Virtual Signing: Capture, Animation, Storage and Transmission – an Overview of the ViSiCAST Project

JA Bangham, SJ Cox, R Elliott, JRW Glauert, I Marshall
University of East Anglia, Norwich
{ab,sjc,re,jrwg,im}@sys.uea.ac.uk
S Rankov, M Wells
TeleVirtual, Norwich
{sanja,mark}@televirtual.com

March, 2000

Abstract

We present an outline of the technical aspects of the recently started ViSiCAST project, which aims to provide deaf citizens with improved access to services and facilities through sign language. We first give an overview of the main application areas, before considering in more detail the two areas of supporting technology development: language and notation, and virtual human signing.

1 Introduction

ViSiCAST is a 3-year project (started in January 2000), funded as part of the EUs Framework V programme, which seeks to support improved access by deaf citizens to information and services in their chosen medium of sign language. ViSiCAST builds on experience gained in two earlier UK projects — described in a companion presentation — involving the project's UK partners, UEA Norwich, TeleVirtual Norwich, the Independent Television Commission and the Post Office, and concerned with the development of deaf signing systems using virtual humans, or avatars: Simon the Signer, and TESSA. To this foundation ViSiCAST brings the participation of leading European experts in broadcast technology (IRT, Germany), in Sign Language Notation (IDG, University of Hamburg), in broadcast imaging and animation standards such as MPEG and DVB (INT, France), and in multimedia content creation (IvD, Netherlands). Evaluation capabilities are provided to the project through the participation of the UK RNID.

At the core of the project are 'enabling technology' workpackages, concerned with the further development of the language technology and avatar technology established in earlier projects. These activities support workpackages covering three different application areas: Multimedia and WWW, Face-to-Face Transactions and Broadcasting. The following sections give brief overviews of the applications, before considering the enabling technologies themselves.

2 Multimedia and WWW Applications

This Workpackage develops applications of virtual signing aimed at the WWW, multimedia and third party software. A major objective here is that of bringing the virtual human signing into the WWW environment, to be achieved by the development of a browser plugin which incorporates the avatar software. This will provide a basis for the provision of signing services for Deaf users of the WWW. For example, using this plugin together with the GML notation described below, the author of a WWW-page will be able to provide signed material as part of the page's content. Alternatively, in the case of the many WWW-pages whose authors have made no explicit provision for deaf users, the plugin provides the basis for a simple translation facility applied to textual content in the page. In its most rudimentary form this would consist simply of finger spelling, but as both avatar and translation technologies mature, translation into natural sign languages in limited contexts becomes a possibility. Finally, we note that the use of such authoring tools is not restricted to the WWW: they may be used to support the development of any computer-based multimedia content.

3 Face-to-Face Transactions

This part of the project will develop applications of the avatar-based signing system in the context of face-to-face transactions, such as those occurring in post offices, health centres and hospitals, advice services, and shops. Development will take place in the specific context of a post office. The initial version of this system translates a limited repertoire of phrases used by a counter clerk into signing, using speech recognition technology in conjunction with the signing avatar. The first enhancement of this system will relax the constraints on the clerk. For example, rather than requiring the clerk to ask specifically "Do you want first or second class postage?", the system would accept a more natural alternative such as "First or second?"

A final version of this system will provide a basis for dialogue between customer and clerk, through the incorporation of available moving image recognition technology to 'read' simple signs made by the deaf customer, which can then be translated into text or speech for the benefit of the clerk.

4 Television & Broadcast Transmission

This workpackage concerns the provision of virtual human synthetic signing capabilities in the context of broadcast television, and has two related aspects: the development and integration of the necessary transmission technology, and the incorporation of ViSiCAST work into the relevant broadcast standards. The technology development includes work on strategies for the deployment of the virtual signer in television set-top boxes, on the transmission of signing represented in the GML notation (described below), and on techniques for low-bandwidth transmission of signing on broadcast television channels. The relevant standards in this area include the Multimedia Home Platform (MHP) within DVB, and the MPEG standards. For example, it is intended to incorporate the capture-based animation systems used by the project into the face and body animation systems defined by MPEG-4, and to integrate the GML notation into the *Multimedia Content Description Interface* framework of MPEG-7.

5 Language and Notation

5.1 Introduction

The language processing aspects of the ViSiCAST project have two main concerns:

- English text to European sign language presentation;
- definition of a ViSiCAST *Gesture Mark-up Language*, or GML, which is intended to be an XML-compliant representation of gestures used to link linguistic analysis work with animation technology.

5.2 English-to-Signing Translation

The initial work on the *Simon-the-Signer* avatar revealed that, though conversion of English text to an SEE signed presentation is feasible, this is not the preferred form in which the pre-lingually Deaf wish to see sign language.

Conversion of English text to natural sign languages has many problem characteristics of general machine translation systems which are typically capable of producing acceptable first draft translations which are completed by human translators or low quality translations acceptable if no alternative would exist.

Natural sign languages, such as BSL, have their own morphology, phonology and syntax [SSW99]. In the case of sign languages, 'phoneme' relates to the meaningful components of sign languages - hand shape, position, orientation and movement, rather than meaningful sound. At the morphological and syntactic levels, sign languages are inherently multi-modal, in the sense that facial expression and body posture contribute significantly to the meaning and ease with which signing is understood. Furthermore, the syntactic and discourse relationships in signing differ from those of English, and signers exploit the 3-dimensional space around the signer in order to communicate efficiently.

For example, BSL uses different types of verb forms, of which one is 'directional verbs', where information about the logical subject and indirect object of the verb is encoded as a pair of pronominal positions around the signer. Hence, 'I give X to you' involves the same hand-shape, but different starting and end positions of the giving motion from those of 'You give X to him'. First, second, and third person pronouns are conventionally located in signing space around the signer, and generated sign sequences should position referents appropriately in this space, so that subsequent references to these can be achieved via pointing and motion directed to and from those positions. It is often noted that BSL has a topic-comment structure, and that this is often reinforced by an organisation analagous to that of a rhetorical question. Hence,

The boy bought the book.

may be signed as:

BOY — BOUGHT — WHAT — BOOK

with a questioning facial expression up to the sign WHAT.

These aspects of sign language suggest an approach in which propositions and textual co-reference are treated as essential aspects, as is the case with Discourse Representation Structures (DRSs) [vEK97].

Some aspects of sign language, however, are under-specified within English. For example, in BSL, the sign for the pronoun 'we' indicates in signing space who is included within the designated group. In particular, the inherent ambiguity in the English use of the word 'we' as to whether the hearer is included or excluded from that group has to be resolved before it can be signed appropriately.

In view of the difficulties of the general problems of natural language understanding inherent within machine translation, ViSiCAST intends to adopt a semi-automatic mode of sign sequence preparation, in which natural language processing techniques are used in an interactive environment, which supports human intervention in the sign sequence preparation where necessary. Natural language processing techniques will be used to construct a DRS semantic representation of the English text, which will then be translated to a form which captures morphological rules of sign construction and semantic conventions used in a sign language such as BSL. Thus, the overall approach is to construct an interactive semi-automatic sign sequence preparation environment in which human interaction can enhance the quality of the outputs of natural language processing modules.

5.3 GML Notation for Signing

The *Gesture Markup Language*, or *GML*, is intended to provide a general framework for the representation and transmission of information about signing sequences, so as to allow signing expressed in this form to be used to drive the avatars described in the following section. The intention is that GML should fit into the framework defined by the XML family of standards currently under development by the W3C [Con00].

The foundation upon which GML will be constructed is the existing notation *HamNoSys* [PLZ+89]. HamNoSys is a well-established, general purpose, pictographic notation for deaf signing, developed by one of the ViSiCAST partners, the Institute for German Sign Language (IDG) at the University of Hamburg. HamNoSys is orientated towards the definition of individual gestures. Each gesture is defined by an initial hand configuration together with the movement to be performed starting from that configuration. The initial configuration in turn is defined by elements representing both form and orientation of the hands, and also, where necessary, the location of the hands relative to the signer's body. The notation also includes some facilities for the definition of non-manual signals, notably facial gestures such as raising the eyebrows, or pursing the lips. However, these are comparatively rudimentary, and it is intended to develop these further in the course of this project. HamNoSys deploys an alphabet of about 200 pictographic characters, for which a computer font and associated keyboard layouts have been defined.

It is envisaged that GML will incorporate information at (at least) the following four linguistic levels:

- glossing
- phonology
- phonetics
- articulation

as well as information covering such issues as signing speed and other temporal constraints. In terms of the above framework, HamNoSys itself operates primarily at the phonetic level.

Initially, and as a means of incorporating GML into our earlier work, GML will function primarily at a rather crude version of the glossing level, in which each 'gloss' element represents a sign, and as such can be used as an index into a lexicon of motion-captured signs, which in turn can be used to drive existing avatars. As appropriate techniques are developed for the synthesis of signs from more elementary components, it will become possible to use GML notation at the lower linguistic levels to drive the avatar.

6 Animation & Modelling

This part of the project is concerned with the development of avatar technology which is both more sophisticated and more flexible than that developed for our earlier projects. The facilities available from those projects are based on a high fidelity motion capture and calibration system, which allows the recording of signing performed by a human signer. The motion data thus captured can then be submitted to post-processing tools which allow it to be decomposed into individual recorded signs. These signs can be blended and 'played back'

through the medium of a digital virtual human, using mainstream computer rendering technology.

One basic way in which this system can be enhanced is to develop the visual quality of the avatar so that it attains the level of photo-realism. As the technology becomes more widely and more frequently used, it becomes desirable to provide users with recording and post-processing tools which permit more detailed control over the recorded content. However, a more fundamental requirement is the development of a more flexible sign generation system to support signing driven by the gesture-based lower levels of the GML notation described in the previous section. There are two interdependent aspects to this task:

- the identification of a repertoire of basic physical avatar features, or 'elements', representing both static configurations of body, face and hands, and motions of those same anatomical features;
- the development of methods of representing and combining these elements so as to generate the full range of realistic gestures which enable the avatar to act as a 'player' for the GML notation.

It is envisaged that, initially, the basic shape and motion elements identified above will be constructed both from real life, using the motion capture system, and using artificial models. Tools based on both these sources can then be developed and evaluated in parallel, prior to final deployment as appropriate.

7 Conclusion

It is clear that the technologies underlying this project will be developed for mainstream applications outside the context of deaf signing. As outlined above, ViSiCAST aims to ensure that these technologies are also utilised to provide improved access to information and services for the Deaf.

References

[Con00] D. Connolly. *Extensible Markup Language (XML)*. World Wide Web Consortium, 2000.

[PLZ+89] S. Prillwitz, R. Leven, H. Zienert, T. Hanke, J. Henning, et al. *Hamburg Notation System for Sign Languages — An Introductory Guide*. International Studies on Sign Language and the Communication of the Deaf, Volume 5. Institute of German Sign Language and Communication of the Deaf, University of Hamburg, 1989.

[SSW99] R. Sutton-Spence and B. Woll. *The Linguistics of British Sign Language*. CUP, 1999.

[vEK97] J. van Eijck and H. Kamp. Representing discourse in context. In J. van Benthem and A. Ter Meulen, editors, *Handbook of Logic and Language*. Elsevier, 1997.