DYNAMIC THROUGHPUT IN DISTRIBUTED MULTIMEDIA

A RESEARCH REPORT
SUBMITTED TO THE FACULTY OF SCIENCE
AND THE DEPARTMENT OF COMPUTER SCIENCE,
UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Ronnie T. Apteker
April 1994

Research Report

Ronen Theodor Apteker
86-04099/N
April, 1994

Research report for the research component of
Master Of Science by coursework and research report.

Department of Computer Science
University of the Witwatersrand

Supervisors : Prof. Hanoch Neishlos
Dr. Valentin Kisimov

Title : Dynamic Throughput in Distributed Multimedia

Dynamic Throughput in Distributed Multimedia

ABSTRACT

Personal computing is currently undergoing radical enhancements with the current technological advancements that have been made in the areas of high resolution displays, GUIs (graphic user interfaces), high quality sound and full motion video. Multimedia stands at the convergence of these technological advances.

The pervasiveness of networks will result in a new generation of distributed services that include multimedia as the fundamental characteristic. The current hyper-activity in the commercial arena is testimony to the future of distributed multimedia services. The anticipation of the data superhighways has led to an industrial scramble filled with takeovers and acquisitions as companies battle to acquire the infrastructure that will set the scene for the services of the future.

This report describes the technologies encompassing distributed multimedia services. Video compression hardware and high-speed networking is the brute force that is required to deliver multimedia services in a distributed environment. However, the large volumes of data that are characteristic of multimedia information is the foremost problem in the realization of on-demand, continuous media services. The success of a multimedia server is determined by the number of users that can be serviced simultaneously. A distributed service is not effective if it can only be used by a small number of users at one particular point in time.

The quality of service (Q_oS) of a distributed multimedia application is described in terms of spatial and temporal parameters. These parameters determine the Q_oS of a multimedia delivery. Maintaining a constant level of quality, or constant throughput, requires a system that guarantees a specified minimal network bandwidth. This requires a capacity based protocol that reserves resources for the establishment of a real-time session. The success of a multimedia server depends on the number of real-time sessions that can be established concurrently.

Dynamic throughput, or dynamic Q_oS, yields better network utilization by deliberately degrading the Q_oS of existing continuous media connections so that new user requests can be accommodated. This distributed multimedia approach involves dynamically controlling Q_oS parameters such as colour depth, frame rate and audio sampling rate. It is obvious that a degradation of the physical characteristics of a real-time service will result in perceived differences in video acceptability. But the relationship between video message importance and video degradation is not that obvious. Through

experimentation it is possible to gain insight into what type of distributed video applications will be more susceptible to a degradation in data throughput.

This paper examines the effects that dropping the frame rate of a distributed multimedia application has on user perception. Findings support the fundamental premise that frame rate reduction itself leads to progressively lower ratings of acceptance, which erodes with each stepwise decrease in the experimental frame rates. Furthermore, findings show that degrading the Q_oS (i.e. adjusting the throughput) of a multimedia session has different consequences for the network bandwidth requirements of different multimedia applications, which is directly related to the number of users that can be supported concurrently by a multimedia server. Results are described in terms of network bandwidth and distributed video applications.

Dynamic throughput of multimedia is by no means an intermediate step in the realization of unconstrained, continuous media on-demand services. Adaptive algorithms will continually be required since there are no upper bounds on the complexity of user requirements.

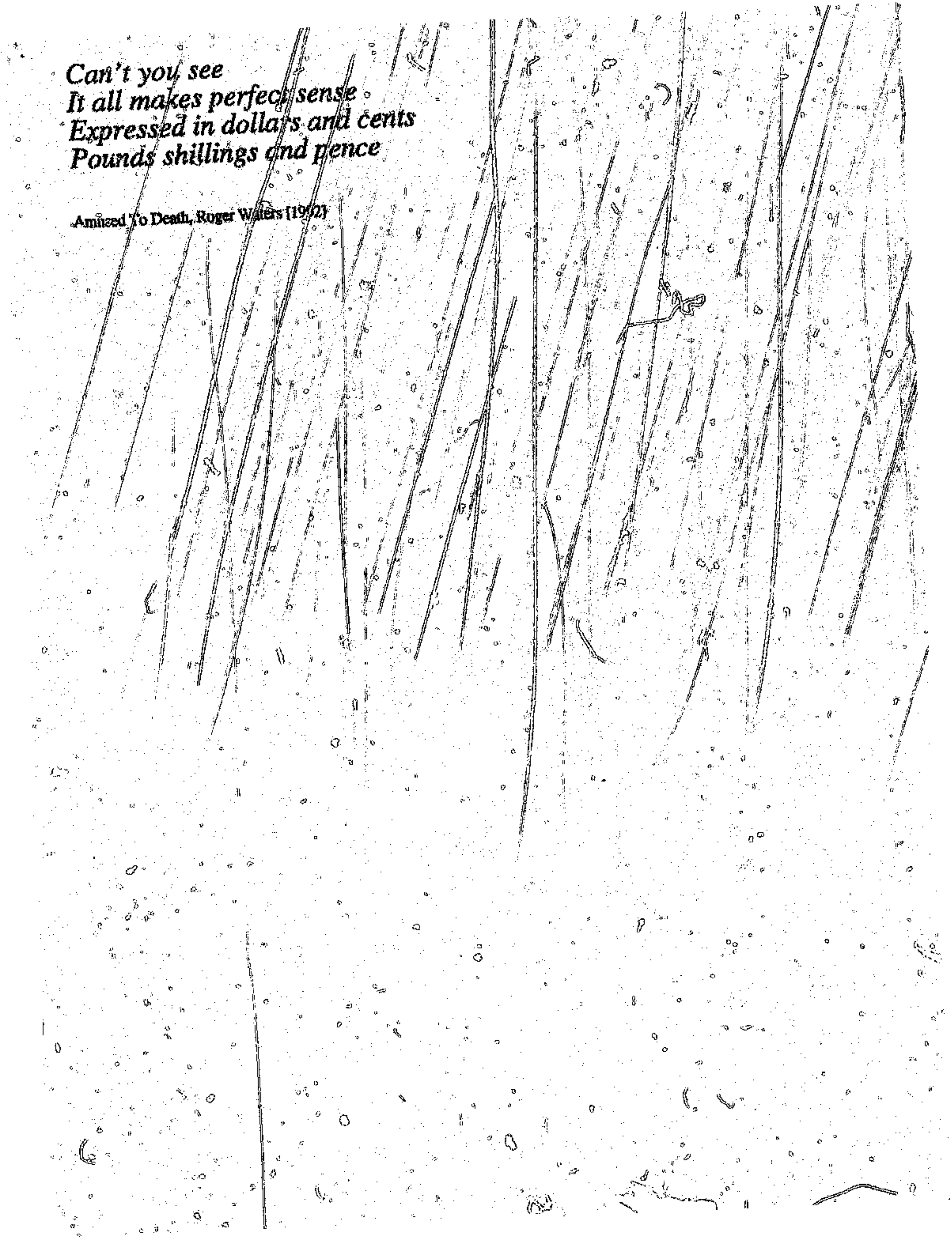
I declare that this research report is my own, unaided work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

Ronnie T. Apteker

20 day of Apr, 19 94.

Acknowledgments

I would like to thank my two supervisors, Hanoch Neishlos and Valentin Kisimov. Their encouragement and patience allowed me to excel in areas far beyond the conventions of this degree. Valentin has been an inspiration to me since the day I met him His positive outlook and motivation made these past 2 years a worthwhile experience, and it has given me strength in new and wonderful areas of life. I thank Hanoch for his wisdom and straight-forward approach to solving problems, and for supporting me in San Jose in February when this work was presented at a conference. Both Hanoch and Valentin have constantly given me ideas and suggestions that have always been interesting and inspirational. I would also like to thank James Fisher of the psychology department for his help in designing the experiment and for his humour and wisdom.



Can't you see
It all makes perfect sense
Expressed in dollars and cents
Pounds shillings and pence

Amused To Death, Roger Waters [1992]

1. INTRODUCTION

Television is the landmark of this century. We need it to communicate. We turn to it daily to be entertained, educated, or simply kept informed of current happenings. Television sits at the heart of our emerging cyber-culture. It spans the world of the familiar, stretching far beyond to the bizarre and the unreal. Our desire to watch has propelled an industry eager to deliver an overwhelming abundance of moving pictures. Yet, in this, the era of "high-tech", it is perhaps almost shocking that the technological infrastructure currently established for the delivery of video information has received minimal change over the past 40 years. The micro-computer explosion that has changed everything from the wristwatch to the motorcar has yet to fully penetrate the television.

The computer distorts the act of viewing motion pictures on a screen as it has traditionally been perceived; so far, that it is now called a computer application, and not a television channel. In essence, a virtual television (an application window) is constructed, which when connected to a data network, can be programmed or scheduled, similar to the way that television networks program their line-ups.

The incorporation of multimedia into distributed applications is the latest challenge in the world of digital computing. Traditionally separated technologies are converging to create new environments that are a prelude to the super-highways that have sent telecommunications, computer and media companies into a bidding war to build the infrastructure necessary for the multimedia on-demand services of the future. The goal : to build a multimedia on-demand server capable of supporting an unlimited number of concurrent users. Described as TV à la carte and virtual VCRs, these distributed services have to offer more than traditional analog broadcast television if they are to be accepted as the services of the future. This requires quality of service (Q.S) guarantees and a digital infrastructure capable of supporting the large volumes of data that are typical of multimedia information.

The motivation for developing digital video communications is simply : desire. The human desire to control the medium, rather than being controlled by it - the freedom to watch what you want, and when you want.

The development of a digital video infrastructure requires the integration of traditionally separated technologies. Although the backdrop may be computers, it is rare that networking experts are experienced in the field of operating systems, and that data compression experts are experienced in the field of communication protocols, etc. Digital computing is the home of a vast many technologies from both the hardware and software arena. Continuous media stands at the convergence of these technological advances.

Digital video, or continuous media, implies data that are time constrained. Digital video consists of two continuous media streams : audio and visual. These data streams have to be delivered in a predictable and synchronized manner. Text based services, by contrast, are not time constrained. Text that arrives at unpredictable times is still text. Its semantics are unaffected by definition. Distributed services that include continuous media covers considerably more parameters than traditional text based services. This report explores some of these parameters.

A multimedia server, or digital video server, is characterized by large volumes of data. A distributed multimedia system is described by data compression schemes, resource managers (workstation and network), and media synchronization mechanisms. Network resource managers maintain network resources by admitting user requests based on available bandwidth and established multimedia sessions. Network admission control is necessary since a new request is only admitted if there is enough network capacity available to create the new session without affecting the Q_oS of existing sessions.

Resources need to be guaranteed for the establishment of a real-time session. Digital video's data requirements have obvious implications for network traffic. Traditional network protocols do not accommodate for data which are time constrained. A capacity based protocol featuring system resource reservation is an approach that is the key to distributed multimedia services.

The Q_oS of a continuous media stream can be described in terms of spatial and temporal parameters [30]. These parameters include : colour depth, resolution, frame rate, audio sampling rate, etc. A continuous media delivery has to guarantee a sustained level of quality with fixed spatial and temporal parameters. This requires a sustained network bandwidth for the duration of the continuous media delivery. The large volumes of data characteristic of multimedia information causes network congestion problems when considering the growing number of users on typical distributed systems. A continuous media delivery, for example, with the following Q_oS levels : 24 bit colour, 640x480 resolution, and 30 frames per second for movie quality playback, must guarantee at least 221 Mbps throughput for the visual stream alone (without video compression). Video compression schemes allow for a greater number of concurrent users by reducing the amount of data needed to be delivered for distributed multimedia applications. The problem of network congestion is addressed further by a dynamic quality of service (dynamic Q_oS) model for continuous media services. Dynamic control of temporal and spatial Q_oS parameters results in a dynamic change in the throughput of multimedia delivery and is the basis of the dynamic Q_oS model. The advantage of dynamic Q_oS is obvious - less traffic means more connections. And the more connections a server can provide concurrently, the more it will succeed as a multimedia on-demand server. For all intents and purposes, dynamic Q_oS and dynamic throughput refer to the same approach of bandwidth reduction, and these two terms be used interchangeably.

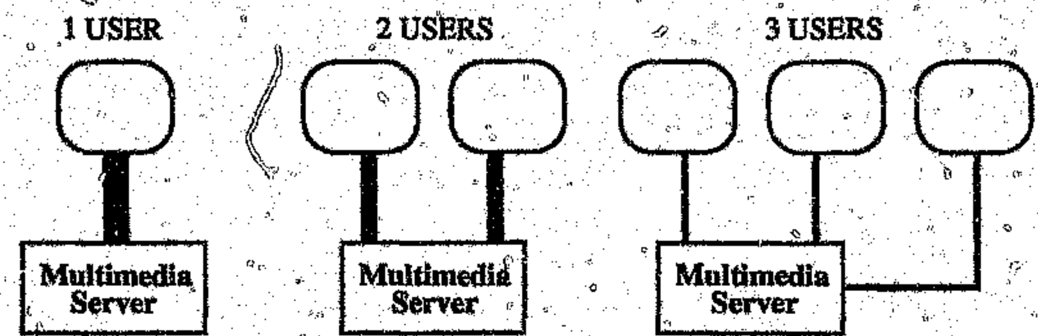


Figure 1 - A conceptual view of dynamic Q_oS

By degrading the Q_oS of existing multimedia sessions it is possible to free system resources so as to accommodate new user requests. The Q_oS of existing sessions are restored when these resources become available again. This scenario involves reducing the quality of existing continuous media sessions, i.e. dynamic throughput, in order to accommodate a greater number of concurrent users. Figure 1 illustrates the conceptual idea behind dynamic Q_oS. The lines joining the multimedia server to the user's workstations are used to describe the network bandwidth usage, or throughput, for the situations of 1 user, 2 concurrent users, and 3 concurrent users, respectively. This figure illustrates that by decreasing the network bandwidth of individual multimedia sessions, it is possible to provide service to a greater number of concurrent users. Degradation of the physical characteristics of a real-time service obviously results in perceived differences in video acceptability. However, the relationship between video message importance and video degradation is not that obvious.

This report explores the effects of the dynamic Q_oS model by analyzing the "watchability" of different motion picture scenarios for different frame rates. Watchability is used as a composite term to embrace aspects of user acceptance of a video message sequence, such as clarity and interpretive acceptability of audio signals, continuity of visual messages, lip synchronization during speech and the general relationship between visual and auditory message components.

A video classification schema (VCS) was developed which categorizes video information based on three dimensions : the Temporal nature of the data, the Auditory message content, and the Visual message content. Temporality describes the static/dynamic nature of a motion picture. Sports footage would be categorized as "high temporal data" due to rapid scene changes. Talk shows would be categorized as "low temporal data" by the static nature of the talking heads. Auditory and visual message content describes the importance of sound and image for different video information based on some elementary principles from psychology [18]. It is this report's argument that different video information (i.e. different classifications in the VCS) implies different acceptance or user watchability with respect to the dynamic Q_oS model.

High temporal data is more demanding on system resources than low temporal data. This is a consequence of the compression ratios that are obtainable for these two classes. Static data (low temporal) is more suited for video compression methods that exploit temporal data redundancy. One such video compression method is MPEG [2,10,15] which uses motion estimation in its frame-differencing approach. This was quantified by compressing numerous video clips using Intel's Indeo RTV 2.1 compression scheme under the Video For Windows environment. It was found that high temporal data consumes significantly more storage capacity than low temporal data after undergoing compression. This exercise illustrates that high temporal data requires significantly more network bandwidth than low temporal data. The relationship between temporality, compression ratio and network bandwidth requirements is described as follows :

$$\text{Temporality} \propto \frac{1}{\text{Compression Ratio}} \propto \text{Network Bandwidth Required}$$

Furthermore, it was also found that the amount of resources required to store visual data is far more significant than its auditory counterpart. This is the reason that we are interested in the compression differences associated with low and high temporal data. It is assumed, for all intents and purposes, that auditory information maintains a constant bandwidth in the VCS. Good quality audio is said to require only about 10% of the bandwidth than its corresponding visual component, and can be included with full motion video with little, if any, additional cost [13].

In the dynamic Q_oS model it is possible to alter the frame rate of a continuous media delivery which has direct consequences for network throughput (traffic). It is obvious that by dropping the frame rate there will be a loss in quality. But technical parameters are not the only determining factors in establishing the watchability of video information in this distributed multimedia model. Rather, more subjective and interpretive factors inherent in the user's perceptual judgments will play a key role in user acceptance.

An experiment was designed to analyze user watchability of different video classifications for different frame rates. This was done by asking a sample of respondents to make structured ratings of test video material compiled in accordance with the VCS. Ratings were based on clarity and interpretability of video content. By simulating a dynamic Q_oS environment using a stand-alone PC, DVI hardware [6,11,13,24] and a VCR, it was possible to gain insight into the practicality of dynamic Q_oS. The analysis of the simulation's results provides insight about the network bandwidth that will be required for particular distributed video applications.

This report describes the effects that changing the frame rate in a dynamic Q_oS system has on user perception. The structure of the report is as follows : The next section, section 2, describes multimedia and the related algorithms and technologies that are characteristic of a continuous media delivery. Section 3 presents a brief description of a multimedia service in the context of dynamic Q_oS. Section 4 presents the video classification schema. Section 5 outlines the details of the experiment. Section 6 presents the outcome of the statistical analysis. Section 7 provides an interpretation of these findings. Section 8 discusses the significance of these results in the context of dynamic Q_oS. Finally, section 9 concludes.

2. MULTIMEDIA, ALGORITHMS AND TECHNOLOGIES

Multimedia

There are as many definitions of "multimedia" as there are multimedia experts. The term "multimedia" conveys different meanings to different people [12]. For all intents and purposes, multimedia implies an integration of audio, video, graphics, animation and text. Multimedia simply means two or more media. It can be any of these in simultaneous combination.

Is important to distinguish between multimedia and *multiple* media. Even with two or more we don't necessarily have what is known as multimedia, per se, but rather something better called multiple media. This form of communication can be considered a parallel mix of various output formats. Media that is physically collocated does not represent multimedia. But integrated media (synchronized) is what we are aiming for. A good working definition [31] would thus be :

MULTIMEDIA = VARIETY + INTEGRATION

A multimedia system should support a variety of media types. This could be as modest as graphics and text or as rich as animation, audio and video. It is emphasized that this alone is not sufficient for a multimedia environment. The variety of media types needs to be integrated into a single system framework. Multimedia presents immense opportunities for computing applications. The introduction of multimedia can enhance the utility of many existing systems. There are also many new areas of application which are made possible by the emergence of multimedia. In particular, distributed multimedia applications cover diverse areas [31] such as :

• office automation
• service industry applications
• retail applications
• domestic applications
• science and engineering
• cultural activities.

There will be an explosion of multimedia applications over the next few years. Many computer companies are betting their futures on multimedia [25]. However, nearly all useful multimedia applications will benefit greatly from networking. It is emphasized that this report concerns multimedia in a distributed context.

Multimedia implies random access. People generally absorb information in a linear manner. You cannot extract the contents of a book without reading it from start to finish. Reading implies visual information transfer. Audio is another form of information transfer. It too is linear. You can't listen to an orchestra recording without listening to an extended series of notes. Multimedia, on the other hand, implies a non-linear approach to information transfer. Interaction requires a random access of information. Digital video technology is paving the way for a new television experience. The act of viewing motion pictures is not likely to change (yet), but the way we interact with the delivery of video information will gain greater sophistication [29]. Greater user control is fueling a need to provide more intelligent services to the famous idiot box. An important design goal of distributed multimedia applications is flexible user interfaces [27]. Access for the handicapped, and multiple language support are important issues. A winning application architecture should support user interface adaptability.

Multimedia is not restricted to the computer screen. It can be viewed on television. Television does offer interaction - the volume can be changed, the set can be switched on and off, an so on. But the degree of interaction is limited due to the nature of the physical device. A computer application providing multimedia information supports a level of interaction that is based on the complexity of the application itself. The degree of interaction is shifted from the computer manufacturer to the software developer. But the multimedia concept is generally associated with the computer, and not the television. Sitting down in front of your computer to watch the news sounds as incongruous as using a multimedia television application.

Multimedia encompasses the world of motion picture. When you go to the cinema you can justifiably state that you are viewing multimedia information. But the word "multimedia" is generally associated with computer systems. Staying with this convention, this report explores multimedia with respect to computers and digital communication mechanisms.

The act of viewing motion picture cannot be considered in isolation. The slogan "it is always better on the big screen" is not merely a marketing strategy. The television offers a radical departure from the sound and image relation that is typical of the cinema. Television images are of a lower quality than cinematic images in terms of spatial information. This lack of detail is apparent, and together with the fact that television shows things smaller than they are, it has profound effects on the audience's attitude [5].

Television audience concentration is not encouraged to the same degree as that for the cinema. The television audience is not subject to darkness, strangers, large images [5]. Television is often not the principle event. It is treated casually rather than concentratedly. Although its viewer's sustained degree of concentration is much lower than that for cinema, it has greater extended watching period and more frequent use than cinema.

Sound plays a greater role for television than it does for the cinema. The context of motion picture therefore has to be put into perspective. The television has been said to engage the glance rather than the gaze [5], and thus has a different relation to voyeurism from cinema's. Television sound presents a strange paradox - although it is an integral part of the motion picture broadcast, television manufactures provide speakers of poor quality [5], even though the broadcast signal often has a wide tonal range. Television sound is massively geared towards the acceptable reproduction of speech. The sound radiates in all directions, whereas the view of the television image is sometimes restricted. Sound can be heard when the screen cannot be seen. The role played by sound stems from the fact that it is used to ensure a certain level of attention, to lure viewers back to looking at the television set. Sound holds attention more consistently than image [5]. It provides a continuity that sustains across momentary lapses of attention.

The above discussion describes a slightly different balance between sound and image from that which is characteristic of cinema [5]. But multimedia, on the other hand, is associated with the computer screen - not a television. Although both are cathode ray tubes, the computer screen, unlike the television, is not the subject of casual motion picture delivery. The computer screen is characterized by a single user sitting very close to it. The quality of the computer screen is much higher than that of the television. Crisp and brilliant images can be displayed. Is it being suggested that the computer screen is similar to the cinema? The answer is no!

The advent of friendly graphic user interfaces has resulted in an explosion of easy to use software that runs in virtual screens or windows. A desktop computer is characterized by multiple applications running simultaneously in separate (sometimes overlapping) windows that are typically much less than the full screen size. So a television application will not only run in a window about less than, say, a third of the screen size, but it will be one of many simultaneous activities that the user is involved in. Thus computer sound will have to play a similar role to that of television sound, in holding the user's attention. Viewing a motion picture in a window will ultimately be one of many simultaneous activities.

It is emphasized that the delivery of multimedia information is characterized in the computer world by digital communications. Addressing quantitative improvements such as doubling the line and pixel rate of existing analog standards, like PAL or NTSC television, would seem like the most obvious reason for the move, but "going digital" strives to journey beyond quantitative enhancements, providing greater functionality inherently limited by existing analog standards [16]. Multimedia implies digital data - data that can be archived and accessed by the use of tools which are efficient and easy to use (similar to the way that information is stored and retrieved from a data base).

In the computer world, multimedia represents a combination of digital information, delivered by digital means. Analog information delivery follows a linear schedule. There is no way to search through a video database other than by viewing a collection of crudely labeled video tapes. Annotation of analog video material is difficult. It is not possible to "write in the margins" of a video tape so that comments on a particular video segment can be passed on to others [16]. Editing video tape at the consumer level is not a popular activity. Digital representation of information is coupled with tools that allow the editing process to be significantly simpler. Think of how easy it is to insert a word into a piece of text using a word processor. An analog comparison includes ink and correction fluid - an obvious limitation.

Finding specific information is a tedious and time consuming event in the analog world. Multimedia's success stems from the potential offered in digital storage and delivery of information. In the future, it will be able to locate specific multimedia information by the symbolic recognition of discrete objects in different media. Furthermore, digital representation of continuous media implies [9,16,19] :

- random access
- compression of information
- encryption of information
- uncompromised duplication of data.

Compression

The audio bandwidth for high-quality stereo sound is about 20 kHz which requires approximately 1.4 Mbps to be carried over a digital network. The bandwidth requirement for broadcast-quality video, on the other hand, is about 10 Mbps, and more than 100 Mbps for HDTV [2]. Because good quality audio only requires about 10% or so of video' bandwidth, and storage and processing requirements, we are only going to look at compression in the visual domain. The additional cost for including audio in a real-time multimedia delivery can be achieved with little, if any, additional cost [13].

Compression in the spatial domain pertains to still images and is known as inter-frame compression, while compression in the temporal domain pertains to a sequence of images (that exist in time) and is known as inter-frame compression.

Compression algorithms take advantage of both spatial and temporal characteristics of an image sequence by exploiting the redundancies between individual frames in the sequence [6], and the non-linearities of human vision [2]. The eye is far more receptive to brightness variation than it is to changes in colour information. Luminance signals are consequently sampled with a higher resolution. Furthermore, the eye is less sensitive to images with high spatial frequency than with low spatial frequency. An image formed by an alternating spatial signal of black and white, for example, would be seen by the human viewer as a uniform gray instead of the alternating checkerboard pattern (the technique used in dithering). Exploiting this deficiency is achieved by coding the high frequency coefficients with fewer bits and the low ones with more bits.

Two standards exist for still and sequential image compression. These are the Joint Photographic Experts Group (JPEG) for still image compression, and the Moving Picture Experts Group (MPEG) for full-motion compression of continuous media [2,10,15]. JPEG image compression is transform based. Each of an image's RGB components are divided into blocks of 8 by 8 pixels and then transformed using the two-dimensional discrete cosine transform (DCT). MPEG specifies a data rate of under 1.5 MB/s by exploiting inter-frame redundancies between image sequences (frame difference compression). MPEG actually addresses audio compression and visual and auditory synchronization. MPEG-Audio specifies auditory compression of a digital audio signal at rates of 64, 128 and 192 Kbps. MPEG-System addresses the issue of media synchronization. MPEG-Video the full name for the component that deals with motion picture compression (MPEG and MPEG-Video are generally used in literature interchangeably).

Still image compression (intra-frame) is conceptually more complicated than sequential image compression (inter-frame). Still images use spatial image compression to compress the information within a single picture frame. Sequential images require temporal compression to compress information between a sequence of picture frames. Still image compression algorithms are characterized by the following 3 criteria :

(1) *Compression ratio*

$$\text{compression ratio} = \frac{\text{size of original image}}{\text{size of compressed image}}$$

This ratio gives an indication of how much compression is achieved for a particular image. Most algorithms have a typical range of compression ratios that they can achieve over a variety of images. Because of this, it is usually more useful to look at an average compression ratio for a particular method, Generally, the higher the compression ratio, the poorer the quality of the resulting image. The tradeoff between compression ratio and picture quality is an important one to consider when compressing images. Furthermore, some compression schemes produce compression ratios that are highly dependent on the image content. For example, a highly detailed image of a crowd at a football game may produce a very small ratio, whereas an image of a pure blue sky may produce a very high compression ratio.

(2) *Image quality*

Compression schemes can be classified as being lossy or lossless. Lossless schemes preserve the original data. Lossy schemes do not preserve the original data – picture information is lost, and it cannot be recovered after compression. Lossy schemes attempt to remove picture information the viewer is apt not to notice. As more and more picture information is removed, the picture quality decreases.

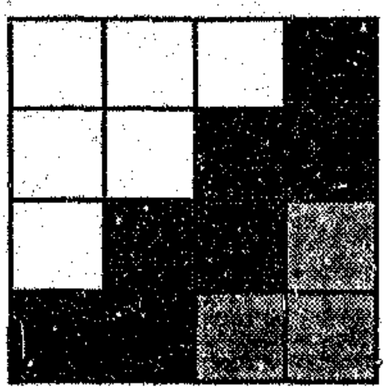
(3) *Compression/Decompression speed*

This is defined as the amount of time required to encode and decode a picture, respectively. This time depends on :

– the complexity of the compression algorithm
– the efficiency of the implementation of the algorithm
– the speed of the utilized hardware

It is obvious that the faster compression/decompression operations can be performed, the better. Fast compression increases the speed with which material can be created. Fast decompression time increases the speed with which the user can display and interact with images.

There is are many different approaches to spatial image compression, DCT-based (discrete cosine transform) methods are the most common, but other methods do exist. Fractal-based compression [10,32], for example, is said to achieve very high compression ratios for natural scenes where the underlying structure is very conducive to this approach. Temporal compression, on the other hand, is generally implemented along the lines of frame difference compression, i.e. the MPEG approach of exploiting inter-frame redundancies. This approach requires a key frame, or reference frame, to be used as the reference point for a specific number of sequential frames that follow directly afterwards. Note that the reference frame itself requires still image compression.



Analog to Digital →

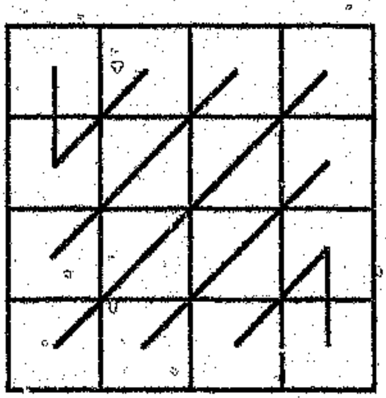
0	0	0	8
0	0	8	8
0	8	8	4
8	8	4	4

Figure 2 - Digital representation of an analog image

A still image is first digitized to form a rectangular matrix of discrete positive integers. These integers represent the colours in the original image. This is classically known as digitizing an (analog) image, and is illustrated in Figure 2. This figure represents a simplistic image as a 4 by 4 matrix of gray levels, and will be used to illustrate the principle of DCT-based image compression (JPEG).

$$\begin{bmatrix} 0 & 0 & 0 & 8 \\ 0 & 0 & 8 & 8 \\ 0 & 8 & 8 & 4 \\ 8 & 8 & 4 & 4 \end{bmatrix} \xrightarrow{\text{DCT}} \begin{bmatrix} 68 & -23 & -4 & -3 \\ -23 & -28 & 17 & 3 \\ -4 & 17 & 20 & -11 \\ -3 & 3 & -11 & -12 \end{bmatrix}$$

Figure 3 - Discrete cosine transformation

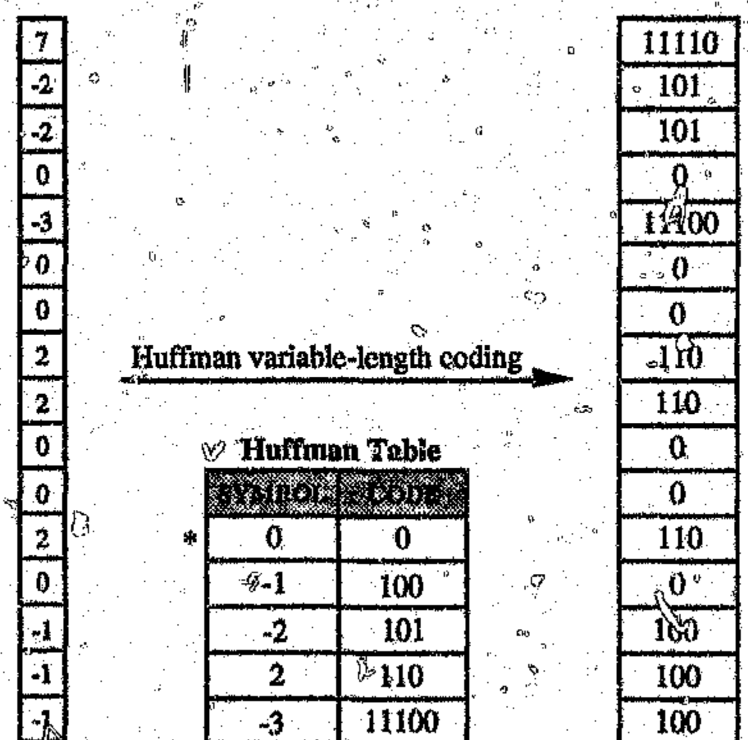


$$\xrightarrow{\text{Zig-Zag Scan}} \begin{bmatrix} 68 \\ -23 \\ -23 \\ -4 \\ -28 \\ -4 \\ -3 \\ 17 \\ 17 \\ -3 \\ 3 \\ 20 \\ 3 \\ -11 \\ -11 \\ -12 \end{bmatrix} \xrightarrow{\text{Quantization}} \begin{bmatrix} 7 \\ -2 \\ -2 \\ 0 \\ -3 \\ 0 \\ 0 \\ 2 \\ 2 \\ 0 \\ 0 \\ 2 \\ 0 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

Figure 4 - Zigzag scan, linear representation and quantization

Figure 3 shows how the DCT maps an image from the spatial domain to the frequency domain. It is a misconception that the DCT itself results in compression of data needed to represent an image. The DCT simply exploits the fact that the human eye is less sensitive to high-frequency components, and thus these data can be eliminated which will then result in data compression.

In Figure 4, the transformed data is scanned in zigzag fashion resulting in a column vector which represents the original matrix in linear form. The column vector then undergoes quantization. In this scenario, quantization simply involves dividing by 10. High-frequency coefficients subjectively matter less than log-frequency ones and may thus be quantized more coarsely [2]. The zigzag scan results in the quantized coefficients being approximately arranged in order of ascending frequency. Note that division by 10 results in lossy compression since the divisor is rounded off.



Huffman Table

	SYMBOL	CODE
*	0	0
	-1	100
	-2	101
	2	110
	-3	11100
	1	11101
†	7	11110

* = most frequently used
† = least frequently used

Figure 5 - Huffman coding

The quantized column vector is then encoded using Huffman style coding. This approach is a well known means of reducing the number of bits needed to represent a data set without losing any information. Notice from Figure 5 how the most frequently used data get encoded with the least number of bits. The resulting data compression should become clearly apparent. Figure 6 further illustrates the principle of Huffman encoding. Tree traversals are greater for data coded with more bits. The most frequently used data are coded with the least number of bits and therefore get decoded with the least number of traversals.

It is emphasized that the DCT process does not result in data compression. The DCT does however represent data in a different domain which is more conducive to compression based on techniques such as Huffman encoding. To reconstruct an image from its reduced data representation the reverse process is followed whereby the Huffman encoded data is firstly decoded into an integer column vector. This vector is then inversely quantized (for our example this would imply multiplication by 10). Note that the resulting data is lossy due to the initial quantization which rounded off the original data. This vector is then reassembled into a matrix after an inverse zigzag scan is applied. Finally, an inverse DCT operation yields the original, albeit lossy, image.

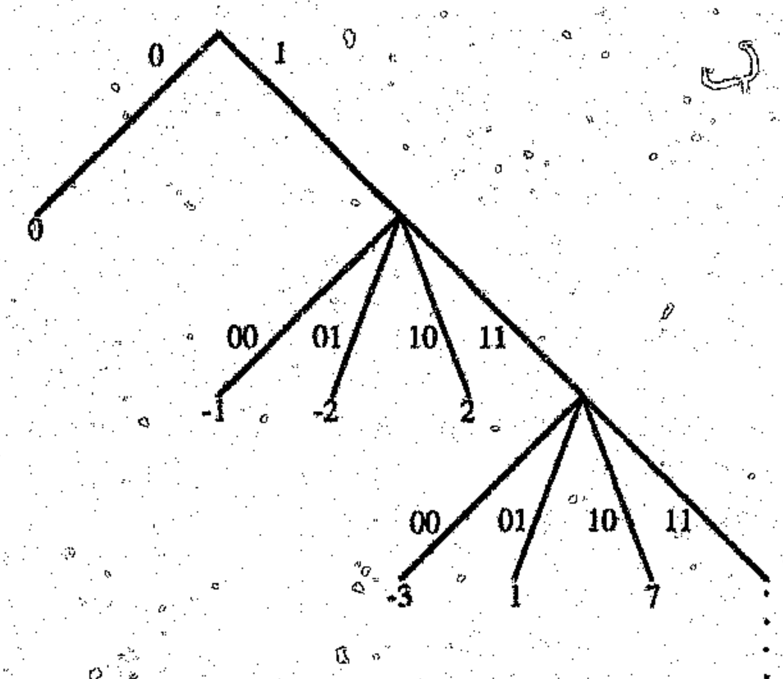


Figure 6 - Coding the most frequently used data with the least number of bits

The above example was simplified for illustration purposes. Typically an image is divided into 8 by 8 non-overlapping blocks of pixels and the above process is applied to each block. This is the basis for the JPEG approach to still image compression. It is also pertinent to MPEG compression because each reference or key frame requires spatial compression. Where MPEG compression takes a radical departure from JPEG is in its exploitation of inter-frame redundancies. Remember that JPEG is only used for spatial or still image compression.

MPEG specifies a reference frame period, i.e. a frame which serves as a reference point for a sequence of frames that follow it directly. The experimental work performed in this report used a frame difference compression scheme (similar to MPEG) that used a reference frame period of 1 in 20. Each frame between 2 reference frames can be referred to as a motion vector. Each of these vectors represents the spatial difference between that frame and the reference frame. By coding only the frame differences, i.e. motion vectors, in an image sequence, it is clear that a high compression of data can result. However, this approach is highly dependent on the nature of an image sequence (as discussed throughout this report). By compressing each motion vector again with the JPEG scheme a further data compression results. This is a very simple overview of motion picture compression. In practice, MPEG may appear slightly more complicated, although its conceptual definition is very straight forward.

Synchronization of Media Streams

Audio and video are perceived as changing continuously over time, and are therefore known as continuous media. These media are characterized by large volumes of data and are typical of mass-market information services such as television, stereo hi-fi and telephone. But continuous media and computer systems do not coexist easily [1]. Processors, buses and networks may not always provide the real-time performance that is required to handle time-based or continuous media. In the future, computer systems will include a facility for I/O of continuous media delivery to and from a user's workstation. However, since multimedia consists of different media in simultaneous combination, explicit synchronization mechanisms are necessary so that multiple media streams can proceed "tightly".

Considering that computer applications can be interactive, synchronization is necessary nearby the sink of the multimedia streams [28]. A user could possibly speed up, slow down, or repeat a multimedia delivery, analogous to the way one controls a VCR. These interactions provide further reasons for the inclusion of explicit synchronization mechanisms. This scenario relies on the independence of different media and introduces greater flexibility with respect to user requirements.

It should be possible, for example, to play an audio fragment without having to access the corresponding images.

Conceptually, synchronization can be described by using the commonly known semaphore with states such as busy and waiting. This simplistic view can then be extended to form a global synchronization mechanism in a distributed multimedia environment. There are many technical reasons why explicit synchronization is needed [1], and include :

- network delays may cause the delivery of related media streams to skew
- motion picture frames are often buffered because of the highly pipelined nature of video decompression hardware, and audio samples, on the other hand, can be played as soon as they are received.
- scarcity of system resources may cause an I/O stream to "starve" and skewing will worsen if I/O continues on one stream and not another

Continuous media streams can be viewed as a sequence of units that must be coordinated in time on a global time system within the distributed system. Each unit has an associated non-negative time stamp. This gives it its *temporal* (inter-frame) position in the stream. An LTS (logical time system) [1] is constructed whereby I/O on each logical device is synchronized to the LTS. The timing resolution is determined by the frequency of the device interrupt. Each new multimedia delivery invokes the creation of an LTS for that session. The initial value of the LTS is set to zero and continues to increase during data transmission. If one of the devices in the LTS becomes blocked, because of network congestion for example, then the skew between the LTS and the device will increase. If a specified upper-bound on the skew factor is reached, then the LTS stops advancing while the device catches up.

Networks Types & Bit Regimes

It has been envisaged that a domestic subscriber will select an HDTV (High Definition Television) channel that will be reachable through an ATM (Asynchronous Transfer Mode) network [3]. ATM-based networks can support both variable and constant bit rate services at very high speeds [10]. Sustained bit rates is a primary objective of high speed network developers since jitter causes delays between packets which implies the need for media synchronization mechanism in the context of distributed multimedia applications. Although a Gigabit network is essentially a billion bits per second, ATM networks achieving bit rates of 600 Mbps are categorized as GigaNets. These networks are currently under development and will be very expensive in their initial employment. Nonetheless these bandwidths are necessary for the realization of multimedia libraries in distributed systems.

A GigaNet is more than just a high-speed pipe; it must provide connectivity to an unlimited number of users with low latencies and acceptable performance. The problem is whether traditional protocols such as TCP/IP can succeed in delivering the performance that is required for distributed multimedia services. The commonly available network protocols generally offer an "all or nothing approach" [20] to distributed services. Distributed multimedia applications need to negotiate their bandwidth requirements with the underlying service network. TCP, for example, which exists on top of IP, provides higher reliability of transmission than UDP [7]. But the protocol and the intended application cannot be considered in isolation. Perhaps UDP, with is less reliable and hence has less overheads, is more conducive to motion picture delivery. Surely the occasional lost packet is less severe than the retransmission delays in the context of digital video communications?

Obviously brute strength will provide greater bandwidths as faster and faster networks continue to be developed. But just as highways become widened, so does the number of cars on the road continually increase. And widening highways is not a cheap solution, and takes time to construct. Likewise, the development of faster networks requires time and resources. Network protocols need to be revised in light of the growing user community and in anticipation of a new generation of distributed applications. This report describes the Capacity Based Session Reservation Protocol as a mechanism for performance guarantees with respect to a distributed multimedia service (see section 3).

Fiber optics is the foundation for Gigabit network, but shielded twisted pair wiring achieves very high speeds for short distances [3]. In the future, transmission speeds of up to 2.4 Gbps will be obtainable over long distances based on development in SONET (Synchronous Optical NETwork). It is important to note that when we speak of public network speeds we are in fact referring to the data rates of switched data services, non-switched private leased lines, or internal networking. Given that the data rates for US public networks have increased by almost 5000% in the last 20 years [23], Gigabit networking is immanent. In the early 1970s, data communications were limited to a speed of 1200 bps. Then we saw the introduction of ISDN (Integrated Services Digital Network) with bandwidth capacity of 64 Kbps, and now T-1 carries with a speed of 1.54 Mbps are quite common in modern switched circuits. Today leased lines capable of 45 Mbps are widely available in France, Japan, Britain and the USA. Public networks are said to increase by roughly an order of magnitude every seven to eight years [23]. Simple analysis suggests that switched circuits will cross the gigabit threshold by the year 2005. The demand for high-performance distributed applications will certainly pre-empt the emergence of such powerful infrastructures. Applications that will require high speed networking include HDTV, remote supercomputer visualizations, and virtual reality. However, the majority of distributed services are text based. These applications are in the kilobit range and will probably still be the dominant application used on the GigaNets of the future in their initial deployment.

Table 1 presents some of the most common networks and their associated bit rates [16,17].

Network	Bit Rate
POTS (Plain Old Telephone Service)	0.3-56 Kbps
DS-0 (fundamental bandwidth of US telephone co.)	56 Kbps
ISDN (Integrated Services Digital Network)	64-144 Kbps
LocalTalk (personal computer LAN)	230 Kbps
T-1 (multiple of DS-0)	1.5 Mbps
Ethernet (packet-based LAN)	10 Mbps
T-3 (multiple of DS-0)	45 Mbps
FDDI (Fiber-optic Data Distributed Interface)	100-200 Mbps
SONET (Synchronous Optical Network)	N*51.84 Mbps

Table 1 - Various networks and their associated bit rates

DVI Hardware & The i750 Processor

In the past 10 years the personal computer has witnessed an explosive increase in the information bandwidth it is capable of handling. Human-computer interfaces have shifted from text and data to include audio, video, photo-realistic images and animation. Personal computers have moved from isolated environments to multi-megabit per second local and wide area networks.

A distributed multimedia application is driven by the network bandwidth that is available on the system. DVI technology is a combination of hardware and software approaches to the problem of integrating continuous media into a distributed computing environment supporting real-time applications.

At the heart of DVI technology is the i750 chip set developed by Intel. The i750PB pixel processor and the i750DB display processor handle real-time compression and decompression of images, audio and video.

GUIs have only been modestly improved upon since their introduction in 1984 with the Apple Macintosh. Even with advances in VLSI (Very Large Scale Integration) technology, we are still limited to 8-bit graphic cards, and even more interesting is that virtually all personal computers still rely on the host microprocessor to create and format the display [13]. The i750 video processor is intended to integrate all processing functions required for multimedia in a single programmable chip set. Separate from the host processor, the i750PB is responsible for most of the data processing that pertains to video compression and decompression. The i750DB processor performs the real-time display functions by generating all the timing signals required to drive display devices.

DVI technology performs its task using the YUV colour schema. A series of 8-bit samples represents each of the YUV components. In YUV, the luminance, or brightness information (Y), is separated from the chrominance, or colour information (U and V). Working in the YUV domain allows various psycho-visual effects to be exploited. The human visual system does not require high frequency data in the colour domain [6]. Colour information does not require the same resolution as the luminance information. Without a significant loss in image quality, the YUV scheme allows each of colour dimensions be stored at one-half or one-quarter of the luminance component. Luma and chroma information are thus placed in different bitmaps because of their difference in resolutions. Chroma information is interpolated up, to the resolution of the luminance data.

DVI technology is assembled and shipped in what is known as the ActionMedia II card for real-time video capture and playback.

3. DISTRIBUTED MULTIMEDIA AND DYNAMIC Q_oS

A continuous media server has to deliver audio and visual data in a timely fashion. To guarantee the continuity of motion picture delivery, the distributed system has to secure the necessary computational resources (i.e., processor and memory) and network resources. A high-speed network is the obvious foundation, but traditional protocols are unsuitable sustaining a continuous media connection.

Protocols can be divided into two classes : reservation- and non-reservation-based protocols. The Capacity Based Session Reservation Protocol (CBSRP), developed at Carnegie Mellon [30], is an example of a reservation-based protocol. The CBSRP guarantees the establishment of a real-time channel with bounded end-to-end delays; it is well suited to the delivery of continuous media. Furthermore, this protocol supports dynamic Q_oS, or dynamic throughput of continuous media in a distributed multimedia environment. It has the capability of changing the throughput, i.e. the Q_oS parameters, of a continuous media connection dynamically. The CBSRP determines whether a continuous media session can be created which meets a user's requested Q_oS parameters. If a minimum quality level can be accommodated then the session is established. In this distributed multimedia environment, a system of priorities is used, whereby the throughput or Q_oS of a session is guaranteed unless a shortage of resources occurs by the introduction of more urgent system tasks.

The CBSRP provides real-time sessions by reserving buffers, processor, and network bandwidth. The reservation of resources is essential for bounded end-to-end delays that are necessary for continuous media services in a distributed system. A multimedia session is established between a sender (multimedia server) and a receiver (user's workstation) if a load is schedulable under a specific deadline. This is determined by the components of the dynamic Q_oS system incorporating the CBSRP.

The components of a dynamic Q_oS system are described in Figure 2. When a user requests a multimedia service, a session manager on the user's workstation reserves the required local resources and forwards the request to a remote session manager on the multimedia server. A session is only successfully established if the required resources can be reserved at both the local and the remote machines. A user requests a multimedia session by specifying a set of Q_oS parameters. A range of temporal and spatial parameters determines the different Q_oS levels that can be used during multimedia delivery. It is emphasized that dynamic control of these parameters is the basis for dynamic Q_oS - it allows a user to create Q_oS levels as a discrete function of the resource load. If a session can be created which meets the user's minimal request Q_oS parameters, then the session is established.

A local session manager handles the creation, termination and Q_oS reconfiguration requests from a user and communicates these with the remote session manager. A network resource manager handles network admission control and resource management. Each machine on the system also has a system resource manager which communicates with the session manager to determine if the local resources available are sufficient for session creation with the minimum Q_oS parameters. This manager provides buffer checks and reports the availability of processor resources to the session manager upon request.

In the Carnegie Mellon [30] implementation of the CBSRP, the session manager, system resource manager, and the network resource manager are all kernel objects in their Advanced Real-Time System (ARTS) distributed operating [illegible], also being developed at this university. Figure 2 also shows the way the various managers in[illegible] in a dynamic Q_oS system.

In Figure 2, the receiver (user) requests a multimedia session to be created. As soon as the local resources are checked and reserved, the receiver's session manager sends a message to the server's session manager requesting a session to be established. The Q_oS parameters are also communicated in the negotiations between local and remote session managers. If both the session managers can service the requests then the session is created. Renegotiation takes place if more user requests are initiated by the other users on the system. If the server can accommodate these users by degrading the Q_0S of existing sessions then these sessions are degraded by session reconfiguration messages between the local and remote session managers.

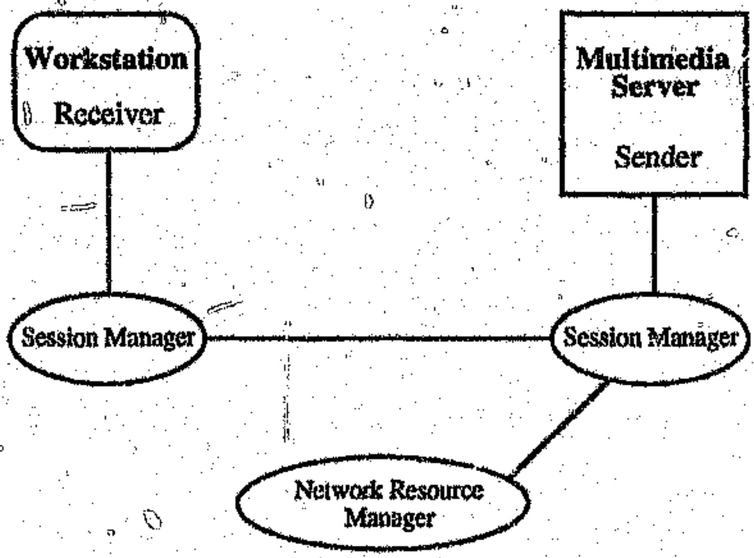


Figure 2 - The components of a dynamic Q_0S system

The network resource manager, illustrated in Figure 2, maintains network resources through a network admission control and bandwidth policy. A network reservation request by the server's session manager is admitted based on the type of network, traffic patterns and existing sessions. A new request is only serviced if the network capacity available is sufficient for the creation of a new session without compromising the Q_0S guarantees of existing sessions.

Q_0S parameters are described by temporal (inter-frame) and spatial (intra-frame) requirements. Visually, spatial parameters are characterized by picture resolution and colour depth. The temporal parameter refers to the frame rate. In the auditory domain, the temporal parameter is used to describe the audio sampling rate, and the spatial parameter refers to mono- or stereo- sound. A user's requested values for spatial and temporal parameters can be expressed as a finite number of Q_0S states. These states are used to control the gradual degradation of continuous media services when the system can no longer sustain resources for optimal service. Q_0S drops during a shortage of resources and is restored when they become available again. Contrast this with a fixed Q_0S which implies that a new service request is rejected when the provision of that service would endanger previously given guarantees [14]. A fixed Q_0S environment can be compared to the telephone system where, if a busy signal is received, the user will have to wait and try again. Adaptive policies such as dynamic Q_oS introduce flexibility but can only guarantee quality levels that have been specified by the user by means of a set of Q_oS parameters.

This dynamic control of spatial and temporal parameters is the foundation for dynamic Q_oS and represents dynamic throughput in a distributed multimedia environment. Its implementation is characterized by session and resource management algorithms, network resource management and a network admission control policy. The work done at Carnegie Mellon with the CBSRP only experimented with dynamic Q_oS changes with respect to network bandwidth. It is, however, possible to extend this to processor cycles and system buffers [30].

Another important aspect of distributed multimedia research is the continuous media storage subsystem. It has been suggested that constrained block allocation [22], whereby successive blocks of data are stored on a disk such that the separation time between them does not exceed playback time for a data block, is conducive to a multimedia server design. Each media stream is stored on a multimedia server in blocks. The way these blocks are organized on the server is important for jitter-free retrieval of continuous media information. The random allocation of multimedia data blocks may result in a situation where the seek time to found successive blocks exceeds the playback time of the media block. The term "scattering parameter" [22] is used to describe the separation between blocks. Constrained block allocation arranges media blocks on the multimedia server in such a way that the scattering parameter does not exceed the playback duration of an individual media block. The final area of interest is media synchronization, which would be implemented on the multimedia server, and on the user's workstation. Audio and visual media streams require asynchrony detection and synchronization algorithms (this was discussed in the previous section).

This report does not provide an in-depth discussion of *all* the components of a distributed multimedia environment. It does, however, attempt to evaluate a dynamic Q_oS environment from a user point of view in terms of clarity and watchability. By experimentation with a specific Q_oS parameter, namely, frame rate, insight is provided as to the type of applications most suited to distributed multimedia services for available network resources.

4. VIDEO CLASSIFICATION SCHEMA

The video classification schema (VCS) used in this study categorizes multimedia, or video information, using three dimensions inherent in video messages. They are : the Temporal (*T*) nature of the data, the importance of the Auditory (A) message component to the user in understanding the message, and the importance of the Visual (V) image to the user in understanding the message.

The following figure, Figure 3, illustrates the VCS. Each of T, A and V has two levels, "*lo*" and "*hi*". Thus we have 8 permutations of T, A and V over "*lo*" and "*hi*".

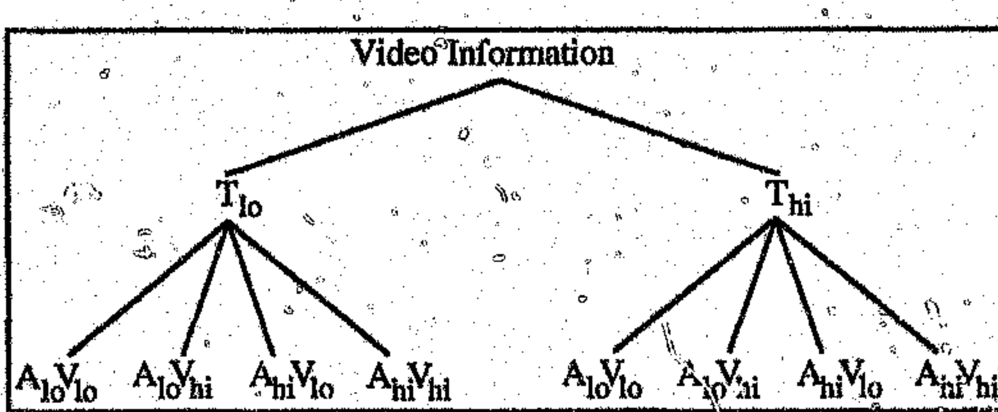


Figure 3 : Video Classification Schema (VCS)

T is the most easily explained aspect of the VCS. T is proportional to the compression ratio that is obtainable for a particular video classification. By compressing different video segments falling in T_{lo} and in T_{hi} we were able to quantify this aspect of the VCS. It was found that video in T_{hi} required significantly more storage space (bytes), on average, than for video in T_{lo}. Thus, the compression ratios for T_{hi} are considerably lower than for T_{lo}. This implies that T_{lo} data is less demanding on a network than T_{hi} data. (It is suspected that even greater resource savings could be measured if the captured video material was from LaserDisc instead of a consumer-grade VCR. These suspicions were justified, in part, by comparing compression ratios of information captured from "test signal" input devices.)

A and V are more difficult to define. It would, of course be possible to use direct physical expression of sound energy and visual matrix, but these options typically make statements only of the transmitted message components, not of the relative importance of these components to the user. Therefore, in the VCS we place emphasis on the structure of a user's subjective appraisals of auditory and visual aspects of video messages. To retain the ecological validity [4] of these judgments the present research uses a method of structured ratings of auditory and visual message components as the basis for the VCS. Twenty respondents were each shown eighty short video clips taken from a wide variety of message types, including sport, actuality, comedy, etc. Users were asked to rate each video clip on a seven point Likert [21] scale for, firstly, the contribution made by the auditory message component to the overall message and then the visual and temporal components. Inter-rater reliabilities on a test trial were assessed as being in the range of $r = 0.65$ to $r = 0.91$. We used these average ratings of the video clip samples to establish the most discrepant clips on each dimension; labeling each "*lo*" and "*hi*". These were then assembled in terms of each $\{A_{lo},A_{hi}\}$, $\{V_{lo},V_{hi}\}$ and $\{T_{lo},T_{hi}\}$. Three representatives of each class were taken, arranged across the levels of reduced frame rates (R) : 5, 10 and 15; thus forming 24 interacting categories ($T_2 x A_2 x V_2 x R_3$). *It is important to note that the categorization into* A_{lo}, V_{hi}, *etc.* was made by raters watching video information at the default VCR frame rate. However, respondents in the experimental trials were asked to view video information at reduced frame rates only.

Examples of video information representative of each class in the VCS, as was used in the experiment, are shown in Table I.

T_{lo}		T_{hi}	
$A_{lo}V_{lo}$	Logo / Test Pattern	$A_{lo}V_{lo}$	Station Break
$A_{lo}V_{hi}$	Snooker	$A_{lo}V_{hi}$	Sporting Highlights
$A_{hi}V_{lo}$	Talk Show	$A_{hi}V_{lo}$	Advertisements
$A_{hi}V_{hi}$	Stand-up Comedy	$A_{hi}V_{hi}$	Music Video

Table 1 - Video classification examples

5. METHOD

The aim of this experiment was to access the watchability of different video classifications (in the VCS) in the context of dynamic Q_oS. The results provide insight into the practicality of this distributed multimedia approach and a guide to the network bandwidth that is necessary for specific multimedia applications.

It is this report's argument that there are perceived differences in the quality of a video window, in a multitasking windowed environment, dictated by the message content in different types of video material. By manipulating the independent variables : Temporal characteristic, Auditory message component, Visual Message component and frame Rate, significant differences are observed with respect to the dependent variable, namely, the measure of watchability.

Again, to preserve the ecological validity of video window viewing tasks, subjects were involved in a spell checking primary task while shadowing a video window as a secondary task. The primary task used Microsoft's Word For Windows 2.0 which is typical of a GUI word processing application. A third task, involving an on-line data collection tool, was used to record a subject's rating of a particular video clip.

Subjects were asked to assess the quality of different video information from the 8 classes of the VCS for the 3 different frame rates (though they were unaware of the VCS). They were asked to make global judgments about the quality of each video clip in terms of message and watchability. In order to form the basis for each subject's judgments, a standardized set of notes was read to each volunteer before they began. The experimental independent variables were based on *T*, *A*, *V* and *R*. The experimental dependent variable was a composite measure of watchability.

The subjects were 60 undergraduate volunteers from computer science who were all familiar with GUI environments. An instruction sheet was read to each subject beforehand. This included an explanation of the tasks they would be performing. Three examples of video clips, similar to those used in the experimental trials, were shown beforehand at a rate of 30 frames per second. These were said to be of "normal" quality and were thus rated 7 on the rating scale. They were told to make judgments as to when they felt that the picture quality was becoming less acceptable in terms of audio and visual message quality and clarity. These dimensions were explained to each respondent in order to focus the basis of their rating judgments. They were also told not to confuse the size of the video window with the quality of its contents. Full screen video windows are more representative of broadcast television, in which case, we would be dealing with analog media. Furthermore, the ecological validity of the experiment is described by multitasking environments illustrated by concurrent windowed tasks.

A 486 PC with a DVI ActionMedia II card running Microsoft's Windows 3.1 was used as the experimentation platform. A VHS VCR was plugged directly into the DVI card which was used to digitize video in real-time for display with Microsoft's Video For Windows. A 640x480 256-colour VGA mode was used with a video window sampling at a resolution of 160x120 pixels. Sound was provided by using a Sound Blaster Pro card. Two VGA screens were attached to the VGA card using a VGA splitting device. The equipment used in the experiment is illustrated in Figure 4.

The two screens pictured in Figure 4 were identical. The second screen is deliberately made smaller in this figure to emphasize that it was only used for monitoring purposes. This gave the researcher an identical view of the subject's screen so that the subject's progress could be monitored. The use of two screens adds to the ecological validity (with respect to the user) of a typical GUI environment. Personal computers got their name by the fact that desktop users are characterized by a GUI featuring one mouse, one screen, and of course, one user.

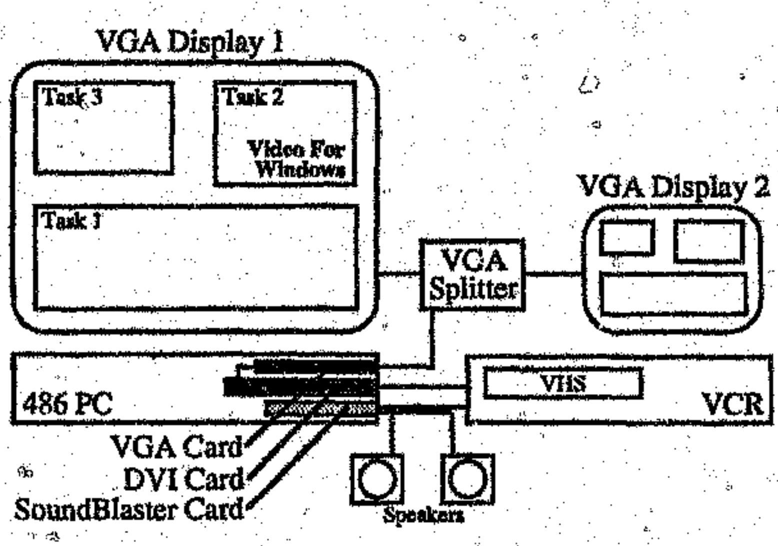


Figure 4 - The components of the experiment

For the primary task, each subject had to spell check a document (deliberately filled with spelling mistakes) at a rate of 10 to 20 words per minute while monitoring the video window. Two monitors were used : one for the subject and one for the researcher to monitor the subject's progress. Each video clip required a randomized frame rate to be entered into Video For Windows without the subject's knowledge of what the frame rate was. Their screen was covered in this process so as to avoid biasing. After setting the frame rate, the VCR was started and the subject began the spell checking task for the duration of the clip.

After each clip, of the 24 clips shown, there was 5 seconds of noise to indicate that the subject must stop the spell checking task and enter a rating of watchability in the rating task. The researcher stopped the VCR and only started it again once the rating was entered. Each randomized frame rate was recorded on paper (by the researcher) for each individual clip. This procedure continued for all 24 clips. A measure was also recorded of the progress that the subject made in the spell checking task. This was used to access if a broadly uniform pace was adhered to for each subject during their participation.

Each subject was exposed to all the test conditions in random order. Each person acted as their own control. Thus a randomized blocks repeated measures design [8] was used. This design examines the variation within each subject.

The data collected for each subject was saved to disk. The recorded frame rates were entered into the computer and saved to disk by the researcher after each subject had completed their final trial, and then, the data was sorted in ascending order from 1 to 24. All 60 data files were then concatenated to produce a 24 x 60 table. This data file was then passed to SAS [26] where a 2-way ANOVA was performed.

6. RESULTS

Data yielded by the experimental method takes the form of a 24 x 60 matrix of rating scores. The 24 independent variable conditions represent exhaustive factorial combinations of *T*, *A*, *V* and *R* factors. Table 2 shows the mean ratings for each condition. Table 3 shows the mean ratings for the levels of each factor over each condition. Table 4 summarizes the results (main effects and interactions) from the repeated measures ANOVA. Table 5 shows the results of planned comparisons between frame rates (levels of *R*).

T_{lo}	5.04	T_{hi}	5.66
A_{lo}	5.55	A_{hi}	5.15
V_{lo}	5.78	V_{hi}	4.92

Table 2 - Mean ratings for each condition

VCS Classification	R_5	R_{10}	R_{15}
$T_{lo}A_{lo}V_{lo}$	6.3	6.2	6.4
$T_{lo}A_{lo}V_{hi}$	3.3	4.8	5.1
$T_{lo}A_{hi}V_{lo}$	3.5	5.2	5.9
$T_{lo}A_{hi}V_{hi}$	3.0	5.1	5.7
$T_{hi}A_{lo}V_{lo}$	5.9	6.3	6.5
$T_{hi}A_{lo}V_{hi}$	4.2	5.6	6.0
$T_{hi}A_{hi}V_{lo}$	4.9	5.9	6.4
$T_{hi}A_{hi}V_{hi}$	4.5	5.7	6.0

Table 3 - Mean ratings for the levels of each factor over each condition

Source	Sum of Squares	d.f.	Mean Square	*F*	Pr > *F*
T	135.06	1	135.06	54.09	0.0001
Error	147.32	59	2.50		

A	60.43	1	60.43	68.40	0.0001
Error	52.11	59	0.88		
V	267.81	1	267.81	178.74	0.0001
Error	88.40	59	1.50		
R	623.11	2	311.56	305.59	0.0001
Error	120.30	118	1.02		
*T*A*	15.01	1	15.01	11.44	0.0013
Error	77.37	59	1.31		
*T*V*	17.12	1	17.12	17.54	0.0001
Error	57.60	59	0.98		
*A*V*	113.91	1	113.91	88.85	0.0001
Error	65.64	59	1.28		
*T*R*	8.15	2	4.08	3.40	0.0368
Error	141.60	118	1.20		
*A*R*	52.19	2	26.10	29.64	0.0001
Error	103.90	118	0.88		
*V*R*	49.13	2	24.56	31.24	0.0001
Error	92.79	118	0.79		
*T*A*V*	24.81	1	24.81	25.28	0.0001
Error	57.90	59	0.98		
*T*A*R*	23.71	2	11.86	11.78	0.0001
Error	119.04	118	1.01		
*T*V*R*	4.03	2	2.01	2.45	0.0905
Error	96.89	118	0.82		
*A*V*R*	27.15	2	13.58	14.51	0.0001
Error	110.43	118	0.94		
*T*A*V*R*	0.46	2	0.23	0.24	0.7883
Error	114.45	118	0.97		

Table 4 - Summary of the results from the repeated measures ANOVA

Source	Sum of Squares	d.f.	Mean Square	*F*	Pr > *F*
R_5 VS R_{10}	5133.75	1	5133.75	356.66	0.0001
Error	849.25	59	14.39		

R_5 VS R_{15}	9225.60	1	9225.60	366.69	0.0001
Error	1484.40	59	25.16		
R_{10} VS R_{15}	595.35	1	595.35	63.44	0.0001
Error	553.65	59	9.38		

Table 5 - Planned comparisons between frame rates (levels of R)

7. DISCUSSION

Table 2 shows the differences in mean ratings for each of the factors T, A and V between the levels "*lo*" and "*hi*". Table 3 shows an interesting set of differences between the respective "*lo*" and "*hi*" levels of the factors T, A and V, all significant at the $p < 0.0001$ level (Table 4). This confirms the basis of the VCS in terms of subjects' disposition to be influenced by each of these factors differentially. Of importance is the opposite effects of T and V on rated watchability, with V_{lo} rated higher in the experimental conditions (all conditions consisting of videos shown at R levels degraded from normal) than V_{hi}, but with T_{hi} rated higher than T_{lo}. This, together with the highly significant TxV interaction ($p < 0.0001$), but the non-significant TxVxR interaction ($p > 0.05$) suggests a differential effect on user acceptance over different levels of R, with T and V working in opposition to viewer perception, which has important implications for dynamic throughput in a distributed multimedia environment.

T_{lo} was found to be less acceptable at lower frame rates for the message content situation of $A_{hi}V_{hi}$. In other words, when a dynamic QoS parameter such as the frame rate R is changed for T_{lo}, the effect is more noticeable than for T_{hi} in the case of $A_{hi}V_{hi}$. This is further evinced by the differential F values and associated p levels apparent in the TxR, AxV, VxR and AxR interactions (Table 4) with much stronger effects brought out by A, V and R combinations than TxR on its own across A and V combinations.

This is contrary to the expectation that T_{lo} data would have less noticeable differences than T_{hi} data because of the static nature of T_{lo}. This finding is interesting since T_{lo} achieves high compression even when sampled at a low frame rate.

Finally, and rather reassuringly, the very strong main effect for R across all conditions ($p < 0.0001$ - Table 4) supports the fundamental premise that frame rate reduction itself leads to progressively lower ratings of user acceptance in terms of overall watchability. Moreover, the significant differences ($p < 0.0001$ - Table 5) between R_5, R_{10} and R_{15} confirm the extent to which watchability becomes eroded with each stepwise decrease in frame rate.

8. FRAMES RATES, DYNAMIC QoS AND BANDWIDTH

Dynamic throughput in a distributed multimedia environment is characterized by the deliberate degradation of continuous media parameters in an effort to accommodate a greater number of concurrent real-time sessions. Dropping the frame rate of a continuous media stream yields desirable effects with respect to available network bandwidth - by decreasing network traffic it is possible to support a greater number of concurrent multimedia sessions.

The results from the experiment provide us with a guideline as to the watchability of video material from the different classes in the VCS for varying levels of frame rate degradation. Table 6 summarizes user perception of different video classifications over degraded levels of *R*. It also includes the default frame rate which by definition corresponds to a quality level, or watchability, of 100%.

	5	10	15	30
$T_{lo}A_{lo}V_{lo}$	90	89	91	100
$T_{lo}A_{lo}V_{hi}$	47	69	73	100
$T_{lo}A_{hi}V_{lo}$	50	74	84	100
$T_{lo}A_{hi}V_{hi}$	43	73	81	100
$T_{hi}A_{lo}V_{lo}$	84	90	93	100
$T_{hi}A_{lo}V_{hi}$	60	80	86	100
$T_{hi}A_{hi}V_{lo}$	70	84	91	100
$T_{hi}A_{hi}V_{hi}$	64	81	86	100

Table 6 - Expected watchability expressed as a percentage for VCS classes VS frame rates

Figure 2 illustrates the effects of frame difference compression (the MPEG [illegible]ach) for video classified as T_{lo} and T_{hi}. These results were obtained by compressing a 30 second sample from each of the 24 video clips used in the experimental trials. Averaging the resulting file sizes yields two distinct compression classes, namely, T_{lo} and T_{hi}. A value of one in every 20 frames was specified as the reference frame period. It is interesting to note that T_{hi} video achieves less impressive compression ratios than T_{lo} when sampled at low frame rates. This can be explained by the fact that T_{hi} is dynamic in nature and by definition this implies that T_{hi} data has more pronounced frame differences when sampled at lower rates. (It is emphasized that more significant differences with respect to compression ratios would be observed if a LaserDisc was used as the input video source instead of a consumer grade VCR.)

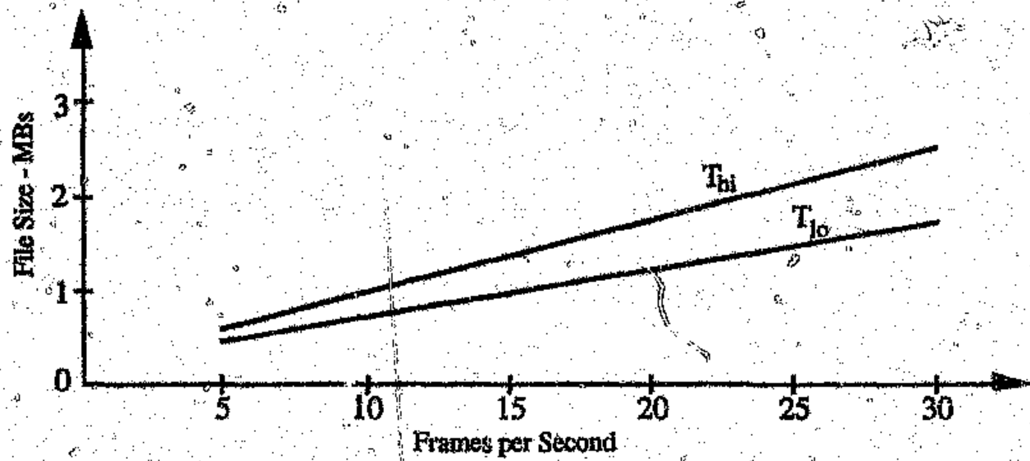


Figure 2 - Video file size VS frame rate for T_{lo} and T_{hi}

Tables 7 and 8 show the bandwidth that is required for video classified as T_{lo} and T_{hi} respectively, over degraded levels of R, and the expected watchability that is associated with each combination of T, A, V and R. The default frame rate is included in each table, which by definition corresponds to a quality level, or watchability, of 100%. These tables are illustrated in Figures 3 and 4, respectively.

	115 Kbps	200 Kbps	260 Kbps	480 Kbps
$T_{lo}A_{lo}V_{lo}$	90	89	91	100
$T_{lo}A_{lo}V_{hi}$	47	69	73	100
$T_{lo}A_{hi}V_{lo}$	50	74	84	100
$T_{lo}A_{hi}V_{hi}$	43	73	81	100

Table 7 - Expected watchability expressed as a percentage for T_{lo} VCS classes VS T_{lo} bandwidths

	165 Kbps	300 Kbps	350 Kbps	670 Kbps
$T_{hi}A_{lo}V_{lo}$	84	90	93	100
$T_{hi}A_{lo}V_{hi}$	60	80	86	100
$T_{hi}A_{hi}V_{lo}$	70	84	91	100
$T_{hi}A_{hi}V_{hi}$	64	81	86	100

Table 8 - Expected watchability expressed as a percentage for T_{hi} VCS classes VS T_{hi} bandwidths

From Figures 3 and 4 it is possible to formulate guidelines as to the network bandwidth that yields a minimal user acceptance level, or watchability, for various video applications in a dynamic QoS environment. For example, a music video application, classified as $T_{hi}A_{hi}V_{hi}$, will achieve more than 80% watchability for network bandwidths of 300 Kbps or more.

Other guidelines can be derived that pertain to the number of concurrent users that can be supported by a multimedia server for a specific level of user acceptance, or watchability, for a particular multimedia application. For example, an FDDI network with a bandwidth of 100 Mbps, can support about 149 users (assuming no protocol overhead) for video classified as $T_{hi}A_{lo}V_{lo}$ with 100% watchability. But, Figure 4 suggests that it can support well over 600 users with about 80% watchability for the same video classification (again, assuming no protocol overhead).

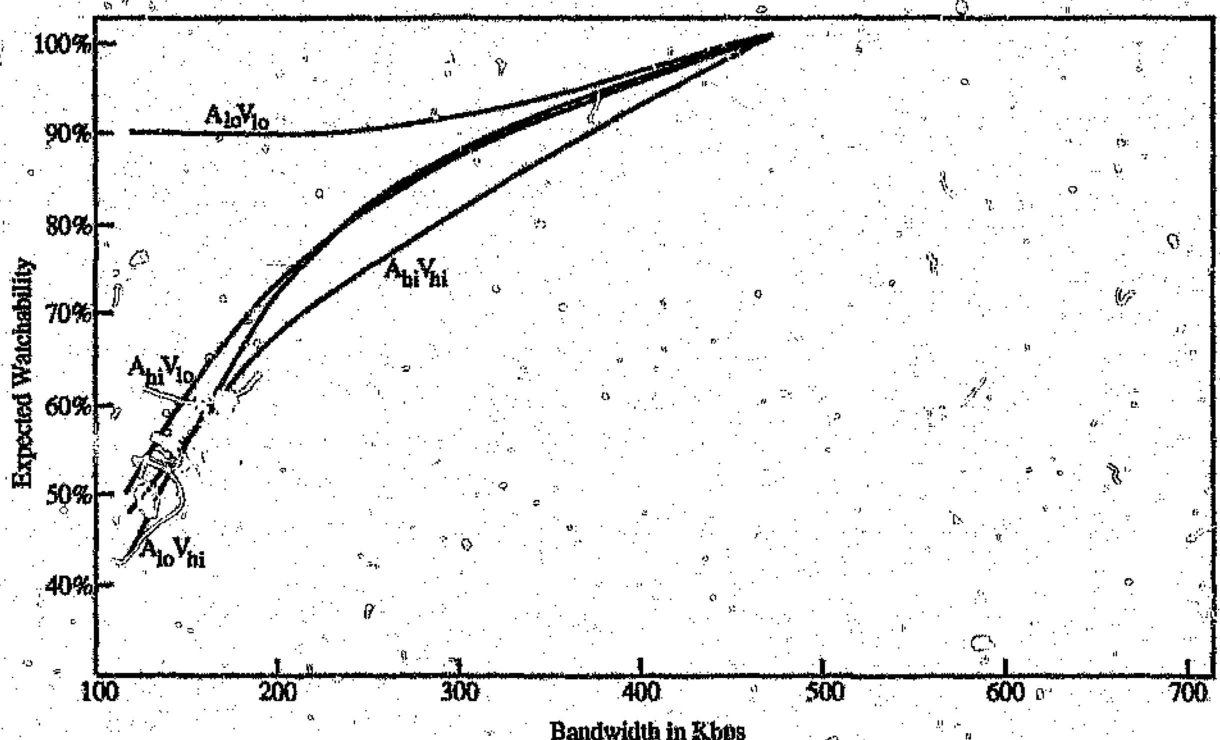


Figure 3 - Expected watchability expressed as a percentage VS T_{lo} bandwidths for T_{lo} VCS classes

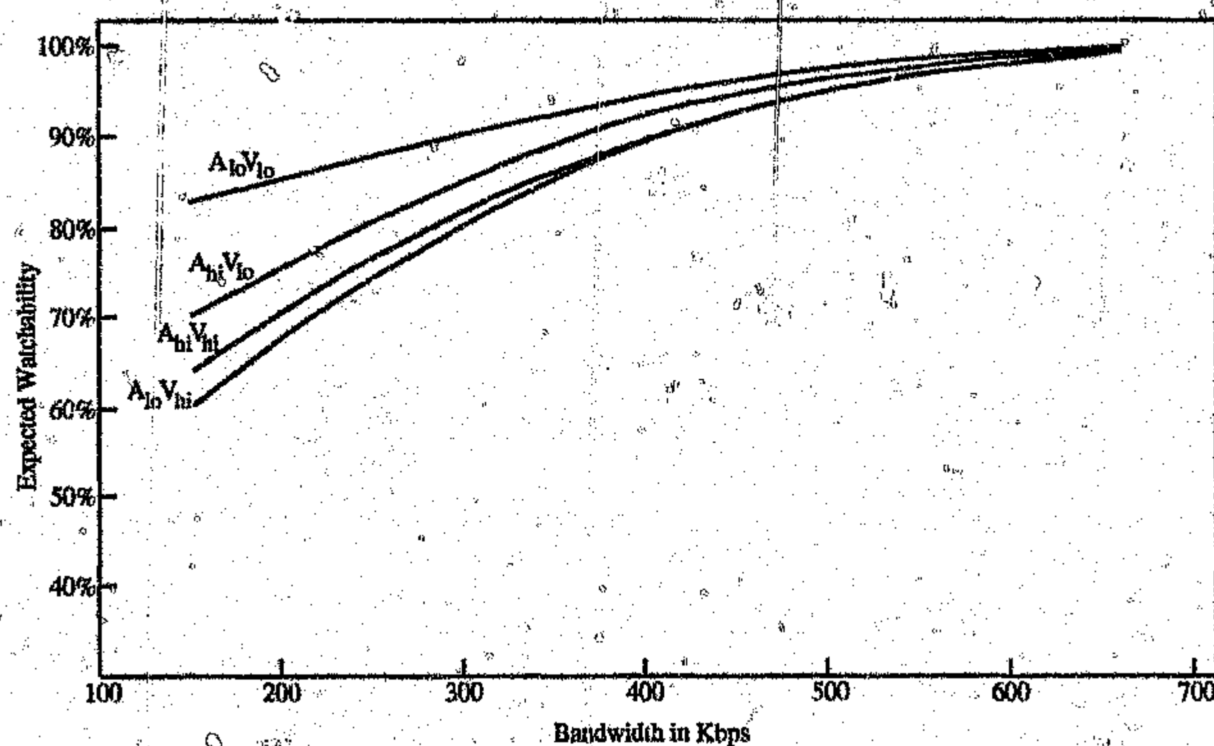


Figure 4 - Expected watchability expressed as a percentage VS T_{hi} bandwidths for T_{hi} VCS classes

Fusthermore, we can conclude that if you cut the bandwidth of T_{hi} VCS classes in half, there will be a 20% decrease (on average) in the watchability of those classes. And if you cut the bandwidth of T_{lo} VCS classes in half, there will be a 30% decrease (on average) in the watchability of those classes. Therefore, T_{lo} is more susceptible to changes in bandwidth, which is directly related to changes in frame rates.

An interesting observation is the logical order of VCS classes in T_{lo} and T_{hi} with respect to the relationship between watchability and bandwidth. In decreasing order (from 100% downwards), we have the following logical order for T_{lo} and T_{hi}, respectively : $T_{lo}A_{lo}V_{lo}$, $T_{lo}A_{hi}V_{lo}$, $T_{lo}A_{hi}V_{hi}$, $T_{lo}A_{lo}V_{hi}$ and $T_{hi}A_{lo}V_{lo}$, $T_{hi}A_{hi}V_{lo}$, $T_{hi}A_{hi}V_{hi}$, $T_{hi}A_{lo}V_{hi}$.

9. CONCLUSION

A multimedia on-demand server is constrained by the number of users that it can support concurrently. This constraint is a result of network congestion in light of the large data load that is characteristic of continuous media. Dynamic Q_oS attempts to alleviate the problem of network congestion, to some extent, by degrading the quality of service of individual user sessions, thus making more bandwidth available. Other bandwidth saving approaches may include intelligent buffering on the users' side by statistically analyzing bandwidth patterns of multimedia information stored on the server. Either way, the effects on user perception are not evaluated in any of these mathematical models. It is obvious that the watchability of multimedia information changes with physical degradation. But not so obvious is the extent of this change with respect to psychological aspects such as auditory and visual message, and temporal classification.

The VCS that has been presented in this report only describes specific continuous media situations. The realization of a full-length motion picture classification is described by a combination of classes from the VCS. Perhaps this is the basis for statistical analysis of bandwidth associated with multimedia information. It would definitely be of interest to investigate such a possibility in future research.

Dynamic throughput is by no means an intermediate step in the realization of an unconstrained multimedia on-demand service. Adaptive algorithms will continually be required since there is no upper bound on the complexity of user requirements.

The telephone companies of the world are the key to digital information transmission. Their move to digital communication mechanisms provides the foundation for distributed multimedia applications in the computing world. But most of these companies are prohibited by law from providing cable television services. Thus it is not surprising that the current technology typical of today's digital telephone networks is not, in its present form, feasible for the large-scale delivery of multimedia information to residential customers [29]. Regulations placed on telephone companies are currently under reexamination with many prospects beginning to surface. In the future, the public switched network will serve a greater role as an information distribution channel [25].

REFERENCES

[1] Anderson D.P. and Homsy G. (1991). A Continuous Media I/O Server and Its Synchronization Mechanism. *IEEE Computer*, 10(10), 51-57.

[2] Ang P.H., Ruetz P.A. and Auld D. (1991). Video compression makes big gains. *IEEE Spectrum*, 28(10), 16-19.

[3] Carpenter B.E., Landweber L.H. and Tirler R. (1992). Where are we with Gigabits? *IEEE Network*, 6(2), 10-13.

[4] Christensen L.B. (1988). *Experimental Methodology*. 4th Edition. Allyn and Bacon.

[5] Corner J. and Hawthorn J. (1986). Communications Studies: An introductory reader. 2nd Edition. Edward Arnold.

[6] Dixon D.F. (1989). Life Before the Chip: Simulating Digital Video Interactive Technology. *Communications of the ACM*, 32(7), 824-831.

[7] Droms R., Haber B., Gong F. and Maeda C. (1991). Report for the Joint SIGGRAPH/SIGCOMM Workshop on Graphics and Networking. *Computer Communications Review*, 21(2), 17-25.

[8] Edwards A.L. (1960). *Experimental Design in Psychological Research*. 2nd Edition. Holt, Rinehart and Winston.

[9] Foley J.D., van Dam J., Feiner S.K. and Hughes J.F. (1990). *Computer Graphics*. 2nd Edition. Addison-Wesley.

[10] Fox E.A. (1991). Advances in Interactive Digital Multimedia Systems. *IEEE Computer*, 10(10), 9-21.

[11] Green J.L. (1991). The Evolution of DVI System Software. *Communications of the ACM*, 34(4), 53-67.

[12] Grimes J. and Potel M. (1991). What is Multimedia? *IEEE Computer Graphics and Applications*, 11(1), 49-52.

[13] Harney K., Keith M., Lavelle G., Ryan L.D. and Stark D.J. (1991). The i750 Video Processor: A Total Multimedia Solution. *Communications of the ACM*, 34(4), 64-78.

[14] Herrtwich R.G. (1992). Summary of the Second International Workshop on Network and Operating System Support for Digital Audio and Video. *SIGOPS* '92, 6(2), 32-59.

[15] Le Gall D. (1991). MPEG : A video compression standard for multimedia applications. *Communications of the ACM*, 34(4), 47-58.

[16] Liebhold M. and Hoffert E.M. (1991). Toward an Open Environment for Digital Video. *Communications of the ACM*, 34(4), 104-112.

[17] Jurgen R.K. (1992). Digital Video. *IEEE Spectrum*, 29(3), 24-30.

[18] Neisser U. (1976). *Cognition and Reality: Principles and implications of cognitive psychology*. Freeman.

[19] Netravali A.N. and Haskell B.G. (1991). *Digital Pictures*. 3rd Edition. Plenum Press.

[20] Oliver D.M. (1992). SIGGRAPH '91 Workshop Report: Computer Graphics in the Network Environment. *Computer Graphics*, 26(3), 209-212.

[21] Oppenheim A.N. (1983). *Questionnaire Design and Attitude Measurement*. 11th Edition. Heinemann.

[22] Rangan P.V., Vin H.M. and Ramanathan S. (1992). Designing an On-Demand Multimedia Service. *IEEE Communications Magazine*, 3(7), 56-64.

[23] Ransom M.N. and Spears D.R. (1992). Applications of Public Gigabit Networks. *IEEE Network*, 6(2), 30-40.

[24] Ripley G.D. (1989). DVI - A Digital Multimedia Technology. *Communications of the ACM*, 32(7), 811-822.

[25] Rosenberg J., Kraut R.E., Gomez L. and Buzzard C.A. (1992). Multimedia Communications For Users. *IEEE Communications Magazine*, 30(5), 20-36.

[26] SAS Institute (1990). *SAS/STAT User's Guide Version 6*. 4th Edition. Volumes 1 & 2. SAS Institute Inc.

[27] Smith K.H. (1992). Accessing Multimedia Network Services. *IEEE Communications Magazine*, 30(5), 72-80.

[28] Steinmetz R. (1990). Synchronization Properties in Multimedia Systems. *IEEE Journal on Selected Areas in Communications*, 8(3), 401-412.

[29] Sutherland J. and Litteral L. (1992). Residential Video Services. *IEEE Communications Magazine*, 3(7), 36-41.

[30] Tokuda H., Tobe Y., Chou S.T.C. and Moura J.M.F. (1992). Continuous Media Communication with Dynamic Q_oS Control Using ARTS with an FDDI Network. *SIGCOMM '92 Conference Proceedings*, 10(4), 11-98.

[31] Williams N. and Blair G.S. (1991). *Distributed Multimedia Application Study*. Internal Report, MPG-91-11, Computing Department, Lancaster University.

[32] Wright A. (1992). Squeezing into the Picture. *Electronics World + Wireless World*, 21(8), 663-666.