Introduction To

Multimedia Conferencing

29 September 1998

Version 1 (DRAFT)

Acknowledgements

This document is based on the UKERNA document called "Introduction to Multimedia Conferencing and the SHRIMP Tools", which will soon be obtainable from http://www.ja.net/service_development/video/service_developments/shrimp/index.html.

In their current form, the documents are nearly identical.

Disclaimer

UCL cannot accept any responsibility for any loss or damage resulting from the use of the material contained in this document. The information is believed to be correct, but no liability can be accepted for any inaccuracies.

Contents

1. Introduction

This document is an introduction to multicast videoconferencing. It introduces multicast technology and a selection of the software tools available for taking part in multimedia conferencing and it tells the reader what to do to get started.

The software tools introduced in this document are all available from the UCL Multimedia Web Site http://www-mice.cs.ucl.ac.uk/multimedia/software/ Documentation for each of the tools is also available from the site.

The tool selection which is introduced in this document consists of:

- a tool for announcing multimedia conferences, and which allows the user to join an announced conference by the click of a button (SDR).
- an audio tool (RAT)
- a video tool (VIC)
- a shared text-based workspace tool (NTE)
- a shared drawing-based workspace tool (WB/WBD)
- a tool for announcing multimedia conferences, and which allows the user to join an announced conference by the click of a button (SDR).

This document is intended for readers who have little or no experience with multimedia conferencing on the Mbone. It gives an introduction to the underlying technology and the software. A glossary of terms is included at the end of the document.

2. Background

Videoconferencing has been possible for some time. Telephone companies offer videophone and videoconferencing services, but most of these require the purchase of special equipment and the setting up of special rooms for videoconferencing. They also require a Multipoint Control Unit (MCU) which causes multi-way conferences to be set up as a series of point-to-point connections between the MCU and each participant. This is an expensive and often cumbersome way of holding remote meetings, and does not scale to supporting large user groups who want to have interactive meetings. The Internet provides an alternative that can be scaled more economically.

The Internet is mainly used for supporting asynchronous communication such as e-mail and the WWW (World Wide Web), but parts of it can also be used for transmitting voice and video in real time, making multimedia conferencing on the

Internet possible. Conferencing on the Internet is different because it uses *multicast* transmission instead of the point-to-point communication used by the telephone companies. It is this multicast technology (explained more fully in the next section) that gives the Internet community an economic way of supporting multi-way multimedia conferences between large numbers of participants.

3. Introduction to the Technology

A computer-supported conference involves the transmission of audio, video and data (e.g. from shared workspaces) – hence "multimedia". In this section we review the process of capturing, sending and receiving these multimedia streams. This is followed by a brief introduction to multicasting over the Internet.

3.1. Transmission processes

Analogue video and audio signals must be digitised before they can be transmitted. The digitised data is then compressed to reduce the quantity of information to be transmitted. There are different ways of digitising and compressing the analogue streams produced by the camera and microphone. These variations are also reflected on the receiving end of the interaction where the process must be reversed to provide analogue video and sound again.

Digitising may be achieved in several different ways. CODEC, as a contraction of COder/DECoder, is the generic term for any device or process performing the conversion of an analogue stream to digital and back, whether hardware or software. A hardware CODEC may handle video or audio or both together. Software CODECs generally refer to a specific coding process, e.g. a G.722 CODEC, a H.261 CODEC, thus a device loosely described as a CODEC may feature several separate CODECs. Many workstations use a frame grabber to supply the video, frame by frame, to a software CODEC program. Audio may be digitised by workstation hardware.

Video compression may involve several stages. Basic compression is applied to each frame. Further compression is achieved by calculating differences between successive frames and sending only those differences. Further savings in bandwidth can be made by predicting such changes and sending only the information which modifies the predictions. This technique uses motion vectors to predict changes due to motion in the picture. In both these cases there is the need to periodically send full reference frames (INTRA frames). These frames provide a new point of reference for the succeeding *difference* frames (INTER frames). The compression process may be performed in hardware or in software.

After compression, the data are segmented into small fragments – *packets* – before transmission over the network.

At the receiver the data are reassembled, uncompressed and converted back to analogue signals. In some cases the uncompressed video is passed directly to the window management software for display (for example in Sun workstations). In other cases it may be converted to a TV standard format such as PAL and displayed on a standard TV monitor.

3.2. The Multicast backbone (Mbone)

Multicast provides one-to-one, one-to-many and many-to-many network transmission services for applications such as videoconferencing which need to communicate with several hosts simultaneously. A one-to-one connection is like a telephone call, where a fixed connection between two telephones is set up for the duration of the call. A one-to-many is a broadcast where one participant is transmitting to an unlimited number of listeners/viewers. In principle, a many-to-many connection allows an unlimited number of participants to communicate with one another at the same time. It has existed for several years on Local Area Networks. On the Internet it is implemented through the Multicast backbone, or Mbone.

When a conference begins, a tree-like structure of connections is established between the participants. This is done dynamically; it is not necessary to set up all the links in advance and when a new participant joins the conference another branch is just added to the tree. The tree can also be "pruned" dynamically, so that inactive branches are detached, making it more economical than point-to-point conferencing.

On the Internet, networks and hosts (e.g. workstations) are located by means of addresses. IP (Internet Protocol) addresses consist of four numbers which, between them, identify the network and host. Transmission between sites is enabled by routers which communicate with one another and hold tables of routes between addresses. There may be several possible routes between two hosts.

The Mbone is a virtual network. It uses the same physical media as the Internet but at present, many Internet routers are not able to deal with multicast addresses. Special multicast routers *mrouters* are therefore needed. An mrouter is simply a workstation, located on a LAN, which can interpret multicast addresses. Multicast routing can also be performed by routers with suitable hardware and software. The mrouter wraps up multicast messages as "normal" Internet traffic – unicast IP packets – and passes them between mrouters. Thus the Mbone consists of connections between these multicast-aware mrouters, along which encapsulated multicast traffic can flow. These connections are known as tunnels. It is important to remember that the traffic in the tunnels uses the real bandwidth of the Internet. For this reason, moderation is always necessary in using the Mbone tools.

4. The Conferencing Tools

Multimedia conferencing on the Mbone makes use of separate tools – a tool for each element of a typical multimedia conference. However, there is an integrated interface (at the time of writing only available for Windows95/NT4.0) which integrates the audio tool (RAT), the video tool (VIC) and two shared workspaces (WBD and NTE) into an easy to use interface. The tools can be classified into the following groups:

- Conference management
- Video
- Audio
- Shared workspace
- Integrated Interface

4.1. Conference management tool

SDR (Multicast Session Directory) is a tool which assists the user in setting up and joining conferences. All conferences set up using SDR are listed in the tool's main window (see Figure 1), rather like a TV listings guide - except that SDR lists conferencing sessions on the Mbone.

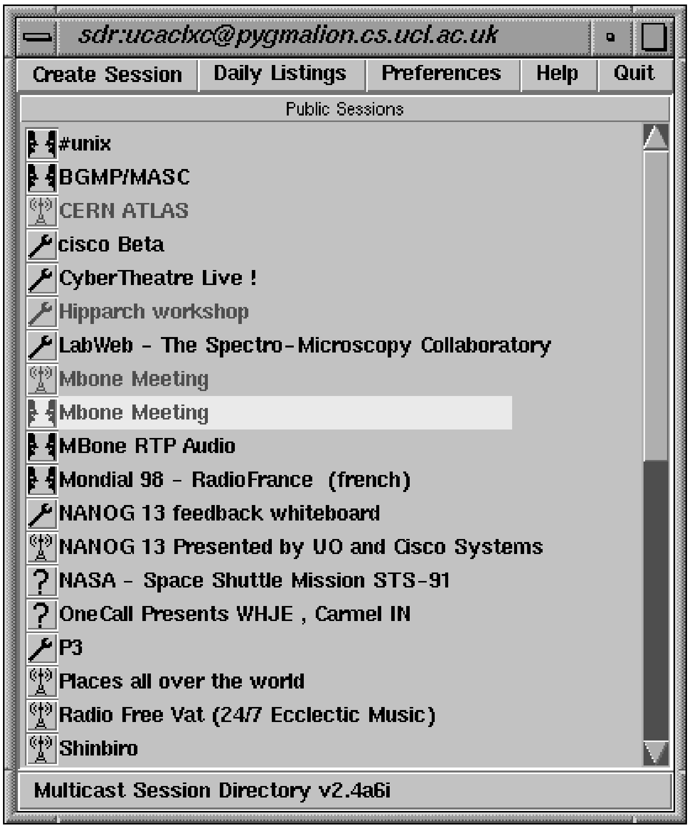

Figure 1: SDR - Session Directory Tool

More information about each session can be obtained by clicking on its name in the list. A window will appear with further details about the event and an invitation to join. When you join a conference using SDR the appropriate tools for that session will automatically be started with the correct parameters. Alternatively, the user can decide to start up only a few of the tools.

With SDR anyone can create and advertise their own conferences, and invite other people to join if they wish. It provides a framework for setting up conference sessions and automatically configures the relevant tools. Once a conference has been set up it will be announced to other users of SDR for a certain period of time (usually specified by the organiser of the conference). During this period other SDR users can join the session as they see fit. It is also possible to invite someone for a quick 5-minute consultation or a full meeting - like making a phone call.

Versions 2.5 and higher of SDR support secure conferencing. This means that it is possible to make session announcements which can only be seen by a certain group of people. It is also possible to encrypt sessions, so only people with the correct password can participate in the sessions.

4.2. Video tool

The video tool is called VIC (see Figure 2). The main window shows thumbnail images of participants who are transmitting video. Next to each image is displayed information about the identity of each user, and technical information about their transmission. Images can be displayed in a variety of sizes.

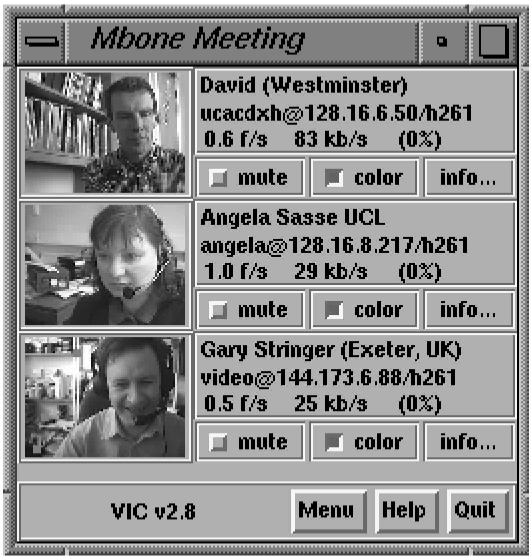

Figure 2: VIC - Video Tool

The *menu* button at the bottom of the tool opens a window from which you can initiate transmission of your own video. Other settings can be selected to suit your requirements.

4.3. *Audio tool*

The audio tool is called the Robust Audio Tool (RAT).

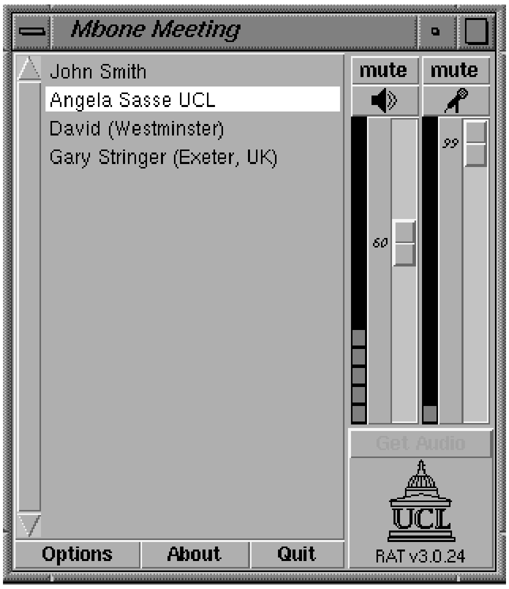

Figure 3: RAT - Audio Tool

It has a main window in which a list of participants is displayed. The current speaker is highlighted. The controls on the right of the window allow the user to control the volume of the incoming and outgoing audio.

Pressing the *Options* button brings up a window that offers a number of different user choices. A useful facility is the possibility of selecting an interactive mode for group work such as meetings, where there will be a number of different speakers, or a single-sender/multiple receiver mode for listening to talks at conferences or other broadcasts.

4.4. *Shared workspace tools*

There are two widely used shared workspace tools, a whiteboard called WB/WBD and a text editor called NTE, both of which are fairly basic. Shared workspaces allow all users to write or draw in the same space, thus supporting collaboration between participants using visual and textual materials. In meetings NTE can be used for displaying the agenda and for taking minutes. In remote tutorials WB/WBD can be used for writing questions, answers and for displaying course material.

4.4.1. WB

WB is a full-colour shared display area on which participants may write, draw and type. Their contributions will be visible to all participants in the session. In a lecture situation it can be used like an overhead projector to display a sequence of slides, using its facility to import documents and pictures generated using other sources.

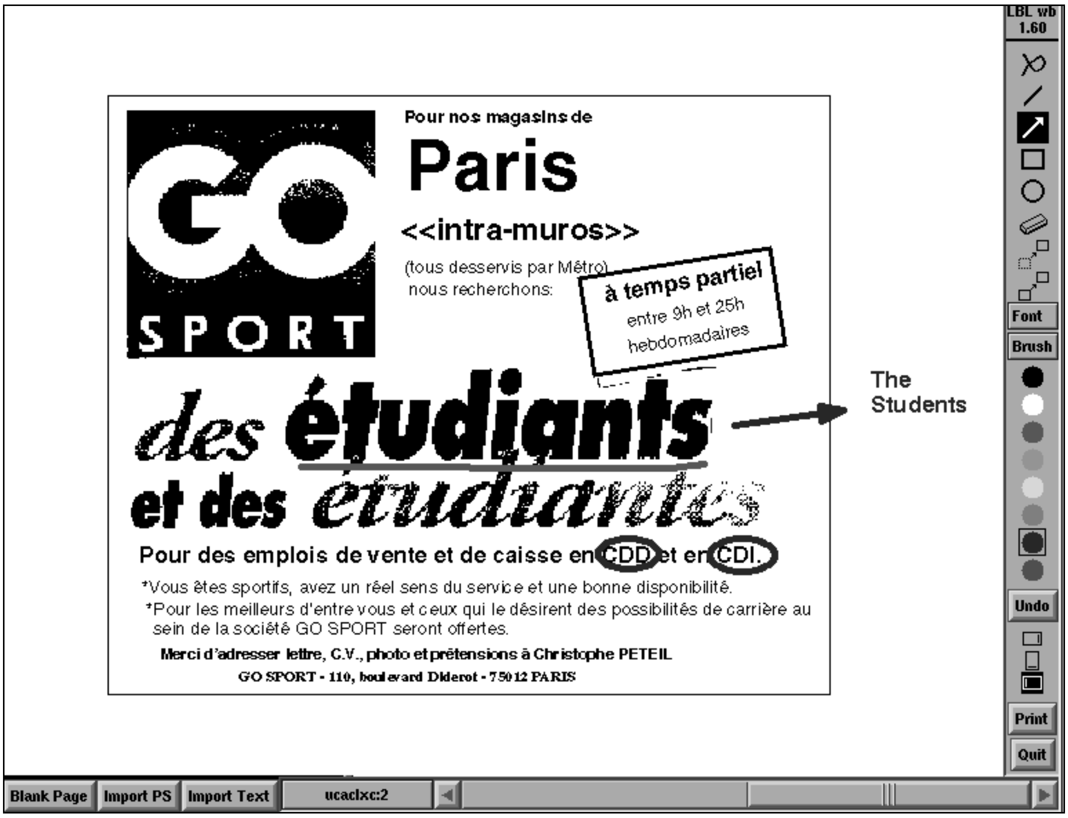

Figure 4: WB - Shared Workspace Tool

WB offers basic drawing tools with a choice of colour, font and brush size. There are controls for changing and creating pages. Pictures and text can be entered either by importing a file or by using the keyboard and mouse. WB is not a word processor. Objects may be annotated but editing facilities are limited (you can delete, copy and move an object, for example, but not resize it). As well as allowing all participants to share the same page, it will also permit several people to work on different pages.

WB is the only tool mentioned here which is not available on Windows95/NT4.0 platforms. Windows95/NT4.0 users are advised to use WBD instead.

4.4.2. WBD

WBD is a clone of WB which can run on all platforms. It is compatible with WB, but as it is far less stable, only WB is recommended for Unix platforms.

WBD is a shared drawing tool with very basic functionality. The tool offers a shared area where all participants can write and draw and import certain types of graphic files. Contributions by any one participant can be seen by all others.

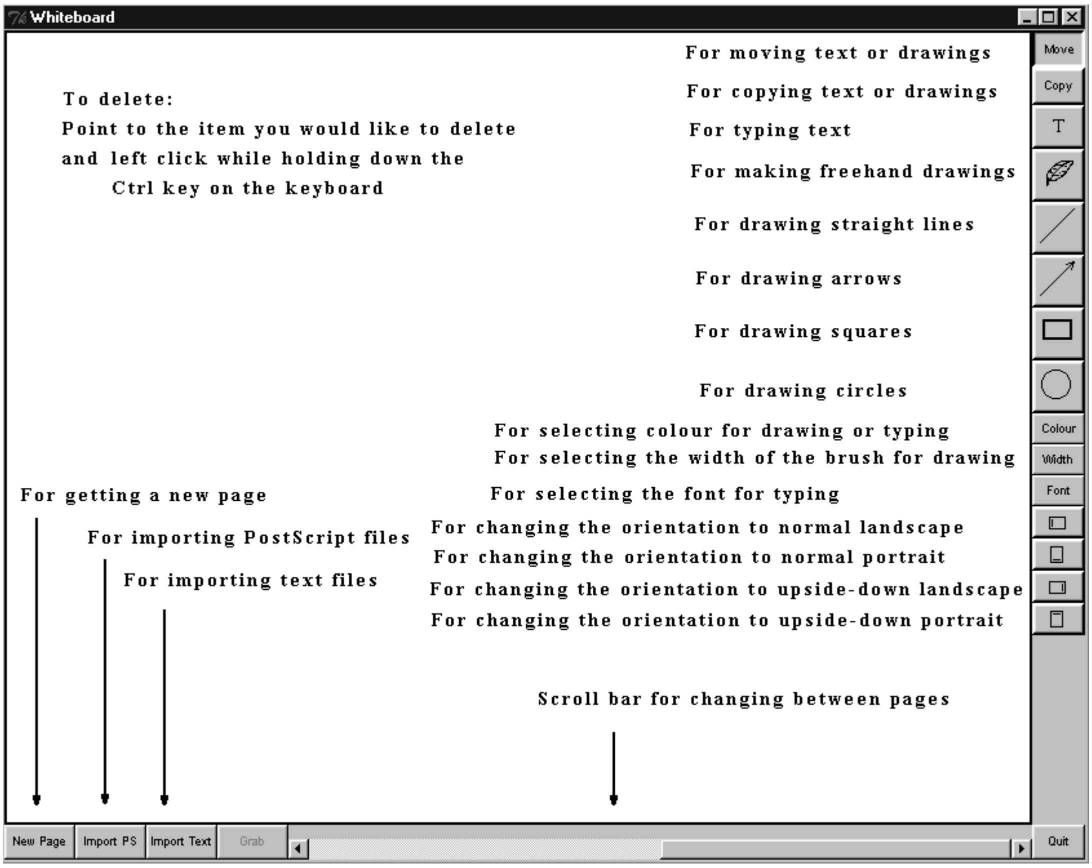

Figure 5: WBD - Shared Workspace Tool

4.4.3. NTE

If more flexible text editing is required and there is no need to display pictures, the NTE shared text editor may be a more suitable choice of workspace. Text is entered via the keyboard or by importing a text file. Unlike WB/WBD, any contribution can be modified by all the members of the group (though it is possible to override this with a locking mechanism).

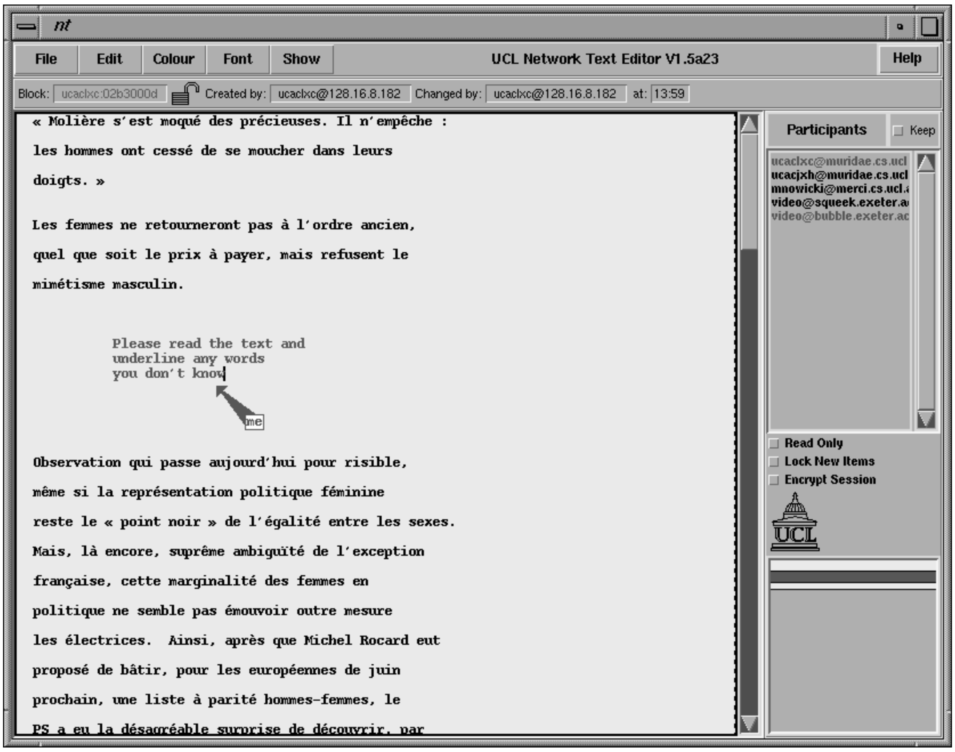

Figure 6: NTE - Shared Workspace Tool

The user can edit, move and point at all parts of a document and as they do so, their actions are seen by all other participants. Participants can select a different colour each to enable easy identification. The pointer shown in Figure 6 appears on the user's screen with the caption 'me', however on the other participants' screens, it appears with the user's name. NTE also allows users to save the text (particularly useful for minutes or when working on a long document over several sessions).

4.4.4. Integrated Interface

At the time of writing, the integrated interface is only available for Windows95/NT4.0. A typical conference will generally involve running the audio tool, the video tool and one of the shared workspace tools. Running three or four tools at the same time means having the screen crowded with windows which can be difficult to manage. An integrated interface has been developed to combat this problem. It integrates the audio tool (RAT), the video tool (VIC) and the shared workspace tools (NTE and WBD) into an easy to use interface which also hides most of the unwanted functionality.

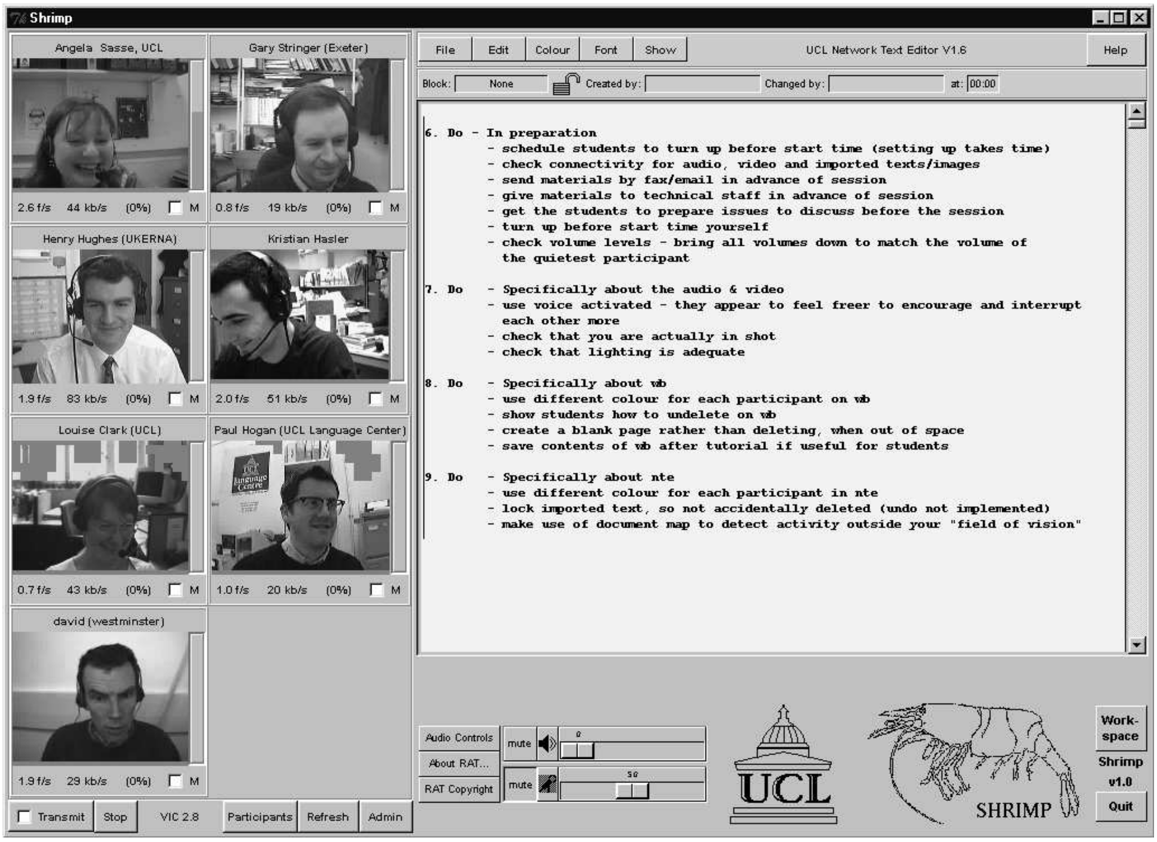

Figure 7: Integrated Interface

5. How to Get Started

5.1. What you need

The prerequisites may be categorised under the headings of justification, connectivity and equipment.

5.1.1. Justification

For many users the most important benefits may be hard to measure (an enhanced experience for students in distance education, or better co-ordination and co-operation in a working group, for example). These days, however, a financial justification is usually needed. This can be made in terms of saved travel expense where there is a need for regular meetings with collegues at remote locations. Multimedia conferencing cannot yet replace the need for *real* meetings, but it can reduce that need.

5.1.2. Connectivity

You and your prospective partners in conferencing will need a good connection to the Internet. The minimum recommended is 500 kbit/s, but it is dependent on the activities you intend to use the Mbone for. Mbone usage shares bandwidth with other Internet usage and the minimum requirement for connecting to the Mbone is therefore also dependent on this other Internet traffic. In particular, video transmission generates a significant network load, even with the efficiency offered by multicast, and you should consider your use in relation to the needs of others sharing the same connection. Most universities and colleges of higher education in the UK are connected to the Internet through JANET, the UK's academic and research network. SuperJANET is a broadband network which is capable of handling the increased amount of data that multicasting realtime audio and video on the network incurs.

5.1.3. Equipment

In order to run the software tools for taking part in videoconferencing, you will need a workstation or PC with one of the following specifications (suggestions given for Linux, Solaris and Windows95/NT4.0):

- A workstation running Solaris 2.5.1 or later with a SPARC processor (Solaris for Intel x86 processors not supported)
  - Sun audio hardware to send and receive audio.
  - In order to *transmit* video a framegrabber card is required - a SunVideo framegrabber is recommended (you do not need a framegrabber card in order to *receive* video).

- A PC running Debian GNU/Linux 1.3.x.
  (Other distributions such as RedHat, are not supported).

  Minimum requirements:-
  Pentium Processor (133MHz or above)
  16MB RAM
  15MB of free hard drive space
  Soundcard (see http://www-mice.cs.ucl.ac.uk/mice/rat/FAQ.html for information on soundcards for PCs)

  Preferred requirements:-
  Pentium P200MMX or better
  32MB RAM
  2D graphics card with 4MB RAM with DirectDraw support. The Matrox Millenium 2 is recommended
  17" colour monitor
  Soundcard (see http://www-mice.cs.ucl.ac.uk/mice/rat/FAQ.html for information on soundcards for PCs)

- BT848 based framegrabber, e.g. Hauppauge Win/TV Primio (optional, but needed to transmit video, though not necessary in order to receive it).
Colour video camera, e.g. Hauppauge video camera.

- A PC running Windows95/NT4.0.

Minimum requirements:-
Pentium Processor or equivalent(133MHz or above)
16MB RAM
15MB of free hard drive space
Soundcard (see http://www-mice.cs.ucl.ac.uk/mice/rat/FAQ.html for information on soundcards for PCs)
"Video for windows" compatible frame grabber card graphics card capable of a minimum resolution of 1024x768x16bit (colour palette)

Preferred requirements:-
Pentium P200MMX or better
32MB RAM
2D graphics card with 4MB RAM with DirectDraw support. The Matrox Millenium 2 is recommended
17" colour monitor
Soundcard (see http://www-mice.cs.ucl.ac.uk/mice/rat/FAQ.html for information on soundcards for PCs)
"Video for windows" compatible frame grabber card. The Hauppauge Win/TV Primio is recommended
Colour video camera. The Hauppauge video camera is recommended.

5.2. What you need to do

This section is only relevant for academic users in the UK.

The JANET Mbone service is operated by the JANET NOSC (Network Operations Services Centre) under the direction of UKERNA. Users wishing to connect to the Mbone should discuss the implications with their local system administrator. Applications for a site connection to the service should be made by the system administrator, and sent by e-mail to JANET Customer Service at *service@ukerna.ac.uk*. See the JANET Mbone Service Technical Guide (http://www.ja.net/documents/mbone3.4.ps) for more information on how to connect to the Mbone.

6. Further Help, Comments, Suggestions and Corrections

The JANET Videoconferencing Advisory Service offers help and support for *academic users in the UK*.

JANET Videoconferencing Advisory Service - VCAS
Telephone: 0191 222 6950
Fax: 0191 222 7696
E-mail: advice@video.ja.net
URL: http://www.video.ja.net/

Glossary

Bandwidth

A term used to show the amount of information flowing through a communications channel. Expressed in units of n $\times$1000 bits per second, kbit/s.

CODEC

COder/DECoder. Converts analogue signals to digital signals and vice versa. CODECs are used for converting analogue audio and video from microphone and camera respectively to digital signals to be sent over the network and converting the digital signals back to analogue signals at the receiver's end for the audio to be played out through speakers and the video images displayed on the screen.

Framegrabber

A device which captures video one frame at a time from an analogue video source.

Frames per second (fps)

The frame rate for video image transmission, measured in frames per second (fps). The higher the frame rate, the better the motion rendition of the video image. 30 fps is the standard necessary for full motion video. To date, most communication channels are not capable of transmitting large numbers of video streams concurrently. With compression and the use of other signal processing algorithms, a videoconferencing VIC between 2 - 10 kilobits per second (kbit/s) is perfectly acceptable.

IP address

IP (Internet Protocol) addresses are unique numerical identifiers for each networked host computer. The IP address is used in conferencing when the conference is point-to-point rather than multiparty.

ISDN

Integrated Services Digital Network.

JANET/SuperJANET

The UK academic and research computer network. SuperJANET is the broadband, highspeed portion of JANET.

Kilobits per second (kbit/s)

The kbit/s rate enables the participants to track how much bandwidth is being used during a session. For videoconferencing, the kbit/s should not be raised above 128 kbit/s, unless using a point-to-point conference or a Local Area Network (TTL=16). See TTL.

Loss Rate

The percentage of packet information lost during transmission.

Mbone

The multicast capable backbone of the Internet. It currently consists of a network of tunnels linking the islands of multicast capable sub-networks around the world.

MCU
Multipoint Control Unit.

MICE NSC
Multimedia Integrated Conferencing for Europe, National Support Centres. A project, funded by UKERNA, to encourage use of the Mbone multimedia conferencing tools piloted in the MICE project. The project ended on March 31st 1997.

Multicasting
Multicasting is sending audio, video etc. on the Internet in a way which ensures that anybody who is interested in receiving the information, *can* receive it, but only people who *are* interested will receive it. Think of it as being in between unicast (like most telephone calls - between two telephones only) and broadcast (TV - the signals are sent to you whether you want to watch or not).

Network congestion
A condition in an IP network where the amount of traffic injected into the network is too great for the routers to handle and some packets are discarded.

NOSC
Network Operations Services Centre.

NTE
Network Text Editor.

Protocol
A set of standards that govern the transfer of information between computers over a network or via telecommunications systems. To reduce errors, the computers at both ends of a communications link must follow the same protocol.

RAT
Robust Audio Tool.

SDR
Multicast Session Directory tool.

TTL
TTL stands for time to live, and determines how far multicast packages can travel over the Internet. The standard values are:
16 - Site
47 - UK
63 - Europe
127 - World

Unicast
A point to point connection between two specific machines.

WB
Whiteboard.

WB

Whiteboard (WB clone).

VIC

Video Conferencing tool, produced at the Lawrence Berkeley Laboratories in California and chosen as the video tool for the SHRIMP package.