References

[Bennett94] Frazer Bennett, Tristan Richardson and Andy Harter. *Teleporting - Making Applications Mobile.* Proceedings of Workshop on Mobile Computing Systems and Applications, Santa Cruz, December 1994.

[Davenport88] G. Davenport *Software Considerations for Multi-Media Video Projects* Proceedings X11 Video Extension Technical Meeting, 1988.

[Eriksson94] Hans Eriksson. *MBone: The Multicast Backbone* Communications of the ACM, Vol.37, No.8, August 1994.

[Harter94] Andy Harter and Andy Hopper. *A Distributed Location System for the Active Office.* IEEE Network, Special Issue on Distributed Systems for Telecommunications, January 1994.

[Hayter91] Mark Hayter and Derek McAuley *The Desk Area Network* Published in ACM Operating Systems Review, 1991.

[Hopper88] Andy Hopper and Roger Needham. *The Cambridge Fast Ring Networking System.* IEEE Transactions on Computers, Vol. 37, No. 10, October 1988. Olivetti Research Technical Report 88.1

[Hopper88] Andy Hopper and Roger Needham. *The Cambridge Fast Ring Networking System.* IEEE Transactions in Computing. Volume 37, No.10. October 1988.

[Hopper91] Andy Hopper *Design and Use of High-Speed Networks in Multimedia Applications* Proceedings of 3rd IFIP Conference on High Speed Networking, Berlin, March '91.

[Jones94] Alan Jones and Andy Hopper. *Handling Audio and Video Streams in a Distributed Environment.* Proceedings of 14th ACM Symposium on Operating Systems Principles, December 1993.

[King92] Tony King. *Pandora: An Experiment in Distributed Multimedia.* Proceedings of Eurographics '92, September 1992. Olivetti Research Technical Report 92.5.

[Leslie93] Ian Leslie, Derek McAuley and David Tennenhouse *ATM Evewrywhere ?* Published in IEEE Network, March 1993

[Lindblad95] Christopher Lindblad, et al. *ViewStation Applications: Implications for Network Traffic.* Proceedings of JSAC, 1995.

[Tennenhouse94] David L. Tennenhouse, et al. *The ViewStation: A Software-Intensive Approach to Media Processing and Distribution* Multimedia Systems Journal, August 1994.

[Ward94] Andy Ward, Andy Harter et al. *The Cambridge Panorama.* Image browsable on the World Wide Web, http://www.cam-orl.co.uk/

[Wray94] Stuart Wray, Tim Glauert and Andy Hopper. *The Medusa Applications Environment.* Proceedings of International Conference on Multimedia Computing and Systems, Boston MA, May 1994. Olivetti Technical Report 94.3.

generating contextual information which can be used to make applications which are adaptive and far more flexible.

5.2 Third Generation Multimedia Applications

The third generation of multimedia applications will not only exploit improvements in the network technology, but more importantly will take advantage of any contextual or location information made available. As mentioned earlier, this information will come from agents interpreting streams, as well as from resource and *active* location databases.

We have experimented with these applications in both Pandora and Medusa, although the integration of location system services has in both cases been retrospective. We consider provision of such information to be a necessary underlying component of any future systems we might build - it should be intrinsic to the system, and available at all levels within it.

The *Teleporting System* [Bennett94] is an example of an application that may be able to exploit third generation networks. It offers users the ability to interact with existing computer applications from any suitably equipped networked terminal. The system presents a means by which we can experiment with applications that adapt and reconfigure based on any contextual and location information available.

At present, multimedia applications do not function within the teleporting system, but the system's integration with third generation networks may make this possible. A subject of ongoing research is the provision of systems support, at all levels, for multimedia applications which adapt to the kind of network they are reign routed on, the kind of peripherals they are using, as well as the environment within which they are operating.

6 Conclusions

We recognise two important details which will influence the further development and wider acceptance of the technologies outlined above. These are *cost* and *scalability*. While networks improve and begin to support the underlying QoS requirements of future multimedia applications, each of these details must be considered.

The *cost* aspect involves charging for network use, which will have to change significantly. There will be a blur between the notions of long and short distance, prompted particularly due to mobile networks which force a relaxation in this distinction. There is also the cost of a network connection itself, which must not limit us in our choice of networked peripherals.

The wide-area propagation of these third generation networks will see their use on many different *scales*. It will extend from the household level, with perhaps even with an *ATM doorbell*, through to the wide area where we will see *information oversupply*, adaptability and ease of connection.

in the sense that re-mapping of stream endpoints will be in general very easy. Figure 3 shows a possible third generation network configuration.

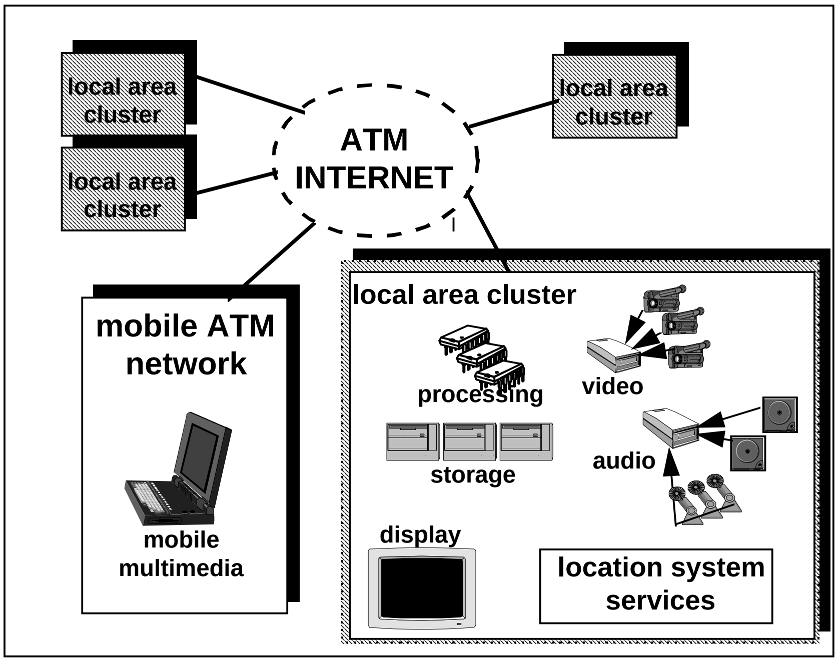

FIG. 3. *A Third Generation ATM Network*

Clearly the most challenging of environments for third generation networks is that within the wireless domain. While we cannot expect the same kind of performance from a wireless network as from a wired one, the network's ability to offer performance guarantees will make it possible for applications to adapt according to the network they are using.

Supporting mobility in the third generation network involves more than just wireless networks, however. We must support *mobile users* who wish to exploit existing, fixed network facilities as they move from place to place. Systems must be able to cope with users' mobility by making use of information about context as well as *active* location information. With this, it will be possible to perform reconfiguration and re-mapping of stream connections, according to a change in environment.

Another important feature of this third generation of network will be the ease with which connectivity is possible. The cost of a network interface should not limit the choice of what can be connected directly to the network.

The improvement in performance will mean that the emphasis may be less on *compression* and more on *annotation* and *expansion*. Since it will not be as necessary to perform compression on streams, *agents* and other applications will be able to easily monitor and maniuplate multimedia data, Our modular approach extends further. An active interface to a peripheral is represented by a *software module*, with input and output nodes. SAMs support multiple connections from these modules and can, in general, simultaneously service several streams, each of a different format. The software modules are connected together in pipelines along which the data will flow.

4.3 Medusa and Second Generation Multimedia Applications

Introducing many more streams into the system makes the traditional multimedia applications of video-phone and video-conference more elaborate. A Medusa video-phone simultaneously offers four different camera views of both ends of the connection. Each of the four images can be viewed at different sizes and frame rates. This *information oversupply* makes the use of these applications much more natural. A multi-view video-phone provides many perspectives of each end of the connection, allowing the users to communicate more freely.

With the ability to manipulate multimedia data in software, inserting *agent* modules into the software pipeline has proved particularly interesting. These agents need not perform any particularly complex function on the data. For example, an agent might monitor several video streams, each a different view of the same office, and select the one with the most movement for viewing. Agents can perform more complex functions, including face, voice and gesture recognition. Agents that enable these are at an early stage of development, but nonetheless serve to demonstrate the flexibility of our approach.

Streams and information provided by agents can be accompanied by contextual information from sensors and *active* location systems[Harter94]. In the setting of a next generation of *ATM Internets*, this information about the nature and state of the environment will, we believe, be highly important.

5 Third Generation Networked Multimedia

5.1 Third Generation Networks

We recognise ATM as a unifying network technology, and anticipate its use on a much wider scale, across many different substrates, in particular in wireless computing networks. The *variety* of networks will be no different to the present, but throughout the full breadth of *third generation* networks there will be provision, at some level, for the timely delivery of continuous media. Streams in the third generation network environment will be abundant and extremely economical. Furthermore, they will be *portable* ously support tens or hundreds of streams, producing perhaps thousands of streams within the local area. Clearly, very complex systems can be imagined in this scenario, and the Medusa project attempted to provide a framework under which some of these ideas could be realised.

4.2 Medusa - Second Generation Networked Multimedia

Project Medusa [Wray94] builds on our experience and design principles adopted in the development of networked multimedia systems [Jones94]. The improvements in our infrastructure already mentioned make it possible to be more flexible in our approach. A modular approach has been taken at all levels, from peripheral design, to system and applications software. This approach promotes easy re-use or replacement of individual components within the system. Figure 2 outlines the components in the Medusa environment.

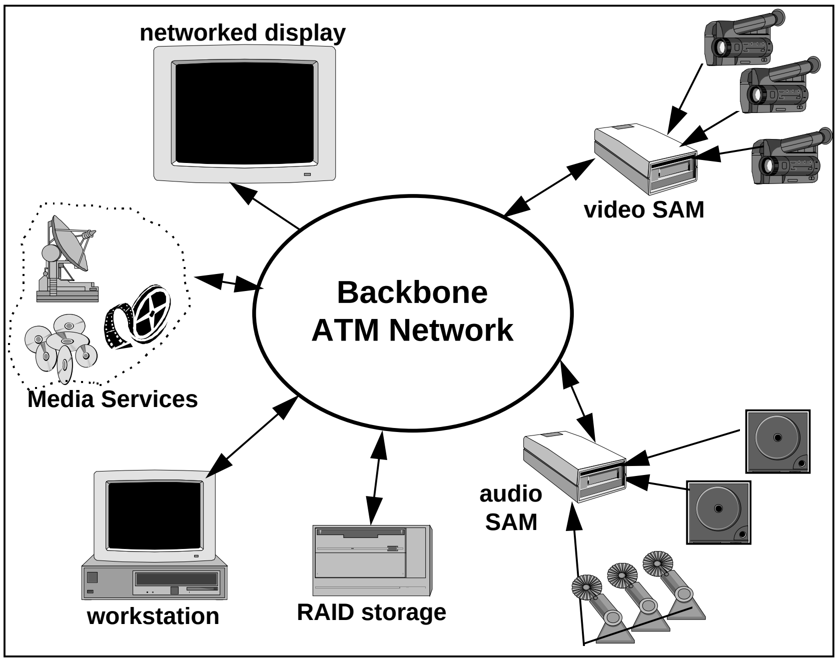

FIG. 2. *Components of the Medusa System Environment*

Our network is constructed from an ATM switch fabric operating at 100Mbs^{-1}. Both network interfaces and ATM switches are common, widely re-used components, and the *direct peripheral* approach permits a collection of ATM cameras, audio interfaces and frame-stores, each independently connected to the ATM network. The modular design means that a standard network interface component is re-used within each of these devices. Each peripheral is known as a *Smart ATM Module* or SAM.

some investigation [Hopper91]. The most popular of these were the *videophone* and the multi-way *video-conferencing* application which permitted four-way video and audio communication. In the four-way video conference, each workstation received four video images, all at full frame rate, and the audio from each end of the conference was mixed in software on the Pandora's box.

A networked multimedia file-server was developed, making applications such as video mail possible. This also proved popular, and patterns of its use suggest that video mail can complement e-mail in enhancing communications facilities.

A Media Server, offering access to TV channels, a CD juke box and the radio was also available from a suitably equipped Pandora's Box. By combining the file server with real-time sources from the media server, we were able to, for example, log news bulletins as they were broadcast, providing the latest news on line at all times.

The *hands off*, approach adopted by the Pandora system made integration with the underlying network much simpler. However, *hands off* can limit the number of potential applications. Furthermore, it was difficult to achieve any different levels of scale with the Pandora system. Only a single set of peripherals could be connected to a Pandora's box, and so adding individual components to the system was not possible. A more flexible approach is needed, and with network interfaces and network fabrics both becoming more economical, alternative approaches become attractive.

4 Second Generation Networked Multimedia

4.1 Second Generation Networks

The general acceptance of ATM as a substantial network technology has resulted in the development of faster, more general purpose ATM networks. This *second generation* of network offers more flexible network configurations, while still offering the important fine-grained sharing necessary for continuous media transfer. Streams in a second generation network are more numerous and more lightweight, offering synchronisation mechanisms and guarantees about timely delivery.

We can also look to extending the role of the workstation within a second generation network. By permitting the *hands on* approach, we can experiment with applications which attempt to make some interpretation of multimedia data. Such applications might include *agents* which monitor sound and motion levels within a room by analysing the data streams coming from it.

At ORL, the development of a general purpose ATM network exhibiting these characteristics has lead to further multimedia experiments. Our work is heading towards systems incorporating peripherals that can simultane-

3.2 Pandora - First Generation Networked Multimedia

The Pandora project was an early attempt at integrating digital video and audio within the desktop computing environment. The system that was developed exploited a general purpose ATM network, based on the Cambridge Fast Ring [Hopper88], using a dedicated sub-system - *Pandora's Box*.

Pandora's box combined audio and video capture and processing boards with a network interface. The box was controlled by a workstation over a high-speed serial link. In these respects, then, Pandora adopted a *hands off*, *near the workstation* approach. Figure 1 shows a Pandora's box and its associated peripherals.

Several sources of reference outline the hardware components of the Pandora system [King92], the network involved [Hopper88] and the design principles adopted [Jones94]. We reflect here on some of the more popular applications that were achieved with the system.

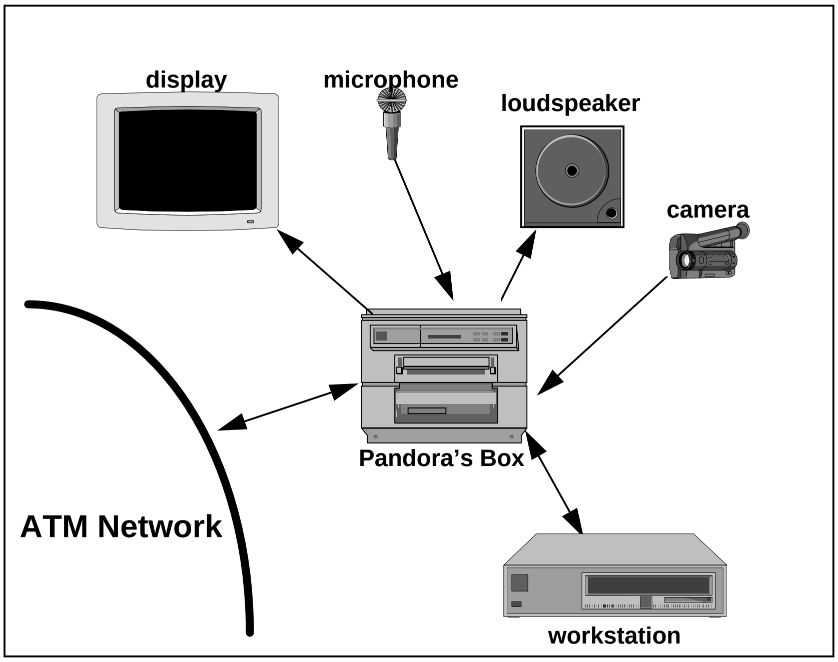

FIG. 1. *Components of a Pandora's Box*

3.3 Pandora and First Generation Multimedia Applications

With a small number of synchronised video and audio connections available, many reasonable applications can be achieved. The Pandora project implemented many applications, and their patterns of use were the subject of region of the image for a close-up. The image is 'nearly live', and in this respect it has a very low-grade real-time requirement.

The *MBone* [Eriksson94] is another example of a mechanism introduced to improve the suitability of zero generation networks for multimedia applications. By providing a multicast backbone to the Internet, the MBone makes wide area distribution of real-time audio and video possible. The MBone is now used regularly by a growing community of users. However, with more than one pair of video and audio channels it can reach a level of high load. There is also no standard mechanism for resource control or real-time traffic control. Without such facilities inherent in the underlying network, the MBone will always suffer shortfalls.

So, our options are limited by zero generation networks. We need a faster network, but more particularly we need a network capable of administering the bandwidth requirements of applications. Applications need to be able to rely on the network to provide a consistent quality of service on these high-speed connections.

3 First Generation Networked Multimedia

3.1 First Generation Networks

A preferred network technology adopted for multimedia experiments is *Asynchronous Transfer Mode* (ATM). ATM has long been recognised as offering a style of networking much better suited to continuous media experiments. The use of smaller data packets, of a fixed size, permits a finer grain of sharing of the network resources. The network can impose restrictions on the rate at which any single endpoint may use the network, and in doing so is able to offer endpoints some guarantees about the allocation of network resources.

With *first generation networks* of this kind, additional functionality can be achieved in our multimedia applications. With sufficient bandwidth, it is now possible to have applications which make use of a small number of real-time streams. Video connections can be bi-directional, accompanied by an audio stream, and additional networked peripherals, like media servers and multimedia file-stores become feasible.

First generation networks can carry a substantial number of streams, and some synchronisation is possible between them. The level of system complexity immediately increases, however, as synchronisation issues arise. There must be sufficient buffering at endpoints to make stream synchronisation possible.

With a faster, fairer network, latency and jitter problems can be reduced to a more tolerable level. However, conditions of high load will still arise, and mechanisms for coping with this must be inherent in both systems and application design.

of the real-time and bandwidth requirements of multimedia applications were, in general, not inherent in the design of these networks.

What can reasonably be achieved depends on the level of load of the network, and in the local area we can expect little more than a single stream of data. This limitation avoids the need for complex synchronisation techniques, and in the case of a uni-directional stream, buffering can be used, where appropriate, to overcome jitter problems.

Some tolerable performance may be possible when the network is lightly loaded. Under conditions of higher load, however, performance soon degrades, and few recovery strategies are possible. Compression of streams is almost mandatory, and while it can contribute greatly to perceived performance, there is a price to pay. *Hands on* strategies for manipulating data become more difficult, and the trade-off between compression and preservation of quality must be considered.

In the wider area, all of these problems are accentuated. The Internet, as a giant interconnection of a wildly disparate collection of networks, offers little opportunity for achieving any real-time, continuous media applications. It has evolved to accommodate the demands placed on it, but we now find it in a state where it can no longer adequately support the kinds of function expected of it.

Examples of applications which highlight this shortcoming include those of the World Wide Web. In offering a simple and extremely flexible global information access mechanism, the Web has demonstrated the need for faster network connections. It also highlights the need for bandwidth reservation mechanisms.

2.2 Applications of Zero Generation Networks

Despite these apparent limitations, it has been appealing to experiment with multimedia over zero generation networks. Without a network that can meet real-time QoS guarantees, applications that perform *bulk delivery* of data are still possible. An example might be a *Video on Demand* style system, where with sufficient buffering, stream synchronisation can be performed at the receiving end. Another approach might be to use self-synchronisation techniques, by combining the audio and video streams into one. With the real-time element removed, however, applications can support only limited interaction between the system and the users.

One example of a *zero generation network* application is the *Cambridge Panorama*[Ward94]. Six video cameras have been placed on the roof of ORL's laboratory. Each camera periodically produces a still image of a different portion of the view across the Cambridge rooftops. These stills are 'stitched' together in software to produce a single, 180^o panoramic image. This *Cambridge Panorama* can be viewed via the World Wide Web, and is interactive in that it is possible for the viewer to zoom in on a particular network interfaces decreases, and as the capacity of the underlying network increases.

The *direct peripheral* approach offers many new flexibilities. Such a system is more tolerant of failure, since individual components operate independently of each other. Furthermore, since these components can be configured in different ways, the system is more easily extensible. As the number of endpoints and streams within the system increases, the load on any single point is reduced, allowing for greater sharing of available bandwidth.

Hands Off/Hands On the Data.

The degree to which any processing of the multimedia data is possible also needs consideration. At a crude level, we can choose either the *hands off* or *hands on* approach. In the *hands off* approach, a workstation may mediate the setting up and tearing down of streams, but the stream data itself cannot be handled directly. This approach may be taken to ensure reasonable efficiency. The scope of applications may be limited in this case, however, since it is difficult to manipulate, monitor or interpret the multimedia data.

By contrast, the *hands on* approach allows closer involvement with the data, either by routing this data through the workstation's network interface, or by making it available to some other, networked, processing component. The *hands on* approach makes possible a class of applications which can manipulate, monitor and interpret multimedia data. We conclude that where network bandwidth and workstation I/O speeds permit, this *hands on* approach is clearly favoured.

The choice of approach to system design is taken in the light of the characteristics of the available network. At ORL, our multimedia experiments have spanned different generations of network, and it is in terms of these generations that we describe the systems we have developed. We continue first with a description of *zero generation* networked multimedia, by reflecting on what can be achieved over today's common network infrastructures.

2 Zero Generation Networked Multimedia

2.1 Zero Generation Networks

Conventional networks such as Ethernet have certainly received attention as candidates for networked multimedia experiments. In general, such networks are able to handle the typical volume of traffic generated by a single, compressed, video stream. However, streams in these *zero generation networks* have no mechanisms for making *Quality of Service* (QoS) or real-time guarantees. There is also no reasonable way of performing synchronisation between streams. This is hardly surprising. Considerations

Analogue/Digital

Early experiments in networked multimedia involved the use of analogue switching techniques. Here, video was shipped on a dedicated analogue network, distinct from control information sent via normal datapaths [Davenport88]. The computer was used simply to control the switching of these video streams. Audio was either switched similarly, on a separate network, or lower quality audio was distributed via the computer's own network.

Analogue techniques offer no reasonable opportunities for scaling. We are bound by the limits of our analogue network, both in terms of capacity and breadth of coverage. A digital approach to shipping video and audio is clearly more favourable, provided the network has sufficient capacity to tolerate it. Using digital video makes it possible to use more easily scaled networks as well as being able to store, edit and retrieve data more effectively.

Inside/Outside the Workstation.

A policy must be adopted on the role of the workstation within a multimedia system. We could choose to route multimedia data through the workstation's network interface, just like any other data. The workstation would be directly responsible for rendering video images on the display, as well as handling any audio through its own devices.

This *inside the workstation* approach may be desirable if the network interface is relatively expensive, since there is no need for any new network endpoints. However, current workstations, and more particularly the operating systems that they use, are not well suited to handling multimedia applications in this way. This is primarily due to the I/O intensive nature of these applications. Without a powerful enough bus, *inside the workstation* integration of multimedia peripherals may be too restrictive.

An alternative might be to route data through the workstation, but in a more direct manner, either through the workstation's main network interface, or through an additional, dedicated, network interface. In this way, it might be possible to route video directly in to the workstation's frame-buffer, for example. A variant of this approach would be to use a dedicated sub-system which handles multimedia traffic and interfaces directly to the local multimedia peripherals. The computer interacts with this sub-system to perform control functions, but the sub-system handles the interface to the network. This approach might be described as a *near the workstation* one.

A third approach is to choose a design which permits peripherals to be connected directly to the network. In this way, the bottleneck at the network interface is reduced, as is the load on the workstation. This *outside the workstation* or *direct peripheral* approach can been adopted as the cost of

Chapter 1
Developments in Networked Multimedia

Frazer Bennett* Andy Hopper*†

May 1995

Abstract

This paper discusses developments in multimedia applications and systems from a network-oriented point of view. We highlight the different generations of network that have been used as substrates for multimedia experiments, focusing particularly on two systems that we have developed.

Finally, we look forward to a third generation of network, the subject of more recent attention, which we hope will make ubiquitous, mobile, networked multimedia applications possible.

1 Introduction - Approaches to Networked Multimedia

As the peripherals components of a networked multimedia system improve, so users' expectations of such a system increase. Furthermore, while the networks become faster and more flexible, network connections are becoming less costly. It is not difficult, then, to imagine elaborate systems involving a great many sources of multimedia data distributed across a wide area. The result is that heavier and far more stringent demands will inevitably be placed on any underlying network.

Clearly, the level of function and performance that can be achieved by networked multimedia systems depends critically on the nature of the network employed. It is not surprising, then, that groups have approached much of this work by trying to develop appropriate network fabrics for the support of continuous media at high speeds. Work on high-speed local area networks has prompted an interest in multimedia applications as a means of 'soaking up' the new bandwidth that was being made available.

Decisions about the composition of a networked multimedia system must be made in the light of the available network technology. We outline below some contrasting approaches that can be taken to system design.

*Olivetti Research Ltd, 24a Trumpington Street, Cambridge, CB2 1QA, United Kingdom. http://www.cam-orl.co.uk/

†University of Cambridge Computer Laboratory, Pembroke Street, Cambridge, CB2 3QG, United Kingdom. http://www.cl.cam.ac.uk/