Visicast Tech Review – WP4 Information

Achievements up to July

There are essentially two core areas of expertise underlying the Visicast Project – those being Graphics and Linguistics. These are represented by WPs 4 and 5 respectively. They feed their product into the WPs dealing with Applications – WPs 1, 2 and 3. For that reason there is an inevitable element of overlap between these "server" and "client" WPs, and lists of their achievements must be considered together.

The first step in WP4 was to establish good mutual understanding of the technologies held and developed by individual partners among the group as a whole. To that end, an early Workshop was held at the premises of Televirtual in Norwich to familiarise partners with motion capture technology and liase on technology transfer. In particular, there was an exchange of data files, animation systems and Virtual Human models between Televirtual and INT, to enable researchers from each organisation to explore common ground. Data and a 3D model were also supplied to IRT, enabling them to explore the issues involved in broadcast transmission of an Avatar system.

Shortly after this, a version of an Avatar player or visualiser which could be run from an HTML page was shown to the consortium, as an indication of what could eventually be achieved by way of a WWW signing viewer. At that stage, the viewer was not built in Direct X7 or Active X, and so could not actually run within the WWW page itself.

The system for capturing movements – hand, body and face gesture – comprising sign language is derived from Mask-VR, an in-house system developed by Televirtual to service its own requirements. Throughout this early period, and indeed continuing through the project, this system was and will be further developed. This is with three principle aspirations in mind. The first is to simplify its operation to enable its use by semi-skilled operatives (rather than the computer scientists who developed it). The second is to improve its robustness (of hardware, software and methodology) to permit its use in industrial settings such as TV studios or Outside Broadcast facilities. A third aspiration is to allow editing of gestures and the creation of new ones. Early work concentrated on streamlining the calibration procedures for setting up the system with a new signer.

The capture system previously used for the project was based on the hardware owned by Televirtual and which formed part of the company's Motion Capture facility. A new shadow system has been set up, based at the University of East Anglia. This will be used for project recording sessions with the various partners (IvD, Hamburg, UEA) requiring virtual sign sequences in their own languages. This has the added benefit of freeing up the Televirtual equipment, both for the company's commercial work and further system development within the project.

It is recognised by the Visicast partners that any eventual Motion Capture System capable of use within television or other industrial facilities will require less intrusive means of data acquisition than those offered by the current body suit, data glove and facial tracker technologies. As a first step, the company has commissioned general research from a partner outside Visicast which will review available technologies and research strands offering interesting motion capture possibilities for the future. ("Markerless Based Human Motion

Capture: A Survey", Joseph Bray, Vision and VR Group, Dept. of Systems Engineering, Brunel University)

Early on, it was decided to go for more subjectively "real" appearing Virtual Humans or Avatars. The earliest work had used VHs created from scratch by a computer graphic artist (Tessa 1). Now, a 3D Laser scan of the head of a young female was conducted, to be the basis for a new VH model.

As the findings of the first appraisal were fed back into the system, information from this exercise was used to refine the design and animation of the new VH (two versions of this character have been created – Tessa 2, wearing Post Office Uniform and badging, suitable for use in the WP3 application and Tessa 3 (also known as "Visia"), a similar character but wearing ordinary, plain clothes, intended for general use).

Within the last two months of the period, Televirtual developed an ActiveX based virtual signing visualiser. The ability to use this software in any number of PC applications makes it an ideal host for a signing player which can be configured for WWW and other multi-media applications – such as those envisaged in WPs 1,2 and 3. The Alpha version is now in its second internal release, a new renderer having been written by Televirtual to overcome problems revealed in the operating system which previously led to a software clash if a second instance of the viewer was instituted in one session. ActiveX controls are being integrated to create versions of the viewer which will allow control of the virtual camera used to view the signing, virtual lighting and Avatar position.

Early in the project, INT wrote programmes to convert between Mask-VR file formats and MPEG4 format, thus establishing the potential for development of a system to comply with that format should that be deemed necessary (see below). At the second Visicast Management Meeting, it was agreed that for the time being, INT and Televirtual should continue development of the Mask-VR and MPEG4 formats in parallel, allowing eventual comparison which would reveal the strengths and weaknesses of each approach. In the very short term, it was agreed that Televirtual would work with IRT to develop a broadcast system for the demonstrator to be shown later in the year.

The activity of the ARTEMIS Project Unit (APU) at INT, focuses on two aspects, both related to the integration of the new MPEG-4 technologies. The first aspect, referred to as *video related activities*, consists in releasing a MPEG-4 video codec allowing scenarios encoding and multiplexing into MPEG-4 streams, and streams decoding into video sequences. The second aspect of the INT work, referred to as *animation related activities*, deals with the conversion between the Mask-VR animation parameters provided by Televirtual and MPEG-4 animation parameters, namely Body Animation Parameters (BAPs) and Face Animation Parameters (FAPs). The conversion scheme to be developed has to ensure visually similar animations of the signing avatar. Within the ViSiCAST project structure, APU video and animation related activities are encapsulated into Work Packages 1 and 4.

Dealing with animation, APU work within the framework of WP 4.4 (Advanced MPEG-4 Animation) aims at assessing MPEG-4 animation as an alternative solution for Consortium requirements. The Synthetic and Natural Hybrid Coding (SNHC) part of the MPEG-4 standard requires two kinds of data to be available for animation purpose:

1. humanoid geometry and texture, referred to as Body Definition Parameters (BDP),

2. humanoid actions, defined by Body Animation Parameters (BAPs) and Face Animation Parameters (FAPs).

A first stage has therefore consisted in elaborating a procedure to convert an arbitrary humanoid model into a MPEG-4 compliant model. This has motivated the development of a user-friendly interface dedicated to this task. The conversion procedure has been validated on the *NewDan* avatar model provided by Televirtual.

Animation parameters, delivered by Televirtual motion capture system, are encoded according to the MaskVR proprietary format. A second stage has therefore consisted in developing a conversion scheme from MaskVR to MPEG-4 compliant representation, namely BAPs and FAPs. This conversion takes advantage of a low bit-rate compression technique for BAPs and FAPs, previously developed by APU within the framework of the MPEG-4 standard. The typical bit-rate for compressed FAPs is 2 kbps, bit-rates for BAPs ranging from 5 to 30 kbps depending on body motion complexity. (The Televirtual Mask-VR Broadcast Transmission system achieves a similar sub-30kpbs bandwidth) The conversion scheme is currently operational for the arm, forearm, hand and fingers. APU is currently completing the integration of the remaining body parts, and is working on face parameters conversion.

Up to now, two solutions have been retained by the Consortium regarding any final animation technique built around the MPEG4 protocols:

1. The first one consists in performing animation using the MaskVR software platform. This implies developing an inverse conversion procedure for the animation parameters after transmission and decoding. APU has just started feasibility studies of this operation
2. The second solution relies on a video-SNHC MPEG-4 player to be developed. A preliminary OpenGL version, running on Silicon Graphics platforms, has already been released. The implementation of the PC version is currently in progress.

Achievements from July to Review

The most recent work carried out by Televirtual and the UEA, has concentrated on producing from the core Mask-VR system a Broadcast sign generator and visualiser suitable for use in the WP1 application and, in particular, for the early demonstration of direct transmission of unmediated signs performed (live or recorded) by a real human signer. Work has also been done on the comporession of this broadcast format (See WP1) which has produced a bandwidth of less than 30 kbps – directly comparable to that claimed for MPEG4 systems (see above).The Broadcast system uses similar ActiveX components to that being developed for WWW and other applications.

At the heart of both systems is a set of COM components: IHOSTCOM takes a collection of motion files, together with each file's calibration data, and generates a BoneSet object. A BoneSet object provides open interface to:

- Retrieve the number of bones;
- Retrieve the value for a single or multiple bone(s);
- Set the value for a single or multiple bone(s);
- To read or redefine the bone hierarchy and
- To transform the bone values between either the global co-ordinate system or localised co-ordinate systems as defined by the bone hierarchy.

A BoneSet is not exclusive to the IHOSTCOM component. Any other program (written in any suitable COM aware language: C++, VB, Delphi, Python), can create and initialise a BoneSet object.

In the Broadcast system, the transmitter essentially uses the BoneSet as derived from IHOSTCOM to generate a compressed representation of the bones for each frame. The receiver reproduces the BoneSet from this compressed format and uses it to animate an avatar.

Animation of an avatar is performed by the following group of components:
Mesh: This describes the polygonal and textural data as well as the attachment of the mesh to the BoneSet. The data for the Mesh is acquired from Televirtual's Fast Motion v2 (FM2) file format.

MultiRenderer: Optimised OpenGL renderer suitable for rendering meshes within one or more windows.

In the case of the ActiveX control, the IHOSTCOM and rendering components are brought together in one package.

Componentisation has meant that other members of the consortium can easily reuse Televirtual's avatar technology from a variety of development platforms. For example, the ActiveX control exposes the BoneSet for the Avatar, enabling third party applications to essentially 'puppeteer' the avatar – this would for example, allow for a MPEG4 humanoid animation stream to drive the Avatar.

The following diagram illustrates the architectural overview presented here.