

VISICAST

SUMMARY OF ANNUAL PROGRESS

1 INTRODUCTION

The aim of this document is to describe briefly how the ViSiCAST project is organised in order to deliver the programme of work it has promised. The project has a relatively large number of partners, cooperating in a sizeable number of workpackages to produce an ambitious set of deliverables.

In order to be successful, ViSiCAST has structured the workpackages into a number of layers: user focussed; application focussed; and technology focussed. ViSiCAST has also structured the workplan into a number of phases: technology transfer and familiarisation; prototype applications; and advanced applications.

2 MANAGEMENT SYSTEM

Workpackage Leadership

ViSiCAST has 9 partners and is organised into 8 project workpackages. Each workpackage is led by a different ViSiCAST Partner. The exception is the ITC, the coordinating partner, which heads up both Management, and Exploitation and Dissemination.

Each workpackage involves a number of partners, so cooperation is essential. Workpackage leaders are responsible for negotiating a detailed plan for achieving the workpackage milestones and deliverables by breaking the work into subtasks that are the responsibility of a single partner.

Microsoft Project is used in a simple way to support project management. A copy of the plan has been provided, showing the relationship between published milestones and deliverables. For illustration, the more detailed plans for some of the workpackages have been integrated with the high-level plan.

Consortium and Workpackage Meetings

Each partner has a project manager who belongs to the ViSiCAST management team that meets in Consortium Meetings on a quarterly basis. Two days are set aside for consortium meetings, although management business will usually take less than one day, so that the rest of the time can be given over to technical work in workpackages. Workpackages arrange additional meetings as and when required.

Extensive use is made of electronic mail for regular communication and exchange of documents. Microsoft Office tools are used by default. A number of mailing lists are used to contact the project management team, or even all staff involved in the project. A consortium Website has been created with the project domain <http://www.visicast.co.uk>. The site is currently used to provide public information about the project and its partners. The intention is to develop the site to provide: demos of ViSiCAST deliverables; a means of archiving and exchanging project information; and facilities to assist project management, by tracking resource usage and progress against the plan.

3 WORKPACKAGE HIERARCHY

The ViSiCAST workpackages, with the exception of Management (WP7), are arranged in three layers.

User Focus

The first layer involves Trials and Evaluation (WP6) and Exploitation and Dissemination (WP8).

The Evaluation package is led by the RNID, which has immediate access to the deaf community and is charged with ensuring that the results of ViSiCAST meet the needs of deaf people and take account of their wishes. The package involves deaf people in the evaluation of all applications.

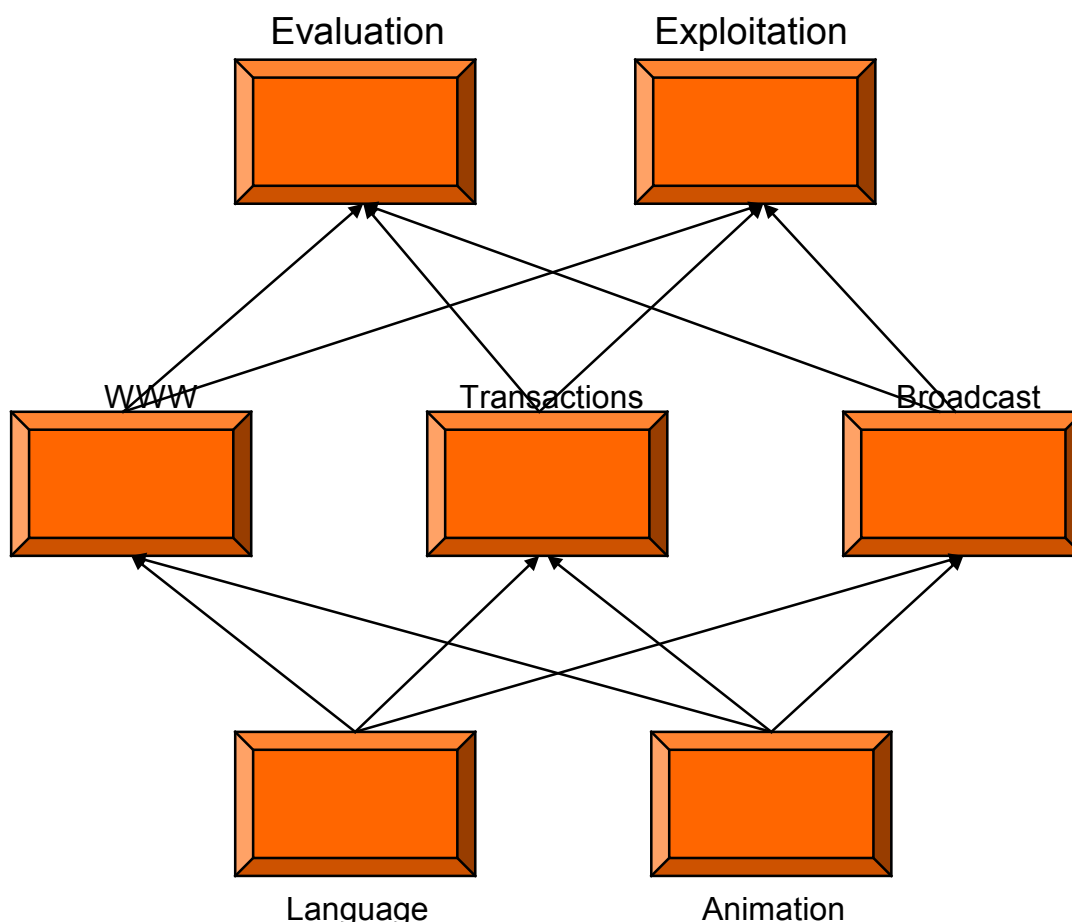
The Exploitation and Dissemination package is led by the ITC, that, with IRT and UKPO, has expertise concerning legislation and regulations relating to deaf people, as well as expertise in the business opportunities for ViSiCAST applications.

Application Focus

The second layer involves applications in three areas: TV and Broadcast Transmission (WP1), Multimedia and WWW Applications (WP2), and Face-to-Face Transactions (WP3). The requirements for these applications are set with reference to the needs of users and aim to provide support for deaf people through the medium of sign language in a range of areas.

The Broadcast package is led by IRT who have expertise in broadcast technology and links with TV companies.

The Multimedia package is led by the IvD who are experienced in providing deaf people with informational and educational materials that could benefit from the addition of signing.



The Face-to-Face package is led by the UK Post Office which manages a very large network of outlets with an obligation to accommodate people with a range of disabilities including deafness.

Technology Focus

The ViSiCAST workplan requires some challenging technical developments to be made. The structuring of the project means that the development of applications is driven by user need and developed with user feedback. In turn, the applications drive the development of the enabling technologies in the third layer.

The packages that focus on technology are: Animation and Modelling (WP4); and Language and Notation (WP5). ViSiCAST applications require high quality animation using 3D graphics. Since the aim is to bridge the hearing and deaf worlds, it is necessary to use advanced computational linguistics if the project is to deliver the natural sign languages which are the choice of the deaf community.

Animation and Modelling is led by Televirtual, a leading company in the area of Virtual Humans driven by motion captured data.

The Language and Notation package is led by the Institute of German Sign Language and Communication of the Deaf at Hamburg University. They are the developers of HamNoSys, a leading notation for cataloguing and analysing sign language.

4 WORKPLAN STRUCTURE

The ViSiCAST project involves some partners who have worked together in the past, but also involves partners bringing very different expertise to the project. In order to make useful progress during early stages of the project, the workplan has been structured so that there is an early phase where existing technology is developed and applied to produce useful applications, while at the same time developments are made in some of the enabling technologies. The new technology is used to develop prototype applications and in the process, the technology will be refined so that more advanced applications can be developed towards the end of the project.

The final section of this report describes each workpackage and picks up some of these themes.

Technology Transfer

In the first stage of the project, which lasts to month 9 or 12 of the project, according to workpackage, the emphasis is on straightforward development of existing technology. This enables partners to become familiar with each other's expertise and to develop patterns of collaborative working.

The animation technology used is an evolution of work from previous projects. Little advanced natural language processing is required because the domain in which the applications work is tightly controlled.

Prototype Applications

In the prototype applications we enhance the animation to handle a much wider range of signs than the fixed repertoire of existing systems. We also aim to add greater realism to the appearance and motion of avatars. The linguistics work aims both to handle less structured input (textual or vocal) and to produce more sophisticated natural signing. However, the systems will only work well in a known domain, and may need some manual intervention to provide faithful translation of spoken language to sign.

This phase starts within the first year of the project and ends at month 18 to 24.

Advanced Applications

The final stage of development will aim to steadily remove the limitations of the prototype phase. In the area of animation and modelling an additional dimension will be the aim of adding recognition of a limited range of signs so that true two-way communication becomes possible. The language processing will aim to provide a high degree of reliability with minimal need for manual intervention to clarify ambiguity or correct badly signed phrases

5 WORKPACKAGE PROGRESS

In this final section we provide information about progress in the ViSiCAST workpackages.

WP1 Television and Broadcast Transmission

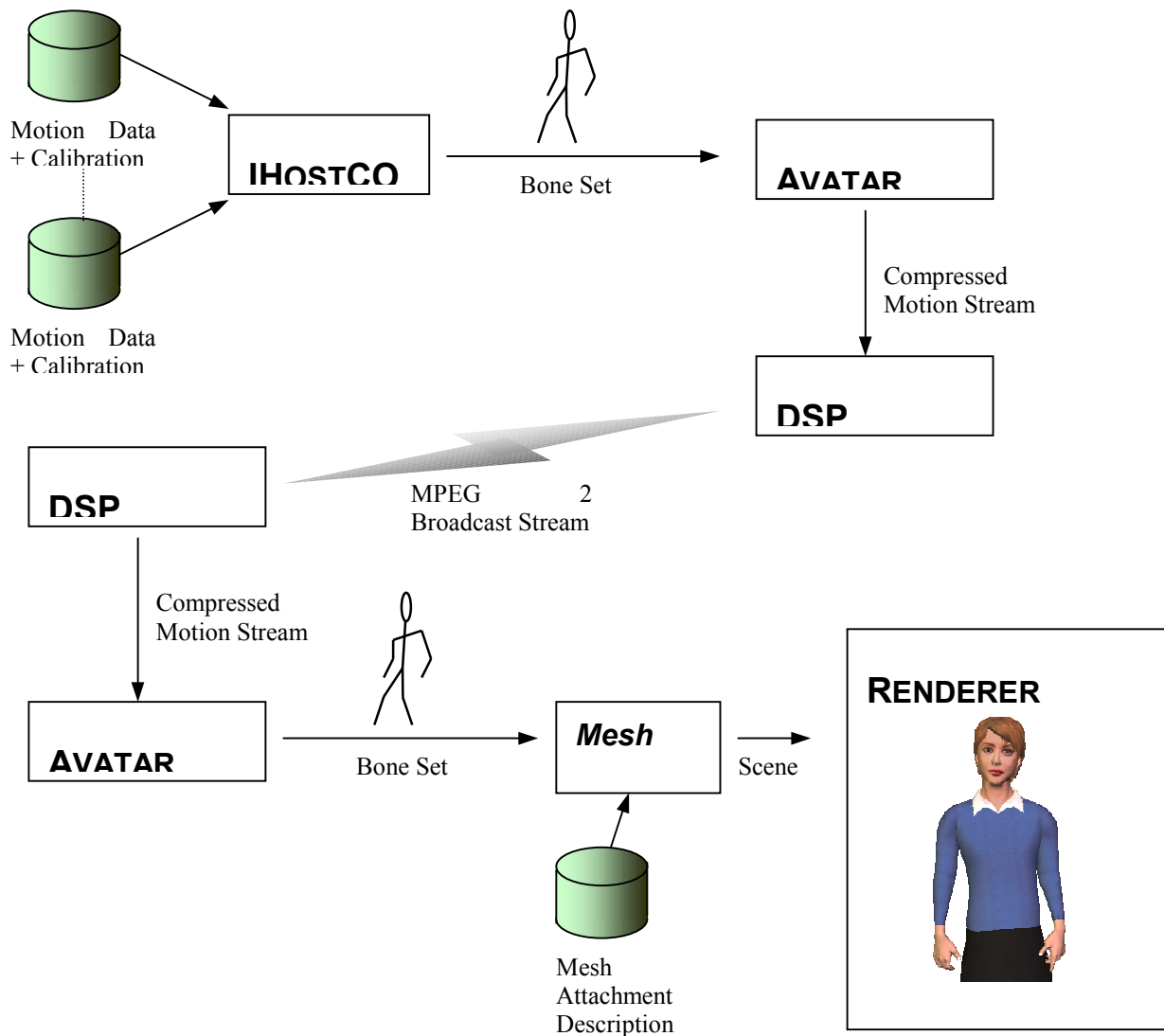
Progress against the Workplan

For the early deliverable in this WP (Direct TV Transmission of Virtual Human Signing), it was agreed by the consortium to build a system based on the Televirtual Mask-VR animation system. The adaptation of the core animation system necessary to permit this is described in the WP4 report.

In the Broadcast system, the transmitter essentially uses the BoneSet as derived from IHOSTCOM (see WP4) 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.

Televirtual, conducting the work within WP4, has also co-operated in early trials of this system, designed to mimic a Broadcast environment. This involved setting up a signing server PC – representing the TV studio or transmission end, and a signing client – representing the consumer and the STB (Set Top Box) likely to be used to decode the signing in the home. This system has been used to test the feasibility of a one-way data feed and its robustness in the face of a TX interruption.

The following diagram illustrates where the Broadcast transmission layer has been inserted in the Mask-VR pipeline.



Within the framework of Work Package 1, the ARTEMIS Project Unit (APU) at the Institut National des Télécommunications (INT), has released version 1 of a MPEG-4 video codec in June 2000. Preliminary results, including evaluation of coding performance and perceptual quality with respect to MPEG-2 compression, have been presented during on the 3rd General Work Package Meeting in Holland (28-30 June 2000). Developments have focused on coding efficiency adaptation with respect to motion amplitude. Specifically, two distinct codecs have been released, respectively adapted to encoding sequences with small and large motion amplitudes. In each case, the superiority of MPEG-4 technologies over MPEG-2 compression in term of bit-rate has been established: MPEG-4 compression of a MPEG-2 stream with high video quality yields bit-rate reduction with a factor varying from 2 to 9. Depending on scene features (in particular, motion activity and texture complexity), the resulting bit-rate for a PAL sequence with standard frame size and rate varies from 500kbps to 2.5Mbps.

APU is currently working on version 2 which will extend version 1 by integrating video object-based coding functionalities, including (i) separate (spatial and motion) encoding of individual moving video objects throughout the sequence, and (ii) coding adaptivity via an automatic partitioning mechanism into frame groups with globally high or low motion activity to be processed by the corresponding codec. Preliminary results obtained with version 2 indicate that a bit-rate reduction could be achieved while maintaining a constant video quality.

Moreover, for small frame dimensions, MPEG-4 compression proves to yield bit-rates compatible with MPEG-2 Supplementary Packets (SP) size requirements. Hence, the idea of encoding avatar animation as a *video* MPEG-4 stream and to embed it into MPEG-2 SP. This leads to propose two

additional solutions for TV broadcasting which maintain full compliance with the MPEG-2 standard, allowing therefore to take advantage of existing hardware/software equipment :

Broadcast video as a MPEG-2 stream, and embed avatar animation, encoded as a *video* MPEG-4 stream, as MPEG-2 SP.

Convert the original MPEG-2 video *and* avatar animation into two *video* MPEG-4 streams, and embed them into a single MPEG-2 stream as two MPEG-2 SP.

In both schemes, MPEG-2 is used as a transport mechanism that takes care of all system requirements (packetizing, multiplexing, synchronisation), avatar animation being combined with video data at the decoding stage. This requires only the insertion of a video MPEG-4 coder/decoder prior/after the MPEG-2 multiplexer/demultiplexer. Solution 2 is clearly superior to Solution 1. The main limitation of these schemes relies on the assumption of video encoding of the animated avatar, which considerably restrains scene composition possibilities at the decoder side.

WP1 is therefore examining a range of possible approaches to broadcasting realistic Virtual Humans, seeking both to exploit emerging standards, while retaining the best possible quality.

WP2 Multimedia and WWW Applications

Achievements to the end of July 2000

The WP2 work focused on developing an Internet browser plug-in (D2-1) for the (semi)automatic translation of text into sign. The achievements of the partners up to the end of July were:

- Selected weather forecast reports as a suitable application domain.
- Ensured the collaboration of the Royal Netherlands Meteorological Institute who provided a set of real-life weather forecast reports for each of the four seasons.
- Developed a model for formatting the information in these Dutch reports. This model is based on an analysis of the real-life reports and contains 248 'concepts' (words or phrases) and 20 'patterns' (corresponding to sentences).
- Translated these concepts into Sign Language of the Netherlands (SLN).
- Translated and adapted the weather forecast model into German Sign language (DGS).
- Defined a set of design specifications for the browser plug-in. The aim of these specifications was that the browser plug-in should be useful and useable for all deaf people, especially those who have difficulty in reading text and who do not yet use the Internet.

Achievements since July including future milestones and deliverables.

After the end of July, WP2 continued to focus on the development of the browser plug-in. The achievements were:

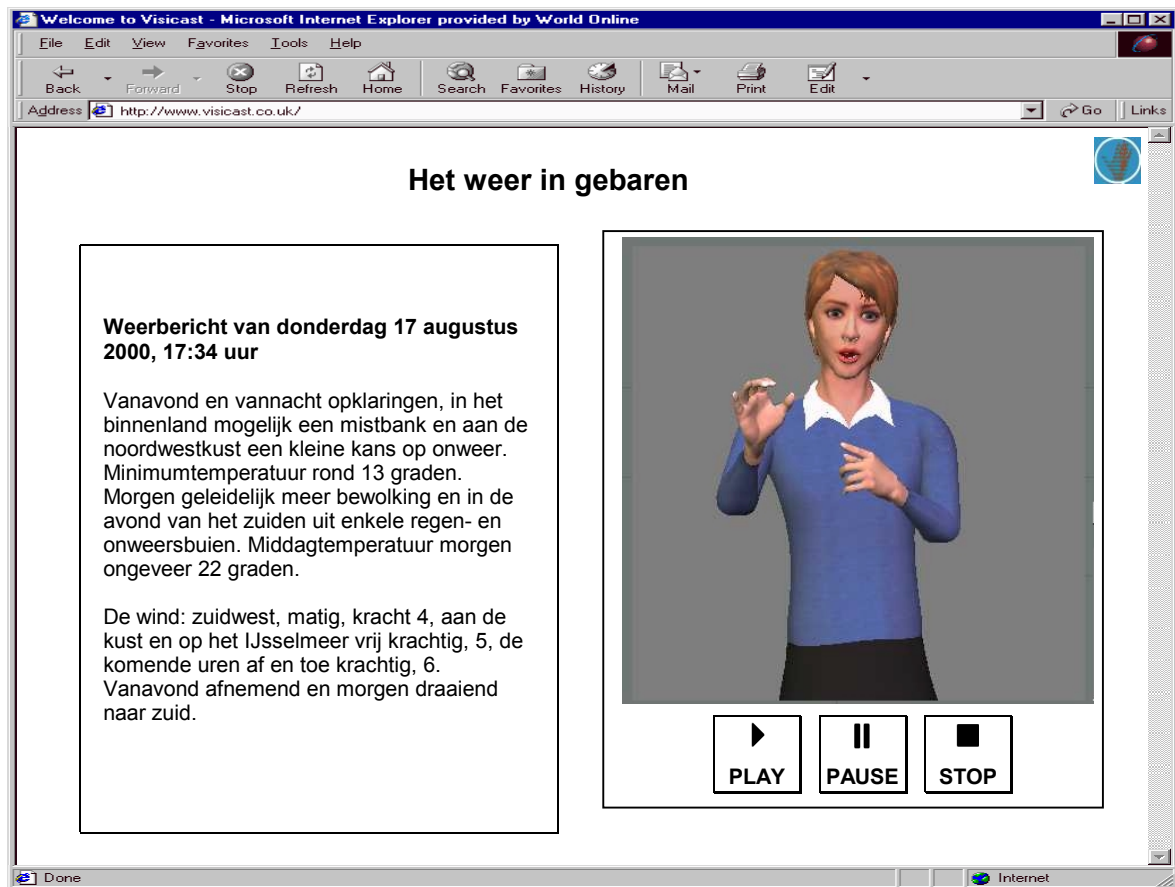
- Implemented the first version of the user interface (see example below).
- Scripting of the motion components needed for the translation of weather reports into DGS.
- Capturing of the motion components needed for the translation of weather reports into SLN.

Within the next months, motion capturing is planned for German Sign Language (DGS) and British Sign Language (BSL). In accordance with the project plan the browser plug-in (D2-1) will be completed at the end of year 1. It will provide (semi)-automatic translation of weather reports into SLN, DGS and BSL.

Progress against plan with any problems identified

For WP2, no delays have occurred or are expected. The only deviation from the Workplan was in the approach to developing the browser plug-in (D2-1). Because this is an early deliverable, it was initially planned to provide simple translation of arbitrary text into sign-supported language (i.e.

following the word order of spoken language). However, because feedback from experts on deaf communication indicated that this could diminish the acceptance of avatar-mediated sign language translation, it was decided to aim already early in the project for translation into real (i.e. grammatical) sign languages. This is also more consistent with the eventual aim of the project, which is translation into real sign language. The language domain chosen was that of weather reports, because it seems reasonable to assume that it will be of interest to most deaf people.



WP3 Face to Face Transactions

Objectives: To provide means for easier communication between deaf and hearing individuals in face to face transactions, such as post offices, banks, shops.

Deliverables to date:

Constrained PO system: Due month 7; Delivered month 7.

Achievements to end of July 2000

Work to date has entailed the production of a prototype system which recognises a series of spoken phrases in a limited domain (that of post office counter transactions), translates the phrase into sign language and signs them to the customer. (Executive summary below)

Recordings of transactions performed in several post offices at different locations around the country were analysed and from this data the most frequently used phrases identified. An automatic speech recognition system was developed to recognise this restricted set of phrases, with provision to insert variable quantities within the phrases where necessary. (eg days of the week and monetary amounts.) The recogniser was implemented such that it may easily be trained to the voice of an individual user

to provide high recognition accuracy despite the effects of background noise and multiple simultaneous speakers that are often found in the post office environment.

A database of motion captured data was recorded to enable the avatar developed in package 4a to sign the recognised phrases, and the avatar interfaced to the speech recognition system. The entire system was then evaluated in collaboration with the Post office and RND as detailed in WP6.

Progress towards next milestones

An investigation into the use of a less constrained recognition system, with free form speech input (still within the post office domain) being mapped to one of the pre recorded phrases. The NaturallySpeaking recognition engine from Dragon Systems is currently being evaluated as an alternative to the (now discontinued) Entropic recogniser. Methods for mapping from free form input to the signed phrases using techniques derived from telephone banking systems are also being investigated. No delays or deviations from the work plan have occurred or are expected with this workpackage.

ViSiCAST: Constrained System for Face-to-Face Communication in the Post Office

Executive Summary

A prototype system to enable a Post Office counter-clerk to communicate with a deaf or hearing-impaired customer using automatically-generated sign-language, and hence to aid completion of a transaction has been developed.

A priori, it might seem that recognising the clerk's speech and displaying it as text to the deaf customer would be an adequate aid to transactions. However, for many people who have been profoundly deaf from a young age, signing tends to be their first language and they learn to read and write more slowly. As a result, numbers of deaf people have below average reading abilities for English text. The system uses British Sign Language (BSL) rather than sign-supported English (SSE) as for the deaf community, SSE is unquestionably less popular than BSL.

Whereas systems to translate text from one spoken language to another are now readily available and work well within a restricted domain of discourse, translation from English text to a European sign-language is still a formidable research problem (this problem is being addressed within the ViSiCAST project). The approach taken to the translation problem in this system is to use pre-stored phrases and to pre-record the signs (as avatar movements) for these phrases. If only a small number of phrases is required, it is possible to record these in BSL. Phrases can be concatenated so that amounts of money can be inserted into a carrier phrase such as "The cost is...".

Although this approach imposes considerable restrictions on the meanings that can be conveyed by the PO clerk and hence on the dialogue, it has the advantage that BSL can be used. Furthermore, the limited nature of the transactions in a Post Office means that most transactions can be completed in this way. Using pre-stored phrases also confers benefits: the speech recognition is very accurate because of the limited number of vocabulary items to be recognised and one can also be sure that the meaning of a phrase uttered is accurately translated into the target language. These gains are important ones, as the "noise" introduced into the information channel by inaccuracies in the recognition process combined with ambiguities in the translation process can make more complex systems fail to translate correctly even simple phrases. By using pre-stored phrases, we in effect trade flexibility for accuracy. The system cannot currently recognise any signs from the customer, however this will be addressed in the third phase of the development of Tessa.

The PO clerk wears a headset microphone. The screen in front of the clerk displays a menu of topics available to him/her e.g. "Postage", "DVLA", "Bill Payments", "Passports". Speaking any of these words invokes another screen showing a list of phrases relevant to this category which can be recognised. However, this is only an "aide-memoire" to the clerk, and all phrases are active (i.e. can be recognised) at any time, so that switching between categories is seamless. Prior to designing the system, we obtained transcripts of recordings of PO transactions at three locations in the UK, in all 16 hours of business. Analysis of these transcriptions was essential for estimating the vocabulary that would be needed by our system to achieve a reasonable coverage of the most popular transactions. At the end of this analysis, we prepared a set of approximately 130 phrases that we estimated were adequate to cover about 90% of transactions.

The speech recogniser used was the Entropic HAPI (HTK Application Interface) system, which incorporates the HTK (Hidden Markov Model (HMM) Toolkit) recogniser [1,2,3]. A network of legal phrases is supplied to the recogniser, which uses a dictionary to decompose each word within a phrase into a sequence of triphones. Decoding of the speech signal is done using a Viterbi decoder that uses the speech models and the network supplied to output the most likely sequence of words given the acoustic input. The network constrains the speech recogniser to a finite number of predefined paths through the available vocabulary. These paths define the set of allowed phrases, and consist of a start node (usually denoting silence, or background noise), followed by a number of word nodes, or sub-networks, (that define, for instance, the legal ways of saying the integers between one and 100), finishing with an end node (again denoting silence). Sub-networks are useful ways of defining phrase segments which can vary. For instance, a sub-network called "one2hundred" represents the legal ways of and this can be inserted at any appropriate point into the network. There are other sub-networks called "amounts-of-money", "days-of-the-week". By constraining the grammar in this way, recognition accuracy is significantly improved over using a looser grammar. Also, because the recogniser operates on a "best-match" basis, a phrase which is phonetically "close" but not identical with a phrase in the network will be recognized as the latter, which confers some flexibility on the speech of the clerk. (For instance "Put that on the scales, please" would be recognised as "Please put it on the scales"). The network was constructed using a graphical network-building tool, graphHVite. This tool enables easy construction and editing of a network of phrases.

An important point about the operation of the recognition system is that both the speech models and the network can be varied. The speech models are adapted to the voice of each user using maximum-likelihood linear regression (MLLR) adaptation [4], a process which takes about twenty minutes, and the individual's models are then stored for later use. Speaker adaptation of the models greatly increases the recognition accuracy and hence the usability of the system.

Once the speech has been recognised, the correct sign sequence is selected by means of a lookup table which maps each spoken phrase to a motion capture file. For phrases with variable quantities such as 'days of the week' and monetary amounts, a series of files including signs for the appropriate day or value are played by the avatar, and the avatar seamlessly blends the various files together into a single sign sequence.

The motion-data stream is displayed using a virtual human. In common with many avatars, a three-dimensional "skeleton" is driven directly from the motion-data. The "skeleton" is wrapped in, and elastically attached to, a texture mapped three-dimensional polygon mesh that is controlled by a separate thread (event loop) that tracks the "skeleton". Currently we employ the RIVA TNT chip, by nVidia, to render the resulting 5000 polygons at 40 frames/s using Direct-X on a Pentium class PC. As a full three-dimensional model, the pose of Tessa can be changed on-the-fly by the user as can the identity of the virtual human and other characteristics. To give the best chance of creating smooth movements on a PC, Tessa was developed using DirectX. Tessa is capable of signing in real time with a refresh rate of 35 frames per second.

The system has been formally evaluated by several members of the deaf community, in collaboration with the RNID. Informal demonstrations have also been conducted as part of the Post Office's DDA awareness road shows.

Bibliography

[1] S. Young, J. Odell, D. Ollason, V. Valtchev and P. Woodland. *The HTK book*. Cambridge University Technical Services Ltd., 1997.

[2] J. Odell, D. Ollason, V. Valtchev and D. Whitehouse. *The HAPI book*. Entropic Cambridge Research Laboratory Ltd., 1997.

[3] S.J. Cox. Hidden Markov models for automatic speech recognition. *British Telecom technical journal*, 6(2):105-115, April 1988.

[4] C. Leggetter and P. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language*. 9(2):171-185, April 1995.

WP4 Animation and Modelling

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 liaise 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-30kbps 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 compression 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:

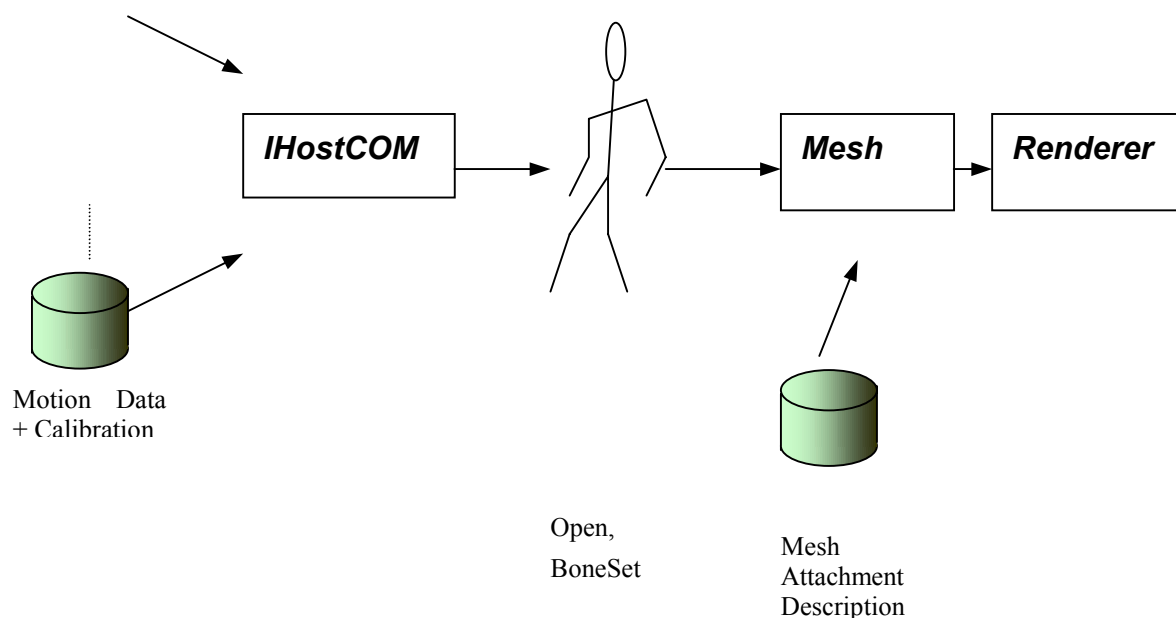
Mesh: This describes the polygonal and textural data as well as the attachment of the mesh to the BoneSet. The data for the Mesh is acquired from Televirtual's Fast Motion v2 (FM2) file format.

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.



WP5 Language and Notation

Achievements to the end of July 2000

The Semantic Interface Language has been chosen, namely, Discourse Representation Structure (DRS) with certain extensions (M5-1). Connections from the previously used CMU parser to DRS are being investigated.

The initial version of the 'Signing Gesture Markup Language' (SiGML) has been defined (M5-10).

SiGML-processing tools are under development, in particular, tools for HamNoSys-to-SiGML translation.

Demonstrator for streaming capabilities in initial SiGML has been developed.

A draft definition of HamNoSys refinements to be incorporated into SiGML has been written: internal evaluation and verification are ongoing.

The initial domain for natural language processing has been chosen based on several pilot studies.

Potential grammar development tools have been analysed in the light of the special needs of sign language grammar and phonology; communication has been initiated with the authors of the most promising systems.

Achievements since July including next milestones and deliverables

Work is continuing towards milestones M5-2, M5-3, and M5-4 to be integrated with the already finished M5-10 into D5-1.

Work is progressing on DRS generation from simple English constructs via the CMU parser.

Progress against plan with any problems identified

In order to compensate delays in recruitment at both UH and UEA, those milestones with immediate dependants have been given priority. This resulted in delays for milestones M5-2 and M5-3. As there was a lay time of 5 months for the output of M5-2 to be used, no corrective actions were needed for that milestone. Work on M5-3 was organised in a way that the extra-WP dependant (part of WP4) is not affected. No influence is seen on the next deliverables.

Work to date on SiGML development has highlighted the importance of the timely availability of a prototype SiGML synthesis tool (milestone M4-5), in order to support WP5's own internal and external evaluation requirements.

The most promising grammar development tool for ViSiCAST, LinGo from Stanford U, is still under development, with one component necessary for ViSiCAST usage not yet available. (The developers' view was recently revised so that the next version should become available within the next weeks.) It was decided by the workpackage members that we wait a maximum of two more months for the system to become usable before choosing the next-best solution. In the meantime, an ad-hoc grammar formalism will be used, with re-coding necessary when switching to LinGo (or any other HPSG-based environment). This extra effort seems justified by the potential gain of efficiency should LinGo become available with the required component.

WP6 Trials and Evaluation

Objectives: Qualitative and quantitative evaluation of transport mechanisms for and the underlying mechanisms of virtual signing.

Milestones to date:

Evaluation of Constrained PO system due month 9; delivered month 7.

Achievements to end of July 2000

Work to date has primarily entailed a formal evaluation of the Constrained Post Office system developed in WP 3 (executive summary attached). With UEA and the Post Office, formal evaluations were conducted in May. Six profoundly deaf people and three Post Office clerks took part. The evaluations indicated that there is scope for improvement of TESSA, gave some insight into how these improvements could be achieved and provided baseline outcome measures against which improvements could be assessed. Modifications are planned for all aspects highlighted as needing improvement, including implementation of an unconstrained version, where phrases need not be repeated word for word, which should enable much more natural communication. There is a need for community evaluations to assess the views of more deaf people and further evaluations of a modified,

unconstrained version of the system (scheduled for next year) to establish the ultimate potential benefits of TESSA.

Progress towards next milestones

Recruitment is in progress for a Community Evaluation Officer who will carry out community evaluations, set up focus groups and obtain general feedback from as many deaf people as possible in the UK. While there have been delays in this recruitment process, this is unlikely to affect completion of milestones as the majority of RNID evaluation work is scheduled for later in the project. It is anticipated that someone will be in post by November.

ViSiCAST: Evaluation of the constrained system for face-to-face communication in the Post Office - 10th July 2000

Executive summary

Evaluations are reported for a system – “TESSA” – developed to help sign-language communication for face-to-face transactions in the Post Office (PO). TESSA recognises what a clerk says, from a restricted list of phrases, and plays an appropriate pre-recorded phrase signed by an avatar on a screen. Six profoundly deaf people whose first language is British Sign Language (BSL) and three PO clerks took part. The main findings were:

- On average, 80% of the signs produced by the avatar and 61% of whole phrases were identified correctly.
- For ratings of ease of identification on a 5-point scale from 1-“Very difficult” to 5-“Very easy”, 79% of phrases were rated 3 or higher.
- For ratings of acceptability on a 3-point scale from 1-“Low” to 3-“High”, 63% of phrases were rated as 2 or 3.
- On average, the time taken to complete staged transactions was longer with TESSA than without, and the deaf participants, and to a lesser extent the clerks, rated communication with TESSA as more difficult and as less acceptable than without TESSA.
- Two of the six deaf participants said they would prefer to have the system available in the PO for use when communication became difficult. The other four said they would prefer to communicate without TESSA in its present form.
- The three deaf participants who usually experienced some worry or upset using the PO said communication with TESSA in the PO would not bother them at all.
- Aspects identified as needing improvement included facial expressions, clearer handshapes, finger configurations and lip patterns (especially for numbers and finger-spelling), the delay between spoken and signed phrases and a clearer distinction between face/hands and plain clothing.
- All clerks said they would prefer to have the system available as they thought it would make communication with deaf customers easier and more effective, though may take more time. Use of the system for multiple languages would ensure more frequent use and hence more likely use with deaf people.
- The clerks suggestions for improvement were primarily access to more phrases and an unconstrained system where phrases need not be spoken verbatim.

In conclusion:

The evaluations indicated that there is scope for improvement of TESSA, gave some insight into how these improvements could be achieved and provided baseline outcome measures against which improvements could be assessed. Modifications are planned for all aspects highlighted as needing improvement, including implementation of an unconstrained version, where phrases need not be repeated word for word, which should enable much more natural communication. There is a need for community evaluations to assess the views of more deaf people and further evaluations of a modified, unconstrained version of the system, eventually in a real PO setting, to establish the ultimate potential benefits of TESSA.

WP7 Project Management

The activity of this Workpackage has been covered above.

Jan Dobson, coordinator of ViSiCAST from the start of the project, moved on from ITC at the start of August 2000. Interim arrangements are in place to manage the project until a new project manager is recruited by ITC. Dr. Nick Lodge, who was fully involved in developing the project proposal, has resumed the role of project coordinator. UEA has freed up some time to allow Dr. John Glauert to provide support.

WP8 Exploitation and Dissemination

The ViSiCAST Marketing and Exploitation Plan (Deliverables D8-1 and D8-2) was produced to schedule in June 2000. The Periodic Reports refer to continuing interest in ViSiCAST, especially from the broadcast industry.

Promulgation of ViSiCAST will continue throughout the project, although most activity does not take place in this workpackage until after the next annual review.

