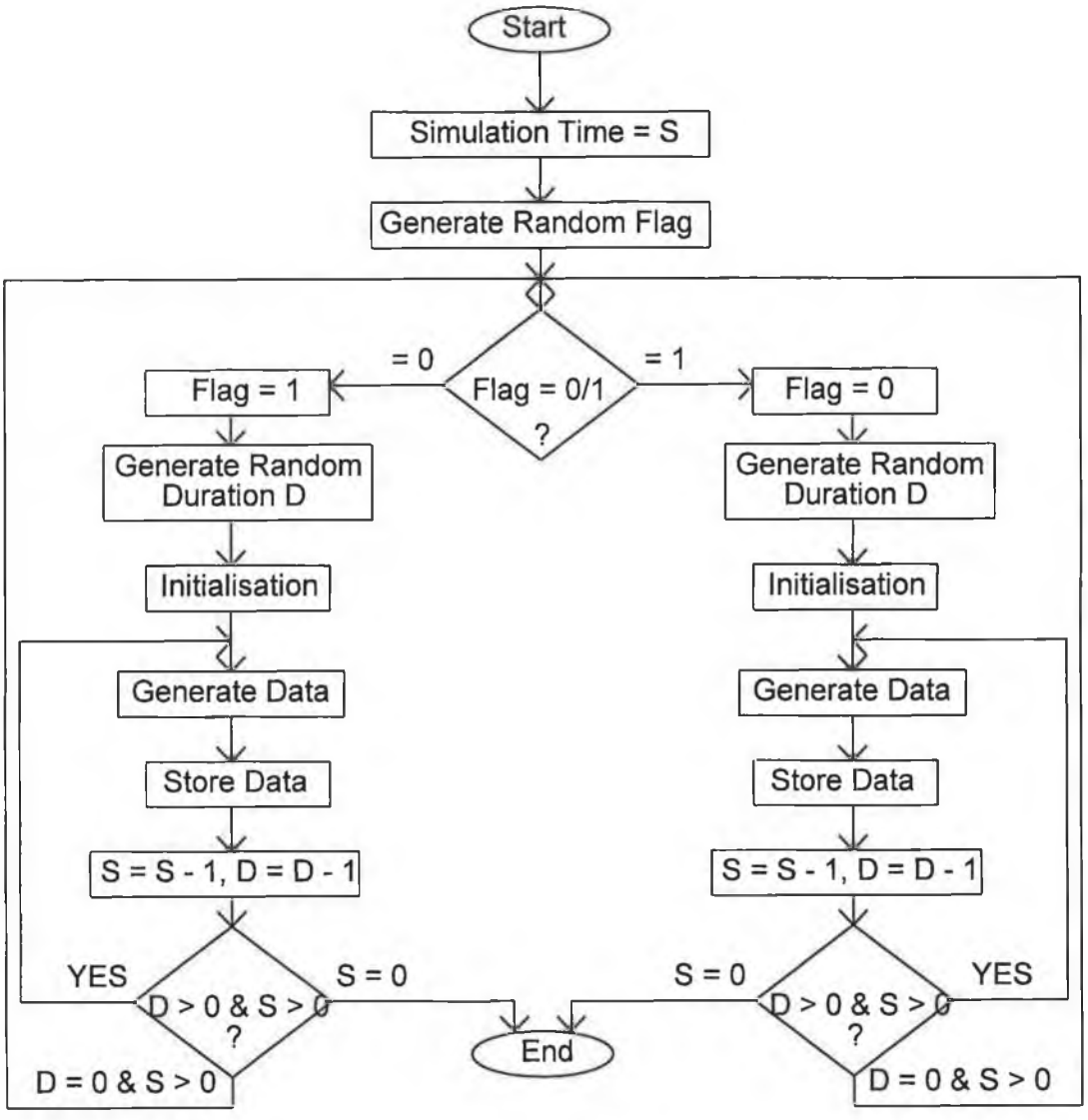


Fig. 6.1 *Two Layer video model software simulation flow-chart.*

By dividing the returned Gamma distributed random variable from the Gamma random variable routine in [47] by λ, the desired Gamma distributed random variable is obtained.

Exponentially distributed random variables are generated by the following function,

$$\xi = \frac{-\log_e U}{\lambda},$$

where $\lambda = 1/\mu_{data}$ and U is a uniformly distributed random variable. If there is a *shift* associated with the distribution imposed during real data analysis in chapter 4, it is also added to generate the correct random quantity.

The *Simulation Time* was also open to variation but remained constant at a value of 15000 frames for each simulation.

6.3 Discrete Autoregressive Video (DAV) Model Simulation

From the analysis of video traffic produced by a 1-Layer VBR H.261 video codec in chapter 4 it was found that short-term sample autocorrelation characteristics were autoregressive or extremely similar(Figs. 4.11, 4.20, 4.29 and 4.36). Therefore, from the selection of hierarchical models presented in Chapter 5 the most obvious model to consider for the task of capturing these characteristics would be a Discrete Autoregressive Video(DAV) model (see {5.3.4.3 (iv)}). In this case $k = 2$ to simulate video traffic for each strategy (Talk-Listen and Motion Level) while the duration in each state, depending on the strategy being considered, is an independent random variable that has a distribution determined by the analysis in Chapter 4.

Within each lower layer state of the DAV model there are the following First Order Autoregressive processes :

$$B_i(n) = a_i B_i(n\text{-}1) + b_i w_i(n), \quad i = 1,2 \tag{6.1}$$

where $w_i(n)$ has a mean, μ_w, and a standard deviation, $\sigma_w = 1$. The coefficient a_i is estimated by two methods : i) the Ratio Method in Appendix A and ii) the Least Squares Method in Appendix B, and the coefficients b_i and μ_w are subsequently estimated by the Method of Moments in Appendix C. Throughout the video model simulation, the data in each state will be independent from the data in the previous state as was roughly determined by analysis in Chapter 4, and during each simulation $B_i(-1)$ will be initialised to μ_i.

6.3.1 The Talk-Listen Strategy

Talk

The talking duration is characterised by an Exponentially distributed random variable, TD, with a mean(μ_{TD}) = 51 frames(see {4.2.3.1.1}). A shift = -3 frames imposed during analysis in Chapter 4 must be re-introduced to re-create talking periods of realistic duration. The mean(μ_{td}) and standard deviation(σ_{td}) of the shifted talking data were found to be 8135.8 bits/frame and 5257.89 bits/frame respectively from the analysis in Chapter 4(see {4.2.3.1.2}). Also to re-produce realistic bits per talking frame, the shift = -1773 bits/frame introduced during analysis must be re-introduced to give a 9908.8 bits/frame mean for talking traffic.

Two sets of parameters were estimated for the First Order Autoregressive Process within the talking state of the DAV model. The parameters are given in Table 6.1 along with the mean(μ_w) of the Gaussian random variable, *w(n)*, that is driving the process which has a standard deviation(σ_w) = 1. The table also contains the mean($\hat{\mu}$) and standard deviation($\hat{\sigma}$) expected from the model. For both sets of data, the coefficient a_i is estimated only by the Least Squares Method in Appendix B. Estimation of a_i by the Ratio Method in Appendix A resulted in $a_i > 1$ which in turn prevented b_i from being estimated by the Method of Moments in Appendix C.

The second set of parameters differ from the first set due to a change in $\hat{\sigma}$. This change was necessary because the original $\hat{\sigma}$ in the first parameter set is associated with a Gamma distributed random variable and this in turn has implications when applied to a Gaussian random variable, *w(n)*, which is implemented in (6.1). By implementing $\hat{\sigma}$ from the first parameter set in Table 6.1, there is a 6% probability($< -1.547\hat{\sigma}$) of producing negative quantities during model simulation. Since the objective is to simulate 1-Layer VBR H.261 Video Codec traffic, negative traffic is prohibited therefore negative traffic must be truncated to zero. This truncation effect in turn will cause an imbalance in the statistics of simulated 1-Layer VBR H.261 Video Codec traffic during talking periods. To overcome the threat of this imbalance, the talking period practical data was re-analysed for variation below the mean only, because the variation of data above the mean is believed to have caused the variation to increase which subsequently will cause difficulties during simulation as already explained. By implementing $\hat{\sigma}$ from the second parameter set in Table 6.1 there is a 4% probability($< -1.75\hat{\sigma}$) of producing negative traffic.

After simulation, the simulated traffic was analysed in the mean and standard deviation which were obtained over a sample of 5000 frames and are given in Table 6.2. Comparisons were also made in the sample autocorrelations which were taken

over the period μ_{TD} and compared to the short-term sample autocorrelation of H.261 Video Codec traffic in Fig. 4.11.

Table 6.1

	Parameter Set #1	Parameter Set #2
a	0.71111	0.71111
b	3696.7214	3269.4979
μ_w	0.635793	0.718872
$\hat{\mu}$	8135.799	8135.799
$\hat{\sigma}$	5257.89	4650.245

In comparing both models, both perform extremely well in the mean(μ_{sim}) and standard deviation(σ_{sim}). The adjustment to $\hat{\sigma}$ was not of great benefit because from Table 6.2, μ_{sim} produced by the first parameter set indicates better performance than the second parameter set. This would also be expected due to the probabilities of negative traffic occurring being very similar in both cases. To perform autocorrelation analysis, sample autocorrelations for both simulations were taken over the duration μ_{TD}. Figs. 6.2 and 6.3 show the short-term sample autocorrelations of simulated traffic by both parameter sets in Table 6.1 respectively.

Table 6.2

	Parameter Set #1	Parameter Set #2
μ_{sim}(bits/frame)	8122.573	8187.52
σ_{sim}(bits/frame)	4832.575	4572.651

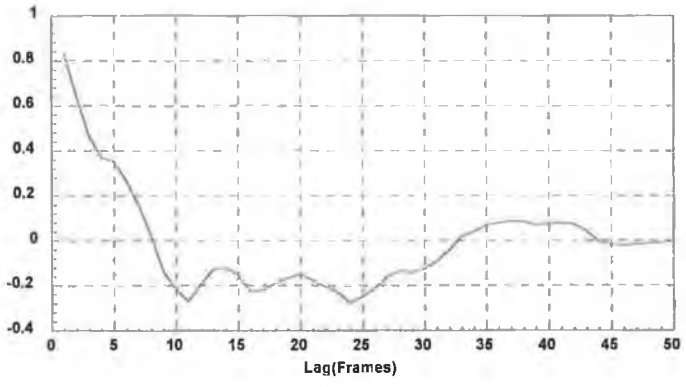


Fig. 6.2 *The sample autocorrelation over a period μ_{TD} for the first parameter set in Table 6.1.*

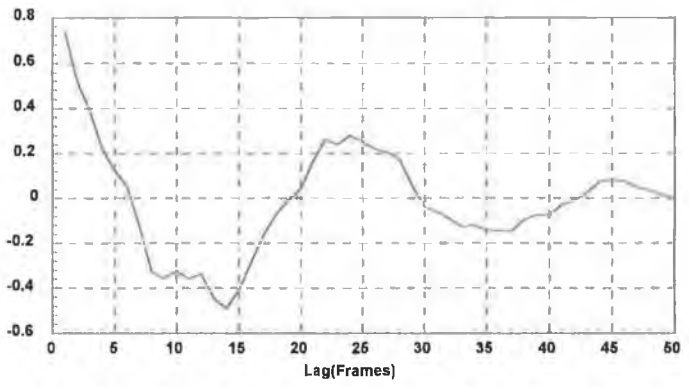


Fig. 6.3 *The sample autocorrelation over a period μ_{TD} for the second parameter set in Table 6.1.*

In comparing these characteristics to Fig. 4.11, the short-term sample autocorrelation of 1-Layer VBR H.261 Video Codec talking data, both short-term sample autocorrelations in Figs. 6.2 and 6.3 exhibit a large degree of similarity to Fig. 4.11, Fig. 6.3 being slightly better. Therefore, both parameter sets give an all-round good performance in the mean, standard deviation and autocorrelation.

Listen

The listening duration is Gamma distributed, determined by the analysis in Chapter 4(see {4.2.3.2.1}). The Gamma distributed listening duration random variable, LD, has a mean(μ_{LD}) and a standard deviation(σ_{LD}) of 47 frames and 39.9 frames respectively which must also be shifted by +2 frames due to the introduction of a shift = -2 frames during analysis in chapter 4. The mean(μ_{ld}) and standard deviation(σ_{ld}) of the shifted listening data were also found to be 5755.354 bits/frame and 5344.8 bits/frame, respectively, by analysis(see {4.2.3.2.2}). A shift = +1322 bits/frame must be re-introduced, due to the shift being imposed during data analysis, to re-produce a realistic listening data mean = 7077.354 bits/frame.

Similar to the approach adopted to model talking data, two sets of parameters for the First Order Autoregressive process in the listening state of the DAV model were estimated to model listening traffic. The parameters are given in Table 6.3 along with the mean(μ_w) of the Gaussian random variable, *w(n)*, that is driving the process which has a standard deviation(σ_w) = 1. The table also contains the mean($\hat{\mu}$) and standard deviation($\hat{\sigma}$) expected from the model. Again, for both sets of data, the coefficient a_i is estimated only by the Least Squares Method in Appendix B.

Estimation of a_i by the Ratio Method in Appendix A resulted in $a_i > 1$ which in turn prevented b_i from being estimated by the Method of Moments in Appendix C.

The second set of parameters differ from the first set due to a change in $\hat{\sigma}$. This change was necessary because the original $\hat{\sigma}$ in the first parameter set is associated with a Gamma distributed random variable and this in turn has implications when applied to a Gaussian distributed random variable, *w(n)*, which is implemented in (6.1). By implementing $\hat{\sigma}$ from the first parameter set in Table 6.3, there is a 14% probability($< -1.077\hat{\sigma}$) of producing negative traffic during model simulation. Since the objective is to simulate 1-Layer VBR H.261 Video Codec listening traffic, negative traffic is prohibited therefore negative traffic must be truncated to zero. This truncation effect in turn will cause an imbalance in the statistics of simulated 1-Layer VBR H.261 Video Codec traffic during listening periods. To overcome the threat of this imbalance, the listening period practical data was re-analysed for variation below the mean only, because the variation of data above the mean is believed to have caused the variation to increase which subsequently will cause difficulties during simulation as already explained. By implementing $\hat{\sigma}$ from the second parameter set in Table 6.3 there is a 7.1% probability($< -1.468\hat{\sigma}$) of producing negative traffic.

Table 6.3

	Parameter Set #1	Parameter Set #2
a	0.81111	0.81111
b	3126.1394	2292.9268
μ_w	0.347754	0.474123
$\hat{\mu}$	5755.354	5755.354
$\hat{\sigma}$	5344.8	3920.246

For both simulations, samples of 5000 frames were taken to determine the mean(μ_{sim}) and standard deviation(σ_{sim}) produced by each parameter set. These results are given in Table 6.4. In comparison with $\hat{\mu}$ and $\hat{\sigma}$ in Table 6.3, the performance of the second parameter set betters that of the first parameter set. This would have been expected due to the higher probability of negative quantities occurring implementing the first parameter set. During simulation negative traffic was truncated to zero subsequently increasing the simulated mean and decreasing the variation from the mean. For further comparison, short-term sample autocorrelations were taken over the period μ_{LD}. Figs. 6.4 and 6.5 are sample autocorrelations of simulated listening data implementing the first and second parameter sets in Table 6.3 respectively. Both short-term autocorrelations have a similar characteristic to that of Fig. 4.20, the short-term sample autocorrelation of 1-Layer VBR H.261 Video Codec listening traffic. The autocorrelation of Fig. 6.5 shows a stronger similarity to

the decaying characteristic of Fig. 4.20 than does Fig. 6.4 which displays a decaying oscillation characteristic but both decay to zero within a short period. Therefore, due to the second parameter set in Table 6.3 achieving better performance in the mean, standard deviation and autocorrelation, the listening traffic from a 1-Layer VBR H.261 Video Codec should be modelled by a First Order Autoregressive process implementing the second parameter set in Table 6.3.

Table 6.4

	Parameter Set #1	Parameter Set #2
μ_{sim}(bits/frame)	6094.088	5794.287
σ_{sim}(bits/frame)	4578.289	3665.498

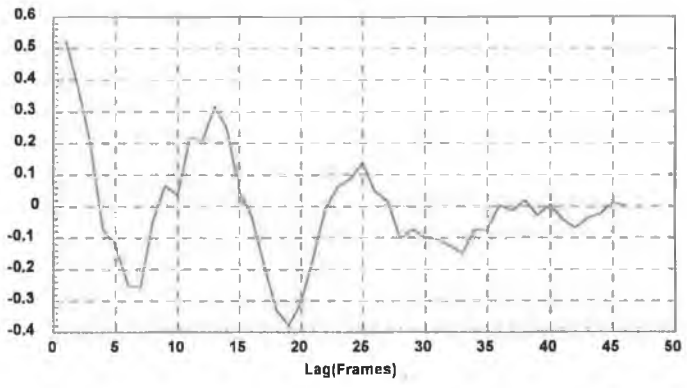


Fig. 6.4 *The sample autocorrelation over the period μ_{LD} for the first parameter set in Table 6.3.*

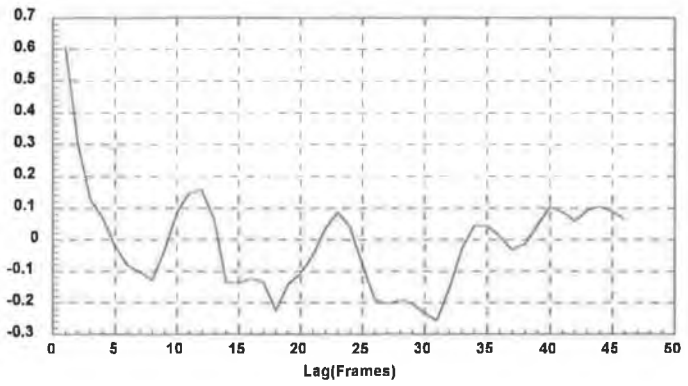


Fig. 6.5 *The sample autocorrelation over the period μ_{LD} for the second parameter set in Table 6.3.*

The Long-term Correlations

In the analysis of videoconference/videophone source traffic, there are long-term correlation characteristics inherent in the traffic(Fig. 4.38). Chapter 5 introduces several hierarchical models to capture these long-term correlations. One such model, the DAV model, has been applied in the Talk-Listen strategy. From this model simulation, the following analysis was made. The simulation data is the result of previous talk and listen traffic model simulations implementing the first and second parameter sets in Tables 6.1 and 6.3. The talk and listen models alternate between each other under the control of the higher talk and listen duration layer. The talk and .

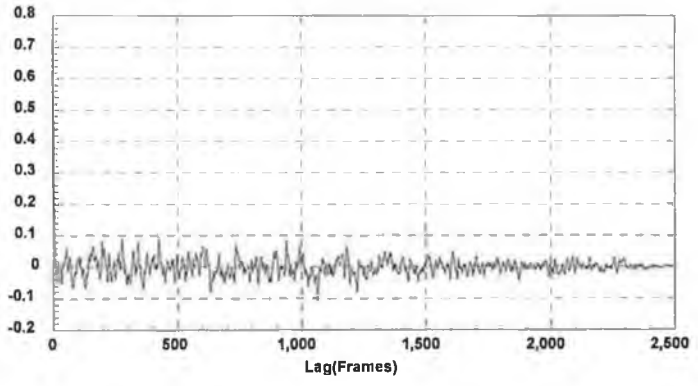


Fig. 6.6 *The long-term sample autocorrelation of Talk-Listen data modelled by the DAV model with the first parameter sets in Tables 6.1 and 6.3.*

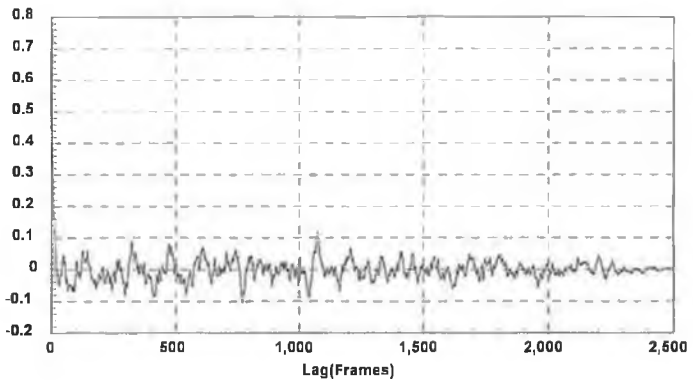


Fig. 6.7 *The long-term sample autocorrelation of Talk-Listen data modelled by the DAV model with the second parameter sets in Tables 6.1 and 6.3.*

listen durations are characterised by independent random variables which are Exponentially and Gamma distributed, respectively(see Chapter 4)

From the simulation of both models, Figs. 6.6 and 6.7 are their long-term sample autocorrelations. In comparison to Fig. 4.38, Figs. 6.6 and 6.7 fail to exhibit any sustained long-term autocorrelations similar to those present in Fig. 4.38. The reason the long-term autocorrelations are suppressed in both figures is related to the ratio $\frac{\hat{\sigma}}{\hat{\mu}}$, where in both talking and listening traffic cases $\frac{\hat{\sigma}}{\hat{\mu}} \approx 1$. In Fig. 6.6 the ratio $\frac{\hat{\sigma}}{\hat{\mu}}$(for both talking and listening traffic) is closer to 1 which is clear to see from the first parameter sets in Tables 6.1 and 6.3. In Fig. 6.7 on the other hand, long-term autocorrelations are slightly more noticable due to the adjustment of σ_{td} and σ_{ld} hence causing $\frac{\hat{\sigma}}{\hat{\mu}} < 1$ in both talking and listening traffic cases. Again this is clear to see from the second parameter sets in Tables 6.1 and 6.3. Several autocorrelations were made over various periods within the simulated traffic and all failed to reveal similar features to those in Fig. 4.38.

To analyse the simulated 1-Layer VBR H.261 Video Codec traffic further, several samples of the mean and standard deviation of the simulated data were determined over a sample length of 2500 frames and are given in Table 6.5. The unshifted mean of 1-Layer VBR H.261 Video Codec traffic is 8.88 kbits/frame(see Table 4.2), but throughout these simulations both the simulated talk and listen traffic had been shifted 1773 bits/frame and 1322 bits/frame respectively. Therefore, to estimate the mean of shifted 1-Layer VBR H.261 Video Codec traffic, the duration in

each state must be considered. The percentage of time the H.261 Video Codec traffic is in a talking state is determined by

$$\frac{\mu_{TD}}{\mu_{TD}+\mu_{LD}} \text{x} 100.$$

In this case, 1-Layer VBR H.261 Video Codec traffic is in a talking state 52% of the time while 48% of the time the traffic is in a listening state. Hence, the shifted 1-Layer VBR H.261 Video Codec traffic mean, μ, can be estimated by

$$\mu = 0.52 \text{ x } \mu_{td} + 0.48 \text{ x } \mu_{ld}$$

which gives μ = 6993.18 bits/frame. In comparing the μ_{sim} values in Table 6.5 to μ, the nearest performance to μ is achieved by the traffic model implementing the second parameter sets. The μ_{sim} values achieved by the first parameter sets overshoot μ as a result of negative traffic truncation to zero. The model implementing the second parameter sets has a lower probability of negative data truncation to zero due to $\hat{\sigma}$ adjustment explained previously during talk and listen data modelling. The standard deviation(σ) of real 1-Layer VBR H.261 Video Codec traffic is equal to 5476.8 bits/frame determined by analysis. The standard deviations in Table 6.5 under-estimate σ due to negative traffic truncation to zero in simulated models implementing the first parameter sets and reduction of $\hat{\sigma}$ to avoid negative data generation in simulated models implementing the second parameter sets. Therefore, overall, the model performs well in the mean and standard deviation but fails to exhibit sustained long-term autocorrelations.

Table 6.5

	Parameter Sets #1 from Tables 6.1 & 6.3	Parameter Sets #2 from Tables 6.1 & 6.3
$\mu_{sim}1$(bits/frame)	7284.121	7009.94
$\mu_{sim}2$(bits/frame)	7368.248	6782.833
$\mu_{sim}3$(bits/frame)	7376.187	6667.491
$\sigma_{sim}1$(bits/frame)	4903.075	4365.124
$\sigma_{sim}2$(bits/frame)	4843.072	4254.32
$\sigma_{sim}3$(bits/frame)	4790.859	4221.319

6.3.2 The Motion Level Strategy

High Level Motion

The duration in high motion from the analysis in Chapter 4 is characterised by an Exponentially distributed random variable, HMD, with a mean(μ_{HMD}) of 56 frames(see {4.2.4.1.1}) which must be shifted an additional +15 frames to re-create realistic high motion periods. The corresponding high motion data is characterised by a Gaussian random variable with a mean(μ_{hmd}) and standard deviation(σ_{hmd}) of 10141.69 bits/frame and 4962.69 bits/frame respectively(see {4.2.4.1.2}). This random variable must be shifted +2549 bits/frame due to the introduction of a shift = -2549 bits/frame during high motion data analysis in chapter 4 to re-create a realistic high level motion mean = 12690.69 bits/frame.

Two sets of parameters for the First Order Autoregressive process in the high motion state of the DAV model were estimated to model high motion data. The parameters are given in Table 6.6 along with the mean(μ_w) of the Gaussian random variable, *w(n)*, that is driving the process which has a standard deviation(σ_w) = 1. The table also contains the mean($\hat{\mu}$) and standard deviation($\hat{\sigma}$) expected from the model. The coefficient a_i for the first parameter set is estimated by the Ratio Method in Appendix A while in the second parameter set it is estimated by the Least Squares Method in Appendix B. b_i and μ_w are subsequently estimated by the Method of Moments in Appendix C for both parameter sets. Unlike the talk and listen data distributions, the requirement to adjust $\hat{\sigma}$ in the first parameter set to reduce the probability of negative traffic occurring is unnecessary in this case. The high motion data distribution determined by analysis in chapter 4 only has 2.1% probability($< -2.044\hat{\sigma}$) of producing negative traffic during model simulation. Therefore such a small probability will not have a large impact on the simulated model statistics.

Table 6.6

	Parameter Set #1	Parameter Set #2
a	0.92951	0.85111
b	1830.2395	2605.3494
μ_w	0.390604	0.579575
$\hat{\mu}$	10141.69	10141.69
$\hat{\sigma}$	4962.69	4962.69

Samples of the mean(μ_{sim}) and standard deviation(σ_{sim}) of the simulated models taken over a sample length of 5000 frames are given in Table 6.7. In comparison with μ_{hmd} and σ_{hmd} both models perform quite satisfactorily. μ_{sim} for both parameter sets will always slightly over-estimate μ_{hmd} the majority of time due to the truncation of negative traffic to zero which is quite evident in Table 6.7.

Truncation will also reduce traffic variation from the mean in comparison with σ_{hmd} which again is evident in Table 6.7.

To analyse the model data on a correlation basis, the simulated data was autocorrelated in the short-term. The short-term sample autocorrelations for the first and second parameter sets in Table 6.6 are shown in Figs. 6.8 and 6.9 respectively. Both short-term sample autocorrelations are significantly similar and more importantly resemble the 1-Layer VBR H.261 Video Codec traffic short-term sample autocorrelation in Fig. 4.29. Therefore, with both parameter sets in Table 6.6 giving good performance in the mean, standard deviation and autocorrelation, either parameter set can be applied to the 1-Layer VBR H.261 Video Codec high motion traffic model.

Table 6.7

	Parameter Set #1	Parameter Set #2
$\mu_{sim}1$(bits/frame)	10369.34	9956.497
$\mu_{sim}2$(bits/frame)	10290.15	10239.17
$\sigma_{sim}1$(bits/frame)	4803.725	4691.164
$\sigma_{sim}2$(bits/frame)	4563.224	4666.125

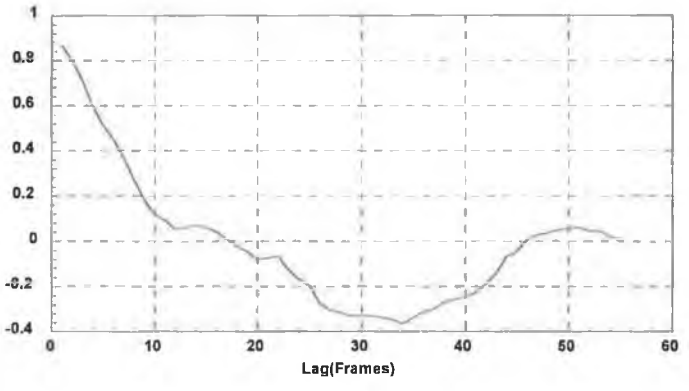


Fig. 6.8 *The sample autocorrelation over the period μ_{HMD} for the first parameter set in Table 6.6.*

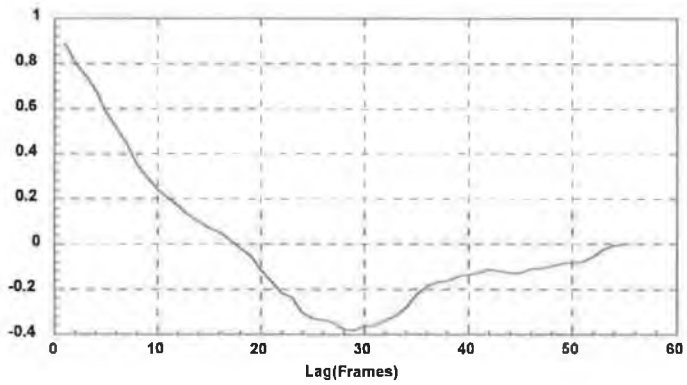


Fig. 6.9 *The sample autocorrelation over the period μ_{HMD} for the second parameter set in Table 6.6.*

Low Level Motion

The duration in a low level motion state is characterised by an Exponentially distributed random variable with a mean duration of 84 frames determined by the analysis in Chapter 4(see {4.2.4.2.1}). The Exponential random variable must also be shifted +15 frames due to a shift = -15 frames introduced during analysis in Chapter 4 to re-create a realistic low level motion duration. The low level motion data distribution is also characterised by a Gamma distributed random variable with a mean(μ_{lmd}) and standard deviation(σ_{lmd}) of 5042.89 bits/frame and 4584.9 bits/frame, respectively, as was determined by analysis in Chapter 4(see {4.2.4.2.2}). A shift = +1322 bits/frame must also be re-introduced to re-create a realistic low level motion data mean = 6364.89 bits/frame.

Four sets of parameters for the First Order Autoregressive process in the low motion state of the DAV model were estimated to model low motion traffic. The parameters are given in Table 6.8 along with the mean(μ_w) of the Gaussian random variable, *w(n)*, that is driving the process which has a standard deviation(σ_w) = 1. The table also contains the mean($\hat{\mu}$) and standard deviation($\hat{\sigma}$) expected from the model implementing each parameter set. The coefficient a_i for the first and third parameter sets is estimated by the Ratio Method in Appendix A while in the second and fourth parameter sets it is estimated by the Least Squares Method in Appendix B. b_i and μ_w are subsequently estimated by the Method of Moments in Appendix C for all parameter sets. For the first and second parameter sets $\hat{\sigma} = \sigma_{lmd}$. In the third and fourth parameter sets $\hat{\sigma}$ has been adjusted to reduce the probability of negative traffic occurring. This high probability of producing negative traffic is due to the application of statistical parameters $\hat{\mu}$ and $\hat{\sigma}$ in parameter sets #1 and #2, which are parameters

describing a Gamma distributed random variable, to a Gaussian distributed random variable. The probability of producing negative traffic with $\hat{\sigma} = \sigma_{lmd}$ is 13.6%($< -1.1\hat{\sigma}$) but by adjusting $\hat{\sigma}$, similar to the method implemented in modelling talk and listen traffic, to $\hat{\sigma}$ = 3001.952 bits/frame a reduced probability of negative traffic occurring of 4.7%($< -1.68\hat{\sigma}$) is obtained.

Table 6.8

	Parameter Set #1	Parameter Set #2	Parameter Set #3	Parameter Set #4
a	0.75589	0.86411	0.75589	0.86411
b	3001.7262	2307.5935	1965.3728	1510.8912
μ_w	0.410103	0.296967	0.626354	0.453559
$\hat{\mu}$	5042.89	5042.89	5042.89	5042.89
$\hat{\sigma}$	4584.9	4584.9	3001.952	3001.952

Table 6.9

	Parameter Set #1	Parameter Set #2	Parameter Set #3	Parameter Set #4
μ_{sim}(bits/frame)	5390.562	5823.985	5002.683	5141.919
σ_{sim}(bits/frame)	3953.991	3955.963	2974.706	2823.223

Samples of the mean(μ_{sim}) and standard deviation(σ_{sim}) of the simulated traffic taken over a sample length of 5000 frames are given in Table 6.9. In comparison with the expected $\hat{\mu}$ and $\hat{\sigma}$ from each model in Table 6.8, models implementing parameter sets #3 and #4 perform quite satisfactorily. The improvement over parameter sets #1 and #2 by parameter sets #3 and #4 has come about due to the adjustment of $\hat{\sigma}$. In the simulated models implementing parameter sets #1 and #2, any negative traffic produced has been truncated to zero causing μ_{sim} of the simulated data to rise which is clearly seen in the statistics. This truncation effect also causes σ_{sim} of the simulated data to decline. Negative data truncations to zero also occur in simulated data implementing parameter sets #3 and #4 but this occurrence is rare due to reasons explained previously and hence has a lesser impact on the simulated data statistics.

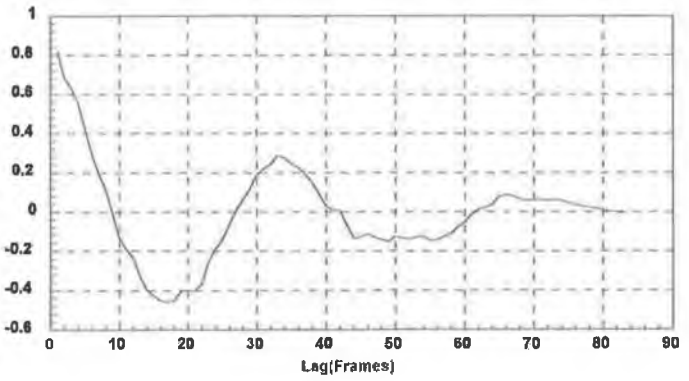


Fig. 6.10 *The sample autocorrelation over the period μ_{LMD} for the first parameter set in Table 6.8.*

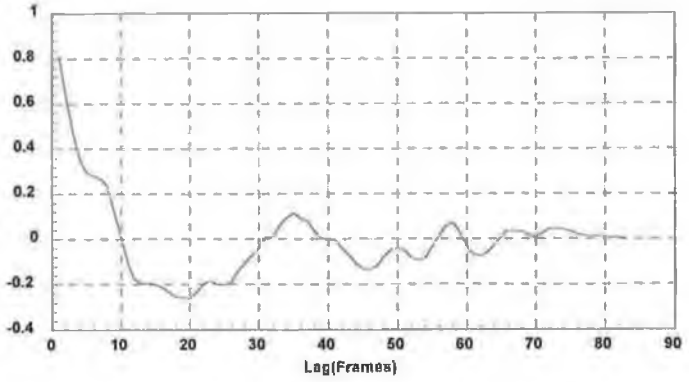


Fig. 6.11 *The sample autocorrelation over the period μ_{LMD} for the second parameter set in Table 6.8.*

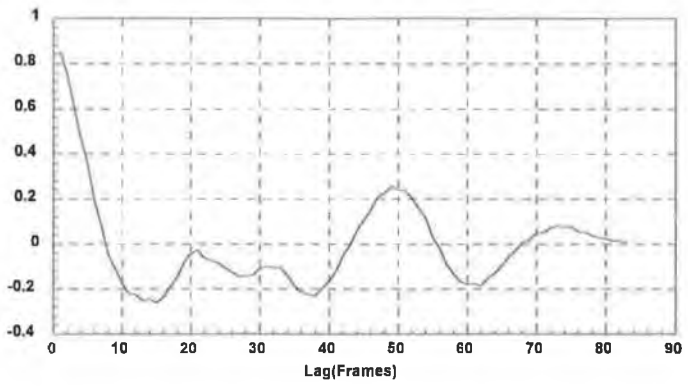


Fig. 6.12 *The sample autocorrelation over the period μ_{LMD} for the third parameter set in Table 6.8.*

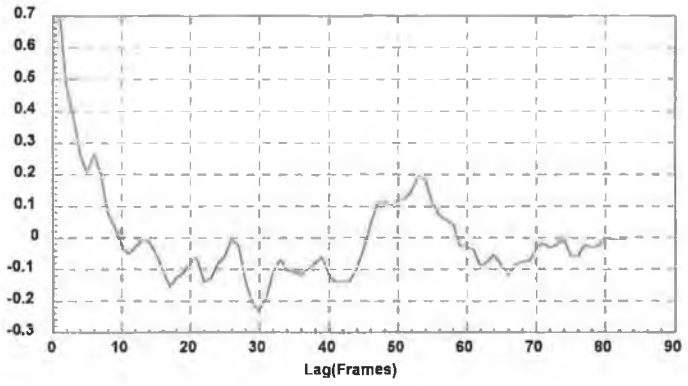


Fig. 6.13 *The sample autocorrelation over the period μ_{LMD} for the fourth parameter set in Table 6.8.*

To analyse the model data on a correlation basis, the simulated data was autocorrelated in the short-term. The short-term sample autocorrelations for the first, second, third and fourth parameter sets in Table 6.8 are shown in Figs. 6.10, 6.11, 6.12 and 6.13 respectively. In comparison to Fig. 4.36, the short-term sample autocorrelation of real low level motion traffic, Figs. 6.10 to 6.13 exhibit a close similarity. All simulated traffic short-term sample autocorrelations exhibit exponentially decaying oscillating characteristics, Figs. 6.10 and 6.12 being particularly good. Unfortunately, Fig. 6.10 is the short-term sample autocorrelation of traffic simulated by parameter set #1 in Table 6.8 which over-estimates $\hat{\mu}$ and

under-estimates $\hat{\sigma}$. Therefore low level motion traffic can be modelled due to parameter sets #3 and #4 in Table 6.8 performing well in the mean and standard deviation and exhibiting satisfactory short-term sample autocorrelations in Figs. 612 and 6.13 respectively.

The Long-term Correlations

In the analysis of videoconference/videophone source traffic, there are long-term correlation characteristics inherent in the data(Fig. 4.38). The hierarchical DAV model introduced in chapter 5 was simulated in the attempt to capture these long-term correlations by the Motion Level strategy. From this model simulation, the following analysis was made. The simulated traffic is the result of previous high motion and low motion traffic models in Tables 6.6 and 6.8 alternating between each other under the control of the higher high motion and low motion duration layers. The models combine to create four various models, which are as follows :

- Model #1 = Parameter Set #1 in Table 6.6 + Parameter Set #1 in Table 6.8;
- Model # 2 = Parameter Set #1 in Table 6.6 + Parameter Set #3 in Table 6.8;
- Model #3 = Parameter Set #2 in Table 6.6 + Parameter Set #2 in Table 6.8;
- Model #4 = Parameter Set #2 in Table 6.6 + Parameter Set #4 in Table 6.8;

The durations in high motion and low motion states are characterised by independent random variables which are Exponentially and Gamma distributed, respectively(see Chapter 4).

The unshifted mean of 1-Layer VBR H.261 Video Codec traffic from Table 4.2 is 8.88kbits/frame and the standard deviation(σ) was also determined to be 5476.8 bits/frame. To estimate the shifted mean(μ) of 1-Layer VBR H.261 Video Codec traffic, which is necessary since the simulated data is shifted, the percentage duration in each state must be considered. This is estimated by the following approach :

$$\frac{(\mu_{HMD} + shift_{HMD})}{(\mu_{HMD} + shift_{HMD}) + (\mu_{LMD} + shift_{LMD})} \text{x100}$$

where μ_{HMD}(mean high motion duration) = 56, $shift_{HMD}$(re-introduction of high motion duration shift imposed during analysis) = 15, μ_{LMD}(mean low motion

duration) = 84 and $shift_{LMD}$(re-introduction of low motion duration shift imposed during analysis) = 15. By applying this approach, 42% of 1-Layer VBR H.261 Video Codec traffic is estimated to be in high motion leaving the other 58% in low motion. Therefore, μ can be estimated as follows :

$$\mu = 0.42 \times \mu_{hmd} + 0.58 \times \mu_{lmd} = 7184.386 \text{ bits/frame.}$$

Several samples of the mean(μ_{sim}) and standard deviation(σ_{sim}) produced by Models #1 to #4 are given in Table 6.10. These samples were obtained over sample lengths of 2500 frames. In comparison to μ, μ_{sim} of Models #1 and #2 perform quite well even though σ_{lmd} has not been adjusted in either case to minimise negative traffic truncation to zero. The only model of the two, Model #1, indicates it's potential to produce a mean > μ while Model #2 exhibits quite a stable mean. On the other hand, the mean values of Models #3 and #4 would be expected to approximate μ due to an adjustment to σ_{lmd} to minimise negative traffic truncation to zero. Model #3 exhibits the ability to achieve this objective where Model #4 fails but indicates it's potential to achieve this objective. A larger sample will undoubtedly reveal more but based on the separate high and low motion simulated traffic analysis, previous to this analysis, Models #3 and #4 will achieve better performance in the mean than Models #1 and #2. With regard to standard deviation, the simulated traffic reveals what was expected. Models #3 and #4 produce traffic with lower deviations from the mean than do Models #1 and #2. This is a result of the adjustment to σ_{lmd} to prevent negative data truncations to zero during low motion periods.

Table 6.10

	Model #1	Model #2	Model #3	Model #4
$\mu_{sim}1$(bits/frame)	8600.549	7220.059	6756.51	7768.07
$\mu_{sim}2$(bits/frame)	7613.89	7264.768	7066.729	6972.801
$\mu_{sim}3$(bits/frame)	7689.122	7513.323	7288.78	7824.249
$\sigma_{sim}1$(bits/frame)	5380.159	4720.053	4445.452	4370.082
$\sigma_{sim}2$(bits/frame)	5062.722	4750.725	4512.707	4257.62
$\sigma_{sim}3$(bits/frame)	4872.967	4919.316	4711.754	4655.075

To further analyse the simulated 1-Layer VBR H.261 Video Codec traffic from the four models, samples of data from each model were autocorrelated to determine whether the hierarchical modelling approach had achieved it's objective in capturing the long-term autocorrelations inherent in 1-Layer VBR H.261 Video Codec traffic(see Fig. 4.38). Long-term sample autocorrelations of data from Models #1 - #4 are shown in Figs. 6.14 - 6.17 respectively. In comparison to Fig. 4.38, Figs.

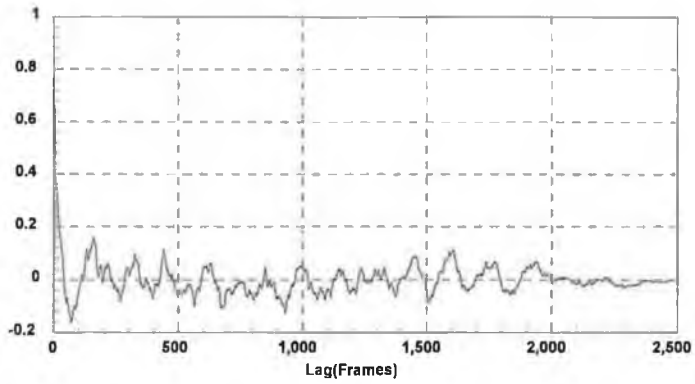


Fig. 6.14 *The long-term sample autocorrelation of High and Low Level Motion data simulated by the DAV model with Model #1 parameters.*

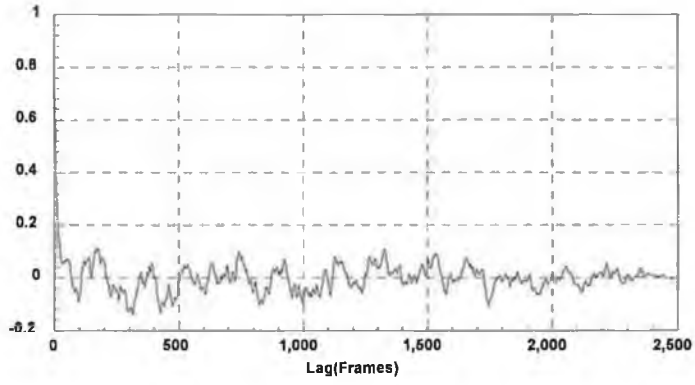


Fig. 6.15 *The long-term sample autocorrelation of High and Low Level Motion data simulated by the DAV model with Model #2 parameters.*

6.14 and 6.15 exhibit long-term autocorrelations but the magnitude of the correlations are suppressed. The cause for this suppression relates back to the truncation of negative traffic to zero, particularly in low motion traffic, thus reducing the magnitude of traffic variation from the mean. In Figs. 6.16 and 6.17, long-term autocorrelations are more prominent and compare highly with those in Fig. 4.38. These long-term correlations are also sustained to within a lag = 2000 frames. Models #3 and #4 succeeded in capturing long-term correlations due to the adjustment of σ_{lmd} to minimise negative traffic truncations to zero thus traffic deviations from the mean were not stunted. These autocorrelation results in Figs. 616

and 6.17 also illustrate the ability of the hierarchical DAV model to capture long-term autocorrelations. Therefore, since both Models #3 and #4 perform well in the mean and standard deviation and succeed in modelling the long-term autocorrelations inherent in 1-Layer VBR H.261 Video Codec traffic, both models may be implemented to artificially generate such traffic.

Due to the presence of these longer term autocorrelations within the simulated 1-Layer VBR H.261 Video Codec traffic, the traffic will require a longer period to reach a stationary value, i.e. the mean. For this reason the μ_{sim} values in Table 6.10 are greater than μ = 7184.385 bits/frame because the analysis should be made over a sample length >> 2500.

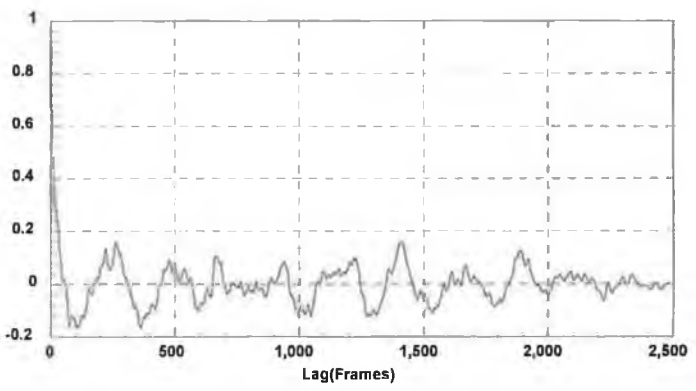


Fig. 6.16 *The long-term sample autocorrelation of High and Low Level Motion data simulated by the DAV model with Model #3 parameters.*

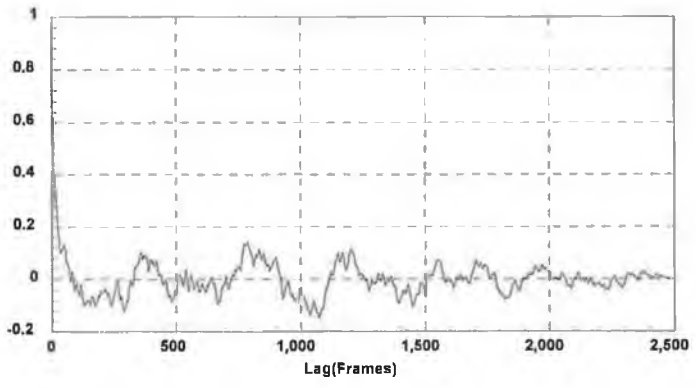


Fig. 6.17 *The long-term sample autocorrelation of High and Low Level Motion data simulated by the DAV model with Model #4 parameters.*

6.4 Discrete State Non-correlated Video Model Simulation

The purpose of this modelling approach is to further analyse the long-term correlations only and to further determine if long-term autocorrelations are captured by both Talk-Listen and Motion Level strategies implementing the layered video traffic modelling approach. The model being simulated is basically that of the DAV model for both Talk-Listen and Motion Level strategies but the First Order Autoregressive processes have been removed from each state. The data generating process within each state is replaced by an independent random variable with a distribution corresponding to the data distributions determined by analysis in Chapter 4. The model therefore could be more accurately described as a continuous time two state model with random variable generation in discrete time i.e. traffic is generated every 40msec. in a PAL standard video sequence.

6.4.1 The Talk-Listen Strategy

For the Talk-Listen strategy model simulation, the talk duration is characterised by an Exponentially distributed random variable with a mean of 51 frames, the listen duration is characterised by a Gamma distributed random variable with a mean and standard deviation of 47 frames and 39.91 frames respectively. To approximate the talk and listen duration distributions by known distributions, they were shifted -3 and -2 frames respectively. During simulation these shifts will be re-introduced to re-create realistic talking and listening durations. The talking data independent random variable is characterised by a Gamma distributed random variable with a mean(μ_{talk}) and standard deviation(σ_{talk}) of 8135.799 bits/frame and 5257.9 bits/frame respectively and the listening data independent random variable is also characterised by a Gamma distributed random variable with a mean(μ_{listen}) and standard deviation(σ_{listen}) of 5755.354 bits/frame and 5344.8 bits/frame respectively. Due to the shifting of both data distributions to allow approximations by known distributions in the analysis in chapter 4, the shifts must be re-introduced during simulation. The talking and listening data shifts are 1773 bits/frame and 1322 bits/frame respectively.

Table 6.11

	Simulated Model Characteristics
$\mu_{sim}1$(bits/frame)	8756.95
$\mu_{sim}2$(bits/frame)	8433.39
$\sigma_{sim}1$(bits/frame)	5563.88
$\sigma_{sim}2$(bits/frame)	5429.55

Since the simulated data is produced by independent random variables, short-term correlations do not serve any purpose. Therefore, long-term autocorrelation characteristics are only analysed. A couple of samples of the mean(μ_{sim}) and standard deviation(σ_{sim}) for a non-correlated video model simulation taken over a sample length = 2500 frames are given in Table 6.11. In comparing these statistics to those of 1-Layer VBR H.261 Video Codec traffic in Table 4.2, with a mean = 8.88 kbits/frame and standard deviation = 5476.8 bits/frame, the model performance is quite good in all respects. Of greater interest though is the long-term autocorrelation analysis of the simulated traffic to determine whether the layered approach to video traffic modelling will capture the longer-term correlations inherent in video traffic. A long-term sample autocorrelation of the simulated video data is illustrated in Fig. 6.18. In comparison to the long-term sample autocorrelation characteristic of 1-Layer VBR H.261 Video Codec traffic in Fig. 4.38, Fig. 6.18 illustrates a highly uncorrelated sequence and indicates a failure in the Talk-Listen strategy layered modelling approach to capture the longer-term correlations inherent in video traffic.

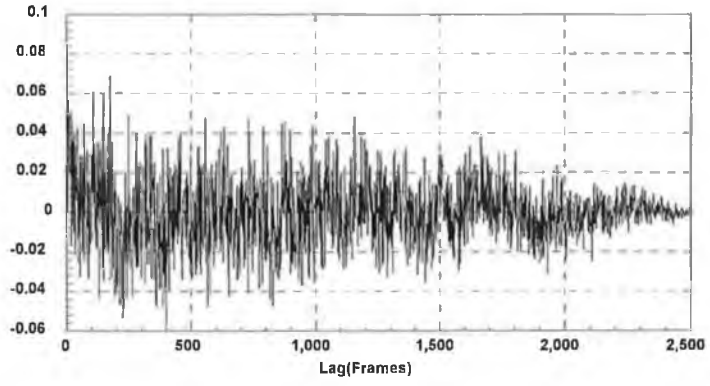


Fig. 6.18 *The long-term sample autocorrelation of simulated 1-Layer VBR H.261 Video Codec traffic by a Non-correlated Video model implementing the Talk-Listen strategy.*

6.4.2 The Motion Level Strategy

To investigate the long-term correlations of simulated Motion Level strategy traffic produced by a Non-correlated Video model, the high level motion duration is characterised by an Exponentially distributed random variable with a mean of 56 frames. The low level motion duration is also characterised by an Exponentially distributed random variable with a mean of 84 frames. The shifts introduced during the duration analysis in chapter 4 must be re-introduced to re-create realistic high and low motion durations during model simulation. Therefore, a shift of 15 frames must

be re-introduced for both high and low level motion duration distributions. The corresponding data within each motion level state of the model is also produced by independent random variables. The high level motion data is characterised by a Gaussian distributed random variable with a mean(μ_{high}) and standard deviation(σ_{high}) of 10141.69 bits/frame and 4962.7 bits/frame respectively and the low level motion data is characterised by a Gamma distributed random variable with a mean(μ_{low}) and standard deviation(σ_{low}) of 5042.889 bits/frame and 4584.9 bits/frame respectively. For both data random variables, shifts of 2549 bits/frame and 1322 bits/frame, respectively, must also be re-introduced after analysis in chapter 4.

Again, short-term analysis would be futile because the simulated video data is produced by independent random variables. Therefore, long-term correlation analysis is of more interest to investigate the effectiveness of the layered modelling approach to model 1-Layer VBR H.261 Video Codec traffic and to capture the long-term correlations present in video traffic. Firstly, the simulated data is analysed to determine the mean(μ_{sim}) and standard deviation(σ_{sim}) of the model. Two samples of μ_{sim} and σ_{sim} from the simulated model obtained over a sample length = 2500 frames are given in Table 6.12. In comparison to the sampled mean of 1-Layer VBR H.261

Table 6.12

	Model Statistical Characteristics
mean1(bits/frame)	8822.7
mean2(bits/frame)	8868.2
standard dev1(bits/frame)	4874.4
standard dev2(bits/frame)	4761.8

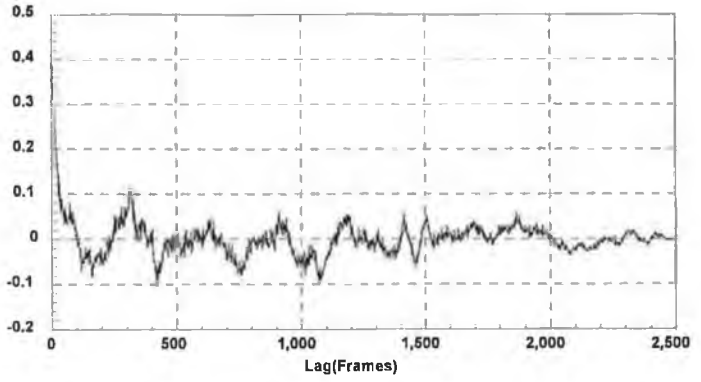


Fig. 6.19 *The long-term sample autocorrelation of simulated 1-Layer VBR H.261 Video Codec traffic by a Non-correlated Video model implementing the Motion Level strategy.*

Video Codec traffic in Table 4.2 and the standard deviation of the traffic = 5476.8 bits/frame, the model performs exceptionally well. A long-term sample autocorrelationof the simulated data is illustrated in Fig. 6.19. In comparison to Fig. 4.38, the long-term correlations are not as prominent but within a *lag* = 1500 the model exhibits the presence of a certain level of long-term correlations. Therefore, since the data generating random variables are independent, these small levels of long-term autocorrelations can only be attributed to the hierarchical layer approach to 1-Layer VBR H.261 Video Codec traffic modelling.

6.5 Impact of Duration Distributions in the Higher Layer

In the analysis of long-term sample autocorrelations in Figs. 6.6, 6.7, 6.14, 6.15, 6.16 and 6.17, the failure or success of the hierarchical model in each case to capture the long-term correlations was put down to the ratio $\frac{\hat{\sigma}}{\hat{\mu}}$. If $\frac{\hat{\sigma}}{\hat{\mu}} \approx 1$ in each traffic producing state of the hierarchical model, the model was very unlikely to exhibit long-term autocorrelations. This was deduced from comparisons between Talk-Listen and Motion Level strategy long-term sample autocorrelations where $\frac{\hat{\sigma}}{\hat{\mu}} \approx 1$ in each traffic producing state and $\frac{\hat{\sigma}}{\hat{\mu}} < 1$ in each traffic producing state respectively. To investigate these deductions further it would be necessary to eliminate the varying durations in each higher state in case they were affecting the outcome.

In the following analysis the Motion Level strategy is only considered. To eliminate the varying duration within each traffic producing state, the durations were fixed to their respective mean durations, i.e. μ_{HMD} and μ_{LMD}, with the shift introduced during analysis in chapter 4 was also added. The model resembles Model #4 of the motion level strategy simulations in {6.3} but durations within each higher layer state remain fixed. The resulting long-term sample autocorrelation of traffic from this simulated model is shown in Fig. 6.20 and immediately noticeable are the slowly decaying long-term correlations produced by the model. (Note - This long-term sample autocorrelation was taken over a sample length of 5000 frames, 2500 frames are shown).

To further investigate the impact variations in duration excert on the long-term autocorrelations, the duration within each state in the previous simulation was allowed to vary about their fixed durations. The varying durations were Gaussian distributed with an initial standard deviation of 5 frames followed by a standard deviation of 10 frames. The resulting long-term sample autocorrelations from both simulations are shown in Figs. 6.21 and 6.22 respectively. (Note - These long-term

sample autocorrelations were taken over sample lengths of 5000 frames, 2500 frames are shown). From both these long-term sample autocorrelations the effect of varying durations within a state are clearly apparent. By increasing the variation around the mean duration causes an increase in the long-term correlation rate of decay or alternatively, smoothes the long-term correlations within the simulated traffic. Therefore, from this analysis the long-term correlations in Figs. 6.14 - 6.17 are suppressed due to the durations within each higher layer state of the hierarchical model. These durations were modelled by Exponentially distributed random variables(4.1) which have standard deviations = mean of the respective exponentially

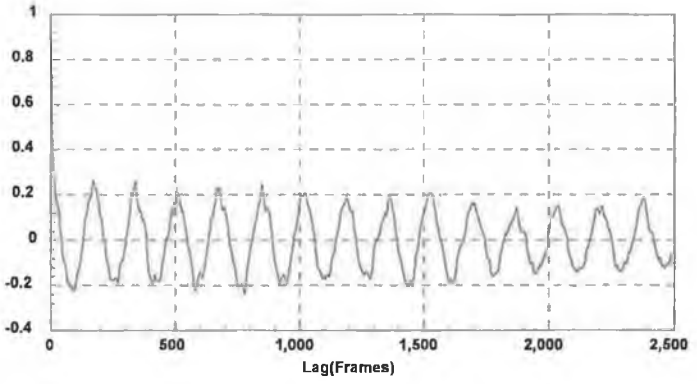


Fig. 6.20 *Long-term sample autocorrelation of traffic produced by Model #4 of the motion level strategy in {6.3} with fixed higher state durations.*

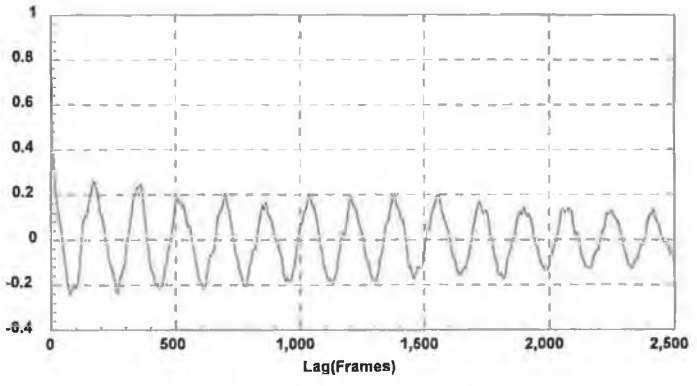


Fig. 6.21 *Long-term sample autocorrelation of traffic produced by Model #4 of the motion level strategy in {6.3} with Gaussian distributed higher state durations with standard deviation = 5 frames.*

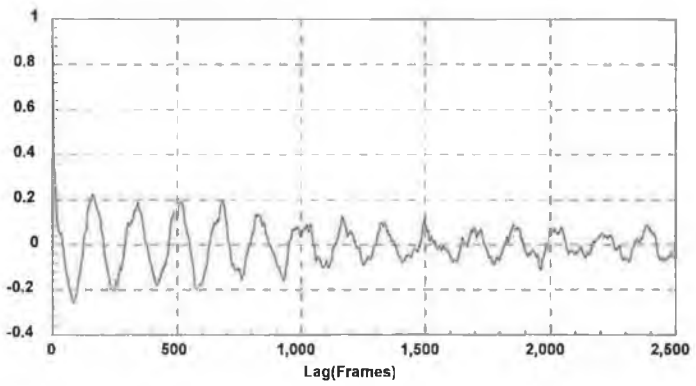


Fig. 6.22 *Long-term sample autocorrelation of traffic produced by Model #4 of the motion level strategy in {6.3} with Gaussian distributed higher state durations with standard deviation = 10 frames.*

distributed random variable, i.e. μ_{HMD} or μ_{LMD}. When such large deviations from the mean duration within each higher layer state of the hierarchical model are considered with respect to the deviations applied in Figs. 6.21 and 6.22, their impact would be extremely significant. The long-term correlations would decay very fast to zero smoothing any long-term correlations present in the simulated traffic. This effect would also apply to the Non-correlated traffic long-term sample autocorrelation in Fig. 6.19. Therefore, one may deduce from this analysis that in order to preserve the long-term correlations inherent in simulated 1-Layer VBR H.261 Video Codec traffic, variations in duration within each traffic generating state must be restricted.

6.6 Discussion

Reflecting on the simulation results in {6.3} and {6.4} and comparing these results to 1-Layer VBR H.261 Video Codec traffic statistics in Table 4.2 and Chapter 4, the performance of the stochastic models are extremely good. Performance in the mean was exceptionally good from models in {6.3} with coefficients estimated by a combination of the Ratio Method and Method of Moments(Appendix A and Appendix C) and a combination of the Least Squares Method and Method of Moments(Appendix B and Appendix C). The standard deviations of simulated data under-estimated the standard deviation, = 5476.8 bits/frame, of 1-Layer VBR H.261 Video Codec traffic which was mainly caused by adjustment of individual state standard deviations to reduce the probability of negative traffic generation within each state. In order to capture the correlation characteristics inherent in 1-Layer VBR

H.261 Video Codec traffic, Gaussian distributed random variables must be implemented. Therefore if video traffic is of any other distribution, such as in this case where talk, listen and low motion data were all Gamma distributed, the statistical parameters associated with the Gamma distribution must be adjusted for application to a Gaussian distributed random variable. In this instance, all standard deviations were reduced by the adjustment.

On the subject of correlations, both Talk-Listen strategy and Motion Level strategy models performed very good in the short-term but in the long-term, the Motion Level strategy models gave an apparent ability to capture long-term correlations better than the Talk-Listen strategy models. In {6.3}, all short-term sample autocorrelations have strong similarities to the short-term sample autocorrelations of 1-Layer VBR H.261 Video Codec traffic in chapter 4. The Talk-Listen strategy models perform well in the mean and standard deviation but the long-term sample autocorrelations fail to exhibit any long-term correlations inherent in the simulated video traffic. The Motion Level strategy models on the other hand perform equally well in the mean and standard deviation and exhibit more prominent long-term correlations inherent in simulated video traffic. This, as already explained, was made possible by the reduction of low motion traffic deviation from the mean but it may also have been achieved by implementing the best process of traffic separation for modelling purposes. The division of 1-Layer VBR H.261 Video Codec traffic into talk and listen states is not as natural as dividing the traffic into high and low motion states. This remark is supported by the comparison of long-term sample autocorrelation results from both modelling strategies in {6.3}.

To further determine the effectiveness of the layered modelling approach to capture the longer-term correlations inherent in 1-Layer H.261 VBR Video Codec traffic, non-correlated video model simulations in {6.4} were analysed. From the results in {6.4}, both modelling strategies achieved excellent performance in the mean and standard deviation when compared to the mean and standard deviation of the sample 1-Layer VBR H.261 VBR Video Codec traffic. The data generated by these models were independent random variables therefore short-term correlation analysis would be unfruitful. The longer-term autocorrelations were of more interest and here the best performance was achieved by the Motion Level strategy layered model. The long-term correlations in both characteristics are not very prominent, unlike the long-term correlations in Fig. 4.38, but the Motion Level strategy simulated traffic exhibits a more correlated characteristic than does the Talk-Listen strategy simulated traffic. But in {6.5} the implications of varying durations within a lower layer state of a hierarchical model were investigated. This analysis revealed the large impact large deviations from the mean duration within a lower layer state have on the long-term correlations of simulated video traffic by a hierarchical model. The simulations in {6.3} and {6.4} are definitely affected by this effect because larger

deviations from the mean duration in each lower layer state to those applied in simulations in {6.5} are implemented and these simulations show the impact of smaller deviations. These simulations show increased rates of decay in long-term correlations because larger deviations from the mean duration within each lower layer state smooth the long-term correlations. Therefore, deviations from the mean duration must be restricted in hierarchical video source models to allow them to exhibit long-term correlations.

Another approach to measure the effectiveness of hierarchical models to model 1-Layer VBR H.261 Video Codec traffic is to consider modelling such traffic by a non-hierarchical model. This approach is adopted in [8, 16, 20, 33, 34, 35, 36, 39, 40] where the most popular non-hierarchical traffic model is a First Order Autoregressive process. To compare the effectiveness between both hierarchical and non-hierarchical models to capture the long-term correlations inherent in 1-Layer VBR H.261 Video Codec traffic the correlation coefficient a for a non-hierarchical First Order Autoregressive model was estimated within a lag = 500 frames by the Least Squares Method in Appendix B to approximate the long-term sample autocorrelation in Fig. 4.38. From this analysis the correlation coefficient was estimated to be

$$a = 0.96.$$

This correlation coefficient is a parameter to the autocorrelation function of a First Order Autoregressive process which is described as follows :

$$R(k) = a^k$$

where $k = 0,1,.....$ This function is shown in Fig. 6.23 for $k = 0,...,2500$ where after a lag = 100 frames the process produces zero correlations. In comparing Fig. 6.23 to

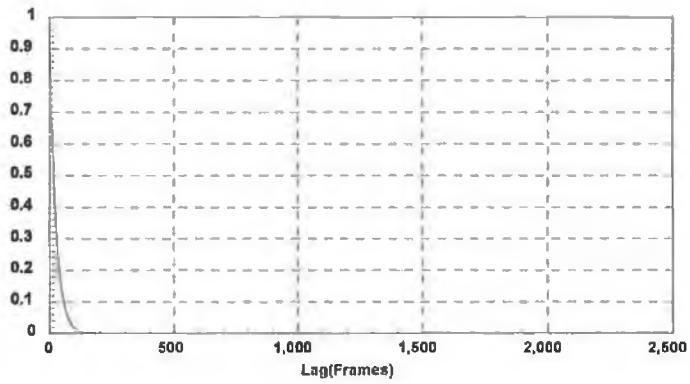


Fig. 6.23

Fig. 4.38, it is quite clear that a non-hierarchical model will not succeed in capturing long-term correlations unlike the hierarchical Discrete Autoregressive Video models which exhibit long-term correlations in Figs. 6.14 - 6.17 and 6.19 - 6.22. Therefore this analysis indicates the superiority of hierarchical models to model 1-Layer VBR H.261 Video Codec traffic.

7. Conclusions and Recommendations

7.1 Conclusions

In [3], several parameters are recommended to characterise 1-Layer VBR H.261 Video Codec traffic(see {4.4}). These parameters are guidelines for the process of video traffic characterisation and are also recommended because video traffic is primarily assumed to be Gaussian distributed based upon previously published studies[8, 9, 22, 20]. This has occurred in [21, 24, 48] where only the statistics of the analysed video codec traffic have been published without determining the distribution of the analysed traffic. Also all analysis, with exception to analyses in [21, 24, 48], was carried out on traffic from alternative video coding algorithms. Analysis here was carried out on variable rate traffic from the adapted H.261 Video Coding Standard. The traffic distribution of 1-Layer VBR H.261 Video Codec traffic in Fig. 4.39 was found not to resemble a Gaussian distribution. Therefore the analysis based upon acquiring the recommended parameters to characterise variable rate video traffic is not adequate unless an indication to the traffic distribution is also included. At this point it could be said that Gaussian distributed traffic might result if the sample video sequence analysed was longer in duration but long duration simulations involving video sources would be ineffective requiring heavy processing.

To overcome this problem alternative techniques were applied to statistically characterise 1-Layer VBR H.261 Video Codec traffic. The objective was to statistically characterise and stochastically model 1-Layer VBR H.261 Video Codec traffic which would subsequently lead to a 2-Layer H.261 Video Codec traffic model. To achieve this aim the contents of 1-Layer VBR H.261 Video Codec coded video sequences were taken into consideration. The contents of video sequences were noticed to vary depending on the audio-visual service being considered : videophone; videoconference; television. Videophone sequences were primarily of interest and were assumed the simplest audio-visual sources to statistically model and characterise. On analysing videophone sequences, directly(subject content) and indirectly(video traffic), two different output traffic processes were clearly inherent in videophone sequences giving rise to longer-term autocorrelations in video sources(see Fig. 4.38).

Direct analysis was thought to have identified different data processes as a result of the talking and listening periods within a videophone sequence. This analysis was expected to reveal a higher and lower mean output traffic level for

talking and listening periods respectively but the analysis in Chapter 4(see {4.2.3}), under the Talk-Listen strategy, indicates that this strategy did not discriminate effectively between the two inherent processes. The mean output traffic for both talking and listening periods were approximate. The standard deviations of traffic from the mean for both talking and listening periods were also approximate and their corresponding traffic distributions were also of similar distribution as determined in Chapter 4. Therefore identifying two inherent processes by subject content analysis within video sequences based upon their statistics revealed poor correlation between output traffic and video sequence subject content, i.e. large traffic periods did not correlate with talking periods nor did low traffic periods relate to listening periods.

Indirect analysis of traffic from the 1-Layer VBR H.261 Video Codec in Fig. 4.22 clearly depicts two levels of mean. Periods of motion are difficult to quantify during a video sequence but the output data captures the effects of motion on the video coding process where periods of high output traffic relate to periods of high motion in a video sequence and vice versa in periods of low motion. This is reflected in the mean traffic level for periods of high motion which is approximately twice that for periods of low motion(see {4.2.4}). The standard deviations for traffic about their respective means during each motion level is approximate in both cases but the traffic distributions were found to differ. High motion traffic distribution was approximately Gaussian after the traffic histogram was shifted during analysis in chapter 4 to accommodate a Gaussian probability density function approximation. Low motion traffic was found to be Gamma distributed after shifting its histogram during analysis in chapter 4. The distributions clearly agree with the coding process where during low motion periods the majority of macroblocks are Interframe coded while during high motion periods Intraframe coding is applied to the majority of macroblocks during the Hybrid coding process.

Traffic autocorrelation analysis for each state(Talk, Listen, High Motion, Low Motion)(see {4.2.3.1.2}, {4.2.3.2.2}, {4.2.4.1.2} and {4.2.4.2.2}) also revealed differences in analysis strategies. The short-term sample autocorrelation for Talk and Listen traffic exhibit, what could only be described as 'general' First Order Autoregressive processes(see Figs. 4.11 and 4.20). The short-term sample autocorrelations in the Motion Level strategy analysis clearly exhibit stronger autoregressive autocorrelation characteristics. This may be associated with the success of isolating and separating traffic containing two distinct different levels of motion. Therefore the Talk-Listen strategy autocorrelation analysis also clearly disagrees with the assumption that high motion is restricted to talking periods alone and low motion is restricted to listening periods alone within a videophone sequence.

Autocorrelation analysis of concatenated traffic produced by a single state(Talk, Listen, High Motion or Low Motion) did not clearly indicate the independence of video traffic during single Talk/Listen/High Motion/Low Motion

durations due to longer term correlations being present which were mainly sporadic, exhibiting complex autocorrelations. This is a result of the video coding process which is mainly influenced by the motion content of video sequences where high correlation periods relate to periods of activity of similar level within a video sequence. To simplify the modelling process and to avoid complexities introduced by the presence of sporadic autocorrelations, video traffic produced within each state(Talk, Listen, High Motion or Low Motion) was assumed independent from video traffic produced by the previous state.

The distribution of time(in frames) a videophone sequence remained in a state(Talk, Listen, High Motion, Low Motion) was approximated by several probability density functions determined by analysis in Chapter 4. These higher level processes in video traffic in {6.5} were clearly seen to be responsible for the longer-term correlations inherent in video traffic and are widely referred to as scene changing processes. In both analysis strategies implemented, the two alternating states in each case were generally regarded as two alternating scene types.

In Chapter 5, several models were introduced to model 1-Layer VBR H.261 Video Codec traffic. The models are hierarchical to enable the capture of long-term and short-term correlations inherent in video traffic which are in the order of seconds and milliseconds respectively. Both Discrete and Continuous Time models were presented. Several models were forwarded for implementation but all were not assessed to measure their capability to model 1-Layer VBR H.261 Video Codec traffic. One model in particular posessed the desired characteristics and was subsequently implemented in simulations, the DAV(Discrete Autoregressive Video) model(see {5.3.4.3(iv)}).

The simulated model results in Chapter 6, for both strategies Talk-Listen and Motion Level, achieve good approximations to the mean of 1-Layer VBR H.261 Video Codec traffic for each individual state within each strategy. The standard deviation also performs quite well but not to the required standard deviation due to adjustments to minimise the occurrence of negative traffic which was necessary to avoid. Several methods were implemented to estimate parameter values for the First Order Autoregressive processes within the states of the DAV model for each strategy. Estimates for parameters by methods in Appendices A, B and C resulted in quite satisfactory approximations, in particular for the Motion Level strategy. The autocorrelation analysis of simulated 1-Layer VBR H.261 Video Codec traffic showed good correlation performance in the short-term for both strategies. The analysis of long-term sample autocorrelations though highlighted a better performance by the Motion Level strategy, in particular when the low motion traffic standard deviation was adjusted to prevent negative traffic occurring.

Further analysis in {6.4} to determine which 1-Layer VBR H.261 Video Codec traffic modelling strategy performs better in the higher layer of the DAV

model further produced evidence of the superior performance in the long-term the Motion Level strategy has over the Talk-Listen strategy. The simulation of non-correlated short-term processes in {6.4}, i.e. independent random variables within each DAV model state, further indicated the ability of the Motion Level strategy to capture long-term autocorrelations.

The impact of varying durations in the higher layer of the model was found to cause suppression and smoothing of long-term autocorrelations as shown in {6.5}. In this analysis it was found that variations around the mean duration within a higher layer state of a hierarchical model must be restricted. If variations around the mean duration are large, as was found in the analysis in chapter 4, the effect on the long-term autocorrelations inherent in the simulated traffic increases the rate of autocorrelation decay and smoothes the long-term autocorrelations. Therefore, by limiting variations around the mean duration present in a higher layer state as shown in {6.5}, the long-term autocorrelations inherent in hierarchical model simulated traffic can be preserved.

Statistical analysis of simulated traffic by both strategies shows very good performance both in the mean and standard deviation. Hierarchical models also model the autocorrelations of 1-Layer VBR H.261 Video Codec traffic better, as shown in {6.6}, but the Motion Level strategy modelling approach gives better performance than the Talk-Listen strategy modelling approach.

7.2 Recommendations for the future

To further strengthen the case that hierarchical models are the best approach to statistically characterise and stochastically model video sources, the additional videophone type(head and shoulders, waist-up) sequences that were recorded, given in Table 4.1, must be processed by a 1-Layer VBR H.261 Video Codec and also analysed. Analysis should implement both strategies, Talk-Listen and Motion Level, presented in this research to determine fully whether the Motion Level strategy models have superior performance. The results of the analysis can be compared for each sequence and it will also determine whether the statistics, coefficients and distributions applied in this research were satisfactory. Gender analysis could also be researched.

Due to videophone and videoconference services being interactive conversational services one party may excert influence on the other interacting party, a characteristic that can be observed every day among people. This from a modelling point of view would reflect reality but to avoid model complexities it would be more attractive to keep each party within an interactive service independent. This effectively has been done in the modelling already carried out in Chapter 6 where it is assumed each scene type(Talk, Listen, High Motion and Low Motion) is

independent from the previous scene type simplifying the modelling process otherwise more complex models would have been required.

On the note of satisfactory approximations to duration and traffic distributions in chapter 4, the Norm Variation function(4.2) is only a general goodness of fit test function, therefore more stringent goodness of fit test functions would be necessary to obtain a definite measure of approximation.

To further improve the accuracy of the model implemented in chapter 6, asymmetric Gaussian distributions may be implemented to replace the Gamma distributions for talk, listen and low motion traffic. The simulations for talk, listen and low motion traffic in chapter 6 applied symmetric Gaussian distributions with reduced standard deviations to prevent negative traffic occurring. This approach satisfies the traffic deviation below the mean but traffic above the mean can achieve higher deviations, particularly in low motion, talk and listen states as seen clearly in their respective traffic distributions. Therefore to improve the realism of the models, asymmetric Gaussian distributions could be considered to replace the applied Gamma distributions.

This research applied the DAV model in Chapter 5 to simulate 1-Layer VBR H.261 Video Codec traffic. Other models presented in Chapter 5 must be investigated, particularly models that will be of benefit to queuing analysis e.g. Discrete state continuous time models and fluid flow models. Other models presented in Chapter 5 may perform better both in the short and long-terms.

To engage these models in ATM network simulations, the simulated traffic must be packetised. The ATM network transports' cells(see {2.2.3.1}) but the presented models simulate video traffic on a bits/frame level. Hence a strategy based on the ATM cell standard must be introduced into the model to pack the simulated traffic into cells. Similarly this can be applied to the recorded videophone type sequences in Table 4.1. This will permit the application of real 1-Layer VBR H.261 Video Codec traffic to network analysis and simulation. Therefore, ATM B-ISDN simulations can be carried out implementing video source models alone or recorded videophone type sequences alone(similar to the technique applied in [34] where one sequence created several sources by varying the start time and start frame in the video sequence) or in combination. These simulations will determine the impact video services will have on network resources(bandwidth, buffer dimensions, transfer delay, switch dimensions, etc.) which will help create techniques to control the Quality of Service(Probability of Cell Loss) of the network.

The lowest level of discretisation reached in this research was the frame level. Since the coding process is applied on the macroblock level(see {3.2.2}), therefore the lowest level of discretisation must be the macroblock level. Macroblocks are either Intraframe or Interframe coded depending on the variation of the macroblock contents(see Fig. 3.3 from the RM8 model). Therefore, macroblocks could be

separated according to the coding process applied which could possibly vary stochastically for both coding techniques from frame to frame.

This research has statistically characterised and stochastically modelled 1-Layer VBR H.261 Video Codec traffic but future video communication terminals are expected to be 2-Layer H.261 Video Codecs to maintain equipment compatibility on the ATM B-ISDN. This research outlines a technique to model 2-Layer H.261 Video Codec traffic via 1-Layer VBR H.261 Video Codec simulated traffic(see {5.4}). This approach to 2-Layer H.261 Video Codec traffic modelling must first be investigated by statistical analysis and if successful can be applied to ATM B-ISDN network simulations.

References

[1] Spragins, J.D., Hammond, J.L., and Pawlikowski, K., "Chapter 12 : Integrated Services Digital Networks", *Telecommunications : Protocols and Design.* Addison-Wesley Publishing Company, Inc., 1991.

[2] Slana, M.F., "Chapter 12 : Integrated Service Digital Network", *Fundamentals of Digital Switching*, edited by J. McDonald, Plenum Press, 1990 New York.

[3] Project COST211bis Final Report, *Redundancy Reduction Techniques for Coding of Broadband Video Signals.*

[4] Sato, Y. and Sato, K.,"Virtual Path and Link Capacity Design for ATM Networks", *IEEE Journal on Selected Areas in Communications*, Vol.9, No.1, pp.104-110, January 1991 .

[5] Sato, K., Ohta, S. and Tokizawa, I.,"Broad-Band ATM Network Architecture Based on Virtual Paths", *IEEE Transactions on Communications*, Vol.38, No.8, pp.1212-1221, August 1990.

[6] Choi, J.K., Choi, M., Chung, T.S., Shin, Y.S. and Kim, K.S.,"Overall Design requirements of Broadband ATM Networks", *GLOBECOM '91*(4B.3.1 - 4B.3.6).

[7] Gallassi, G., Rigolio, G. and Verri, L.,"Resource Management and Dimensioning in ATM Networks", *IEEE Network Magazine*, pp.8-17, May 1990.

[8] Maglaris, B., Anastassiou, D., Sen, P., Karlsson, G. and Robbins, J.D.,"Performance Models of Statistical Multiplexing in Packet Video Communications", *IEEE Transactions on Communications*, Vol.36, No.7, pp.834-843, July 1988.

[9] Verbiest, W., Pinnoo, L. and Voeten, B., "The Impact of the ATM Concept on Video Coding", *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 9, pp.1623-1632, December 1988.

[10] Karlsson, G. and Vetterli, M., "Packet Video and Its Integration into the Network Architecture", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, pp.739-751, June 1989.

[11] Darragh, J.C. and Baker, R.L., "Fixed Distortion Subband Coding of Images for Packet-Switched Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, pp.789-800, June 1989.

[12] CCITT Recommendation I.430, "Basic User-Network Interface - Layer 1 Specification", *International Telecommunication Union Blue Book*, Vol. III, Fascicle III.8.

[13] CCITT Recommendation I.431, "Primary Rate User-Network Interface - Layer 1 Specification", *International Telecommunication Union Blue Book*, Vol. III, Fascicle III.8.

[14] CCITT Recommendation I.412, "ISDN User-Network Interfaces, Interface Structures and Access Capabilities", *International Telecommunication Union Blue Book*, Vol. III, Fascicle III.8.

[15] CCITT Recommendation I.121, "Broadband Aspects of ISDN", *International Telecommunication Union Blue Book*, Vol. III, Fascicle III.7.

[16] Yasuda, Y., Yasuda, H. and Ohta, N., "Packet Video Transmission through ATM Networks", *Proc. GLOBECOM 1989*, pp.876-880, 27-30 Nov. 1989, Dallas.

[17] Marafih, N.M., Zhang, Y.-Q. and Pickholtz, R.L., "Modelling and Queueing Analysis of Variable-Bit-Rate Coded Video Sources in ATM Networks",(Source unknown).

[18] Rodriguez-Dagnino, R.M., Khansari, M.R.K. and Leon-Garcia, A., "Prediction of Bit Rate Sequences of Encoded Video Signals", *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 3, pp.305-314, April 1991.

[19] Kishino, F., Manabe, K., Hayashi, Y. and Yasuda, H., "Variable Bit-Rate Coding of Video Signals for ATM Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, pp.801-806, June 1989.

[20] Nomura, M., Fujii, T. and Ohta, N., "Basic Characteristics of Variable Rate Video Coding in ATM Environment", *IEEE Journal on Selected Areas in communications*, Vol. 7, No. 5, pp.752-760, June 1989.

[21] UK(BTRL,) "Total bit rate requirements of a 2-layer video codec hardware for differing base layer fixed bit rates", *COST211ter Sim/Hw Document No. 91/36*, Madrid, April 1991.

[22] Chin, H.S., Goodge, J.W., Griffiths, R. and Parish, D.J., "Statistics of Video Signals for Viewphone-Type Pictures", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, pp.826-832, June 1989.

[23] Ghanbari, M., "Two-Layer Coding of Video Signals for VBR Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, pp.771-781, June 1989.

[24] Murphy, S., Dublin City University, "Simulation of H.261 based Codecs", *COST242TD(093)*, Warsaw, 24 October 1993.

[25] UCL Belgium, "Requirements analysis for video services : Bit Rate Statistics of TV Distribution Codecs in the VBR Mode", *COST211ter Document No. 92/46*, Hannover, April 1992.

[26] Yegenoglu, F., Jabbari, B. and Zhang, Y.-Q., "Motion-Classified Autoregressive Modeling of Variable Bit Rate Video", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 3, No. 1, pp.42-53, February 1993.

[27] Lei, S.-M., Chen, T.-C. and Tzou, K.-H., "Subband HDTV Coding Using High-Order Conditional Statistics", *IEEE Journal on Selected Areas in Communications*, Vol. 11, No. 1, pp.65-75, January 1993.

[28] Argenti, F., Benelli, G. and Mecocci, A., "Source Coding and Transmission of HDTV Images Compressed with the Wavelet Transform", *IEEE Journal on Selected Areas in Communications*, Vol. 11, No. 1, pp.46-58, January 1993.

[29] Kim, J.T., Lee, H.J. and Choi, J.S., "Subband Coding Using Human Visual Characteristics for Image Signals", *IEEE Journal on Selected Areas in Communications*, Vol. 11, No. 1, pp.59-64, January 1993.

[30] Zdepski, J., Raychaudhuri, D. and Joseph, K., "Statistically Based Buffer Control Policies for Constant Rate Transmission of Compressed Digital Video", *IEEE Transactions on Communications*, Vol. 39, No. 6, pp.947-957, June 1991.

[31] Liu, J., Dublin City University, "A Report of Video Compression Using Multi-Level Fractal Block Coding Techniques", *COST211ter SIM(92)100*, Dublin, December 1992.

[32] Dublin City University Video Coding Group, "Simulation Model for Object-based Coding", *COST211ter SIM(93)50*, Lausanne, October 1993.

[33] Roberts, J.W., Guibert, J. and Simonian, A., "Network Performance Considerations in the Design of a VBR Codec", *Queueing, Performance and Control in ATM (ITC-13)*, Cohen, J.W. and Pack, C.D. (editors), Elsevier Science Publishers B.V. (North-Holland), pp.77-82.

[34] Heyman, D.P., Tabatabai, A. and Lakshman, T.V., "Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 2, No. 1, pp.49-59, March 1992.

[35] Huang, S-S., "Source Modelling for Packet Video", *IEEE Inter. Conf. on Communications 1988*, pp.1262-1267, 12-15 June, Philadelphia, USA.

[36] Cohen, D. M. and Heyman, D. P., "A Simulation Study of Video Teleconferencing Traffic in ATM Networks", *Proc. IEEE INFOCOM '93*, Vol. 3, pp.894-901, 28 March - 1 April, San Francisco.

[37] Frater, M. R., Tan, P. and Arnold, J. F., "A New Statistical Model for Traffic Generated by VBR CODECs for Television on the Broadband ISDN", *Technical Report*, Dept. of Electrical Engineering, University College, University of New South Wales, Australian Defence Force Academy, Canberra ACT 2600, Australia, April 1993.

[38] Sen, P., Maglaris, B., Rikli, N-E. and Anastassiou, D., "Models for Packet Switching of Variable-Bit-Rate Video Sources", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, pp.865-869, June 1989.

[39] La Corte, A., Lombardo, A., Palazzo, S. and Zinna, S., "Modeling activity in VBR sources", *Signal Processing : Image Communication*, Vol. 3, Pt. 2-3, pp.167-178, June 1991.

[40] Huang, S-S., "Modelling and Analysis for Packet Video", *Proc. GLOBECOM '89*, 27-30 November 1989, Dallas, USA.

[41] Filipiak, J., "Structure of Traffic Flows in Multiservuce Networks", *Australian Telecommunications Research*, Vol. 23, No. 1, pp.5-18, 1989.

[42] Gihr, O. and Tran-Gia, P., "A Layered Description of ATM Cell Traffic Streams and Correlation Analysis", *Proc. GLOBECOM '91*, pp.137-144.

[43] Verbiest, W., "Video Coding in the ATD Environment", *Proc. Third Inter. Conf. on New Syst. and Services in Telecommunications*, Liege, Belgium, November 1986.

[44] Kroner, H., "Statistical Multiplexing of Sporadic Sources - Exact and Approximate Performance Analysis", *Queueing, Performance and Control in ATM (ITC-13)*, Cohen, J.W. and Pack, C.D. (Editors), Elsevier Science Publishers B.V. (North-Holland)

[45] Roberts, J.W., "Variable-Bit-Rate Traffic Control in B-ISDN", *IEEE Communications Magazine*, Vol. 29, No. 9, pp.50-56, September 1991.

[46] Botvich, D.D., Curran, T., MacFhearraigh, A. and Murphy, S., "Hierarchical Approach to Video Source Modelling", *Proc. of the 11th UK Teletraffic Conference*, Cambridge, 23-25 March 1994.

[47] William H. Press, Brian P. Flannery, Saul A. Teukolsky and William T. Vetterling, *Numerical Recipes in C : The Art of Scientific Computing*, Cambridge University Press, 1988.

[48] UK(BTRL), "Comparison of bit rates for a 2-Layer video codec hardware with differing quantisers for the enhancement layer", *COST211ter Sim/Hw Document No. 91/37*, Madrid, April 1991.

[49] Athanasios Papoulis, *Probability, Random Variables and Stochastic Processes,* Third Edition, McGraw-Hill International Editions Electrical & Electronic Engineering Series, 1991.

[50] CCITT SGXV Working Party XV/4 Specialist Group on Coding for Visual Telephony, *Description of Reference Model 8(RM8)*, Document 525, 1989.

[51] ISO 7498 : Open Systems Interconnect Basic Reference Model.

[52] ITU-T I.100-Series Recommendations : General ISDN Concept.

[53] ITU-T Recommendation I.211," ISDN Service Capabilities : B-ISDN Service Aspects", 1993.

[54] CCITT Recommendation G.707, "Synchronous Digital Hierarchy Bit Rates", 1992.

[55] CCITT Recommendation G.709, "Synchronous Multiplexing structure", 1992.

[56] Ferguson, S.P., "Implications of SONET and SDH", *IEE Electronics & Communications Journal*, Vol. 6, No. 3, pp.133-142, June 1994.

[57] Popple, G. and Glen, P., "Specification of the Broadband User-Network Interface", *IEE Electronics & Communications Journal*, Vol. 6, No. 2, pp.105-112, April 1994.

[58] Jeffrey, M., "Asynchronous Transfer Mode: the ultimate broadband solution?", *IEE Electronics & Communications Journal*, Vol. 6, No. 3, pp.143-151, June 1994.

Appendices

Appendix A : The Ratio Method

For a First Order Autoregressive process

$$x(n) = ax(n-1) + bw(n) \tag{A.1}$$

where $w(n)$ is an independent Gaussian distributed random variable with mean = μ_w and standard deviation = σ_w, the coefficient a is a measure of the correlation between data produced by this stochastic process. If $\hat{R}(k)$, $k = 0,...,K$, is the sample autocorrelation of a real data set, $\hat{x}(n)$, $n = 0,...,N$, therefore for $R(k)$, $k = 0,...,K$, to approximate $\hat{R}(k)$, where $R(k)$ is the sample autocorrelation of the model data set, $x(n)$, $n = 0,...,N$, then the coefficient a in (A.1) can be estimated by

$$a = \frac{\sum_{i=1}^{K} \hat{Z}(i)}{K}$$

where

$$\hat{Z}(i) = \frac{\hat{R}(i)}{\hat{R}(i-1)}.$$

This estimation is based on the relationship between $x(n)$ and $x(n-1)$ in (A.1) where the coefficient a is effectively a measure of the memory within the process i.e. the present output depends on the previous output by a factor of a. The estimation technique obtains an average value for a over the full sample autocorrelation range K to reasonably satisfy all correlations within the sample range. This estimation technique also depends on the noise source $bw(n)$, whereby if the noise source is negligible the estimation technique would be reasonably accurate and on the other hand if the noise source could not be neglected, this technique would not yield good estimations for a.

Appendix B : The Least Squares Method

The steady state autocorrelation of output data from a Discrete Time Difference Equation with a Noise Source at the input(see (A.1)) is as follows from [49] :

$$R_x(k) = \frac{\mu_w{}^2 b^2}{(1-a)^2} + \frac{b^2 \sigma_w{}^2 a^k}{1-a^2}.$$

(B.1)

The first term in (B.1) is the steady state mean of $x(n)$ in (A.1) squared i.e. $\mu_x{}^2$, and the second term in (B.1) equates to $\sigma_x{}^2 a^k$ because $\sigma_x{}^2 = \frac{\sigma_w{}^2 b^2}{1-a^2}$. Therefore, if $\mu_w = 0$ and $\sigma_w = 1$, then

$$R_x(k) = \sigma_x{}^2 a^k.$$

(B.2)

The autocorrelation of the process therefore only depends on the rate of decay controlled by the coefficient a in (A.1).

To estimate a the following procedure is followed. Initially a is known to be within the range $0 \le a \le 1$. By advancing a up through this range in steps of h and applying a to (B.2), a can be estimated by identifying the minimum error between $\hat{R}(k)$, the sample autocorrelation of a real data set, $\hat{x}(n)$, $n = 0,...,N$, and $R_x(k)$ for various values of a by the following technique :

$$n(a') = \sum_{k=0}^{K} \left(\hat{R}(k) - R_x{}^{a'}(k) \right)^2,$$

where $R_x{}^{a'}(k) = R_x(k)$ with $a = a'$. The minimum $n(a')$ between both correlations gives $a = a'$ which is estimated to five decimal places.

Appendix C : The Method of Moments

The other coefficient b of the First Order Autoregressive process in (A.1) can be estimated by implementing both the Method of Moments and the coefficient a estimated by either Appendix A or B. The Method of Moments derives that

$$\sigma_x{}^2 = \frac{\sigma_w{}^2 b^2}{1-a^2}.$$

(C.1)

Therefore when (A.1) has an independent Gaussian distributed random variable source with mean, μ_w, and standard deviation, $\sigma_w = 1$, then from (C.1)

$$b = \sigma_x \sqrt{1-a^2}.$$

(C.2)

Let $\sigma_x = \hat{\sigma}_x$, where $\hat{\sigma}_x$ is the standard deviation of the data sample, $\hat{x}(n)$, $n = 0,...,N$. The Method of Moments is further implemented to estimate μ_w which is necessary for the source of (A.1). By the Method of Moments

$$\mu_x = \frac{\mu_w b}{(1-a)}$$

therefore

$$\mu_w = \frac{\mu_x (1-a)}{b} \tag{C.3}$$

and by letting $\mu_x = \hat{\mu}_x$, the mean of the data sample $\hat{x}(n)$, $n = 0,...,N$, μ_w can be derived.

From Appendix B, μ_w was set to 0 for convenience because it only performs a data shifting role within the process i.e. $RV_{(\mu w = s,\ \sigma w = 1)} = RV_{(\mu w = 0,\ \sigma w = 1)} + s$, where $RV_{(\mu w = 0,\ \sigma w = 1)}$ is a Normal distributed random variable and s is the shift. Therefore from (C.2) and (C.3) the b and μ_w parameters for (A.1) can be estimated.