The Development of Language Processing Support for the ViSiCAST Project

R Elliott, JRW Glauert, JR Kennaway, I Marshall
University of East Anglia, Norwich
{re,jrwg,jrk,im}@sys.uea.ac.uk *

ABSTRACT

ViSiCAST is a major new project funded by the European Union, aiming to provide improved access to services and facilities for deaf citizens through sign language presented by a virtual human, or avatar. We give here an outline of the project, and describe early work in the area of linguistics and language processing. This work covers two distinct but related areas: first, the development of an XML-compliant notation for deaf sign language gestures, which can be used to drive the signing avatar; and, second, the development of a framework supporting the translation of natural language text into this gesture-orientated notation.

Keywords

Sign Language, Virtual Signing, Language Processing

INTRODUCTION

Virtual Signing, Animation, Capture, Storage, and Transmission (ViSiCAST) is a 3-year project which started in January 2000, funded as part of the EU's 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 involving the project's UK partners, UEA Norwich, Televirtual Norwich, the Independent Television Commission and the Post Office. These projects were concerned with the development of deaf signing systems using virtual humans, or avatars: *Simon*, *Tessa* and *Visia*. To this foundation ViSiCAST brings the participation of leading European experts in broadcast technology (IRT, Germany), in Sign Language Notation (IDGS, University of Hamburg), in broadcast imaging and animation standards such as MPEG and DVB (INT, France), and in multimedia content creation for the deaf (IvD, Netherlands). Evaluation capabilities are provided to the project through the participation of the UK Royal National Institute for Deaf people, IDGS and IvD.

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.

In section 2 we briefly review the achievements of *Simon* and *Tessa* showing how they provide a foundation for ViSiCAST. In Section 3 we discuss improved approaches to the linguistics and language processing required for generating more natural sign language.

Section 4 outlines the way in which this work will be exploited within ViSiCAST. Section 5 concludes the paper.

BACKGROUND

Recent advances in multi-media technology have lead to an increased interest in virtual humans. Rendered off-line, they are regularly used in the entertainment industry. In addition standards are emerging for driving moving virtual humans over networks [1, 2, 3]. In particular, MPEG-4 (Version 2) provides two alternatives, "The Body" [4] and, through the adoption of VRML as a multimedia object, H-anim [5]. Such advances have excited interest in the use of virtual humans in delivering sign language presentations of stories and personal histories to children [2], and of language translation of text to American, Irish and Japanese sign languages [6]. To deliver readable sign language the virtual human has to present movements, gestures and expressions clearly and, if the presentation is to be acceptable reproduction of the original, the rendering has to be high quality with

*The work described here is supported by the European Union; previous work was supported by the ITC and the Post Office in the UK; we are grateful to colleagues in partner institutions for their support.

signs clearly identifiable.

Motion Capture and Replay

To achieve the fidelity appropriate for signing, the movements of a human signer are directly motion captured and are used to drive a virtual human [7, 8, 9], initially named *Simon* but more recently *Tessa* and *Visia*. Methods for capturing signing movements directly from video have been reported [10, 11, 12, 13, 14] but, although this is desirable, such approaches are not yet practical. The alternative is to capture the signs using individual sensors for the hands, body and face (see Figure 1).

Figure 1: Sign motion capture

Movements are motion captured using three methods. Cybergloves, with 18 resistive elements each, are used to record finger and thumb positions relative to the hand itself. Polhemus magnetic sensors record the wrist, upper arm, head and upper torso positions in three dimensional space relative to a magnetic field source. A video face tracker records facial expression. The face tracker consists of a helmet mounted camera with infrared filters, surrounded by infra-red light emitting diodes to illuminate (typically 18) Scotchlight reflectors positioned at regions of interest such as the mouth and eyebrows. The various sensors are sampled at between 30 and 60 Hz. These separate data streams are synthesised into a single raw motion-data stream, that can drive the virtual human directly.

In order to produce smooth movements on a PC, avatars have been developed capable of signing in real time with a refresh rate of 50 frames per second. A '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 PC accelerated 3D graphics cards to render the resulting 5000 polygons and a proprietary storage and transmission format developed by one of the partners (MaskVR). As a full three-dimensional model, the avatar's pose (see Figure 2) can be changed under program control.

Figure 2: *Visia* as a 3D model

Language Processing

Motion capture technology is used to generate data files of both individual signs and sign phrases. Although it can be used to motion capture prolonged signed 'texts', which are subsequently replayed, the prospect of providing access to information and services relies on using shorter sign sequences from which appropriate information and responses can be synthesised dynamically.

However, there are a number of different variants of sign language. These range from natural sign languages (British Sign Language - BSL - in the UK), which are the preferred languages of native (pre-lingually) Deaf signers, through to increasingly artificial forms of sign language such as Signed English (SE) which has often found favour within education as a means of conveying precise information about the local spoken language. A less extreme form is often used by signers who were not pre-lingually deaf and by inexperienced hearing signers which retains the word order of the spoken language but omits many functional words. In the UK this is known as Sign Supported English (SSE).

Natural Language Processing techniques have been investigated to assess the longer term prospects of providing access to textual information [7, 8, 9]. A provisional system (see Figure 3), constructed around the CMU link grammar parser [15], has been implemented to present textual information as SSE. The parser identifies functional words to omit from the signed presentation and resolves some ambiguities in English, such as the use of 'book' as a noun or as a verb which can then be signed differently.

Initial investigations, funded by the Independent Television Commission (ITC), were concerned with the feasibility of signing subtitle streams which accompany television programmes. The original aim was to consider the extent to which existing subtitle streams could be processed at the television receiver/set top box, providing the potential of access through signing to the large number of programmes which are subtitled. Although this approach held the prospect of Deaf viewers controlling the speed of signing, in a comparable way to which a hearing viewer controls the sound volume of sound, this highlighted the problem that SSE signing still has too much content to be signed even after functional words are omitted,

The parsed analysis of sentences from the subtitle stream were analysed to prioritise modifiers and phrases so that low priority elements could be identified. These could then be omitted from the signed presentation if the signing began to lag behind the television image. The individual signs were motion-captured from a sign language interpreter and looked up in a word to sign dictionary for signing. Interpolation between the end position of one sign and the starting position of the next sign achieved a smooth signing presentation. This work demonstrated the feasibility of using an avatar to present smooth signing and provided evidence of the small-scale readability of the signed presentations. Interpolation relies on the ability to manipulate motion capture data which would not be practical if signs were captured as video sequences.

However, difficulties in maintaining a reasonably timely synchronisation to the accompanying television image undermined the extent to which the overall story of the signed presentation was understood. As significantly, it became apparent that the focus on the simpler problem of signing using SSE was not addressing the problem of access for the most significant group who could benefit from the application of this technology. Hence, the focus of current work is to convert English textual information to national sign languages such as BSL.

Speech to Signing in the TESSA System

In co-operation with the UK Post Office (PO), the possibilities of increasing access to customer services through signing have also been explored. The TESSA system aims to aid communication between a PO counter clerk and a signing person by translating the clerk's speech to sign language and displaying it using an avatar. Dialogues of transactions in UK Post Offices were analysed and, from these, a library of frequent transaction sentences and phrases was constructed. RNID deaf signers then translated these sentences and phrases into BSL equivalents and an RNID deaf signer signed the sequences for motion capture [16].

The sentences and phrases spoken by the clerk are recognised by a speech recognition system which passes its output to the signing system. Each sentence or phrase is either complete in its own right, or has place holders which are filled with particular values. For example, the BSL form for the sentence *That will be X pounds Y pence* has place holders that are filled by signs for the numbers recognised from the speech input.

The system was evaluated in May of this year by six pre-lingually deaf people and three post office counter clerks. Feedback from the participants was encouraging and highly constructive - generally indicating areas in which the system could be improved. In particular, the deaf users emphasised the role of facial expression and clarity of handshape as significant aspects affecting comprehension. They also suggested that the system would be most useful for the more complex transactions in which it was difficult for the clerk to convey the correct information. Comments from the clerks mainly suggested that a more 'free form' input system in which they were not constrained to saying certain pre-defined phrases would be preferable, and that they would also like more phrases to be available. These points are being addressed in the next phase of the development of the system.

LANGUAGE AND NOTATION

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

- English text to European sign language presentation.
- Definition of a ViSiCAST *Signing Gesture Markup Language* (SiGML), which is intended to be an XML-compliant representation of gestures used to link linguistic analysis work with animation technology.

In this section we consider the above aspects in turn, together with the related sign synthesis issues.

English-to-Signing Translation

The response to initial work on the *Simon* avatar demonstrated 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. In addition, SSE presentations still exhibit significant natural language processing issues if the signed presentation has to be synchronised with a verbal commentary.

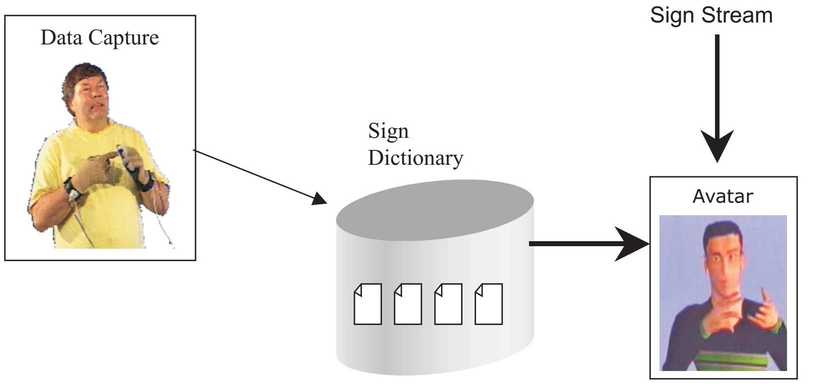

Figure 3: *Simon* System Dataflow

Natural sign languages, such as BSL, have their own morphology, phonology and syntax [17]. 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.

The approach undertaken in the English to Sign Language conversion is to translate English text to an intermediate representation based upon Discourse Representation Structures (DRSs) [18, 19], and then to convert the DRS representation to one of a number of national sign languages (see Figure 4). DRSs have been chosen as the intermediate representation due to the centrality of logical propositions and explicit representation of individuals and objects as arguments to predicates which capture anaphoric relationships.

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 handshape, 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. The DRS representation for these sentences are, respectively :

[X,Y : i(X) & you(Y) & give(X,Y)]

and

[X,Y : i(X) & you(Y) & give(Y,X)]

where the variables used in the propositions are explicitly enumerated. Synthesis to BSL can proceed by associating the *giving* hand-shape with the predicate *give* and motion is signing space between the positions associated with the predicates *you* and *i* dependent upon the argument ordering to the *give* predicate.

In addition, it is often noted that BSL has a topic-comment structure, and that this is often reinforced by an organisation analogous 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, or as

BOUGHT — BOOK — WHO — BOY

depending upon which aspect of the sentence is topicalised. The latter sentence can be characterised by a DRS

[X,Y : boy(X) & book(Y) & bought(X,Y) & comment(X)]

Furthermore, sign languages, such as BSL, allow the signer to locate discourse referents temporarily at one of a number of third person positions in signing space and to subsequently refer to that referent by pointing to its position. The compositional feature of DRSs allows DRSs for a sequence of sentences to be composed into a larger single DRS where the variables of the component DRSs are resolved to capture co-reference relationships. This feature of DRSs directly supports the idea of locating frequently used referents (variables) in signing space with subsequent use by pointing.

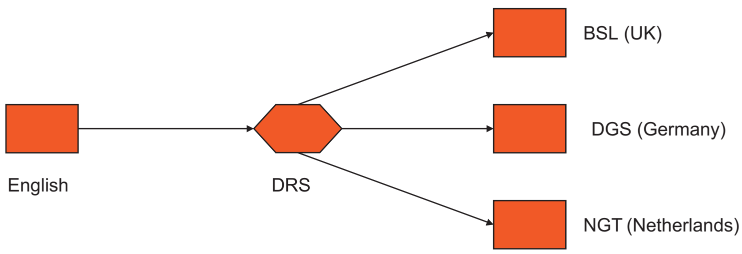

Figure 4: Translation to Sign Languages via DRSs

The conversion process is viewed as a semi-automatic translation with human intervention in order to enhance syntactic and semantic processing when automatic techniques are insufficient on their own. As indicated above, the intermediate representation is envisaged to be a form of DRS, since this representation has a number of characteristics which are oriented towards a serious treatment of sign language generation. In particular the following features of BSL are of note in this context:

1. Classifier handshapes are used as a refined form of pronominal references inside signs for verbs which incorporate subject/object information. In BSL, such 'proforms' are usually associated with information about verb object roles and can be incorporated into particular classes of verbs. Organisation of the back-end dictionary as a 'SignNet' analogous to WordNet [20] (a hierarchically structured lexicon of (American) English including information regarding synonyms, hypernyms etc) gives the potential of generating appropriate classifier handshapes with such verbs.

2. BSL signals temporal phenomena significantly differently from English. In particular, English tenses are not signaled in a comparable way in sign language. Hence, the DRS should be as accurate as possible with respect to temporal phenomena to allow conversion to sign language using, for example, appropriate time lines (such as BSLs three/four major time lines).

3. BSL makes a significant grammatical distinction between a single instance of a one-to-many relationship and many instances of the corresponding one-to-one relationship. For example, the ambiguity of

   > The lecturer speaks to the students.

   as either:

   > The lecturer speaks individually to each student (in turn).

   or:

   > The lecturer speaks to the students collectively.

   needs to be resolved in order to sign one of these alternatives as appropriate. DRSs allow the explicit representation both of a set and of its individual members.

Conversion of English text to natural sign languages exhibits many problem characteristics of general machine translation systems. These 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.

It is unrealistic to expect high-quality translation without some human intervention. Some aspects of sign language 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 addition, some anaphoric references are intimately related to common sense world knowledge which is beyond current language processing systems except in highly restricted domains. For example, the differing anaphoric relationships of the default readings of :

> John drove Bill to the railway station. He then returned home.

and

> John drove Bill to the railway station. He caught the 6.00 train.

requires significant understanding of expected patterns of human behaviour. It may be possible to determine from the surrounding context which of these is appropriate but it is more likely that human intervention will be required; however the DRS representation will have required that such ambiguity is resolved.

SiGML Notation for Signing

The *Signing Gesture Markup Language*, or *SiGML*, is intended to provide a general framework for the representation and transmission of information about signing sequences, so that signing expressed in this form can be used to drive the avatars described in the following section. The intention is that SiGML should fit into the framework defined by the XML family of standards currently under development by the W3C [21]. The other basis upon which SiGML is being developed is the existing notation *HamNoSys* [22]. 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 (IDGS) at the University of Hamburg. HamNoSys has been designed to support the definition of individual signing gestures. Its underlying gestural model and the associated notation are sufficiently general that HamNoSys is effectively capable of acting as a universal sign language notation: with very few exceptions, the signs of any of the many national sign languages can be expressed in HamNoSys. To represent its basic gestural elements 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, in its final form, SiGML will be capable of describing signing sequences at a number of linguistic levels, including:

- glossing (representing the meaning using an English word or phrase);
- phonology (explained in previous subsection);
- phonetics (explained in previous subsection);
- physical articulation (e.g. motion capture data).

The notation will also allow the expression of information covering such issues as signing speed and other temporal constraints. It should be appreciated that, from the standpoint of sign generation, these different linguistic levels can be seen as being in some measure alternatives to one another: for example, if complete articulation information is available for a particular sign, phonetic or gestural level information for that sign (such as may be specified in HamNoSys) would not be needed to drive the signing avatar, although its presence might be helpful and informative in other contexts.

Initially, and as a means of incorporating it into our earlier work, SiGML 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 our current signing avatars. However, as other parts of the project develop appropriate techniques for the synthesis of signs from more elementary components, the intention is that it should become possible to drive the signing avatar via SiGML notation at a lower linguistic level. The motivation here is to provide greater flexibility: instead of being restricted to the presentation of signing sequences for which the necessary signs have previously been motion captured, the signing avatar can present arbitrary signs as long as gestural level definitions of those signs have been provided. SiGML effectively provides this gestural, or phonetic, level of definition by transplanting the HamNoSys gestural model, with appropriate modifications, into an XML framework.

In the HamNoSys model, a typical gesture is defined by an initial hand configuration together with the action 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 in the "signing space" surrounding the signer's body. A variety of primitive actions may be specified, such as a movement of a hand to a new location, or a change of hand configuration. These primitive actions may be combined, both in temporal sequence and in parallel, so as to define more complex actions. All of the gestural features just identified may be specified for a single hand only, or for both hands individually, according to the requirements of the individual gesture. 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, in HamNoSys these are comparatively rudimentary, and it is intended to develop them further in SiGML.

XML [21] allows documents with a well-defined structure to be represented in a textual form which can be conveniently stored in a computer system and transmitted over computer networks such as the world-wide web. An XML document is given its structure by representing it as a sequence of nested *elements*, each of which is delimited by start and end *tags*. (A tag has the form `<...>`.) Any XML document may (but is not required to) specify its structure by means of a *Document Type Definition* (DTD): this is usually achieved simply by specifying a URL (network location) at which the full definition is to be found. The great advantage of providing a DTD is that it effectively specifies how the document is to be interpreted. Thus, returning to our present context, SiGML is effectively defined by means of its DTD, which specifies the allowable forms for valid signs.

Figure 5 shows a simple example of prototype SiGML, based on HamNoSys as described above. As may be seen, this is well-formed XML; indeed it is also valid, conforming as it does to the prototype SiGML DTD, `sigmlv0.dtd`, which is not explicitly shown. To keep this example to a manageable size, it contains the spec-

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE sigml SYSTEM "sigmlv0.dtd">
<sigml>
    <avatar url="Simon.ava" id="A" alt="Simon"/>
    <sign gloss="GOING-TO">
    <!-- Taken from Hamnosys 2.0 manual, p.42, top line. -->
        <hamnosys_sign lr_symm="parallel">
            <handposture
                handshapeclass="ham_finger2"
                thumbpos = "ham_thumb_outmod"
                extfidir="direction_uo"
                palmor="direction_l">
            </handposture>
            <par_movement>
                <straightmovement
                    direction="direction_o"
                    curve="direction_u"
                />
                <handposture extfidir="direction_do"/>
            </par_movement>
        </hamnosys_sign>
    </sign>
</sigml>
```

Figure 5: Prototype *Signing Gesture Markup Language* Example

ification of just a single sign, namely the DGS sign for *going-to*. This sign is defined in terms of the initial hand configuration and position (the <handposture> element), in conjunction with the required hand action starting from this configuration (the <par_movement> element). In this case, the action consists of a hand motion together with a simultaneous change of finger direction. The hand moves to a position "outwards", or away from, the body; but rather than moving in a straight line the hand traces an arc, "upwards" from, or above, the horizontal. Simultaneously with this motion, the direction of the index finger changes from "upwards and outwards" to "downwards and outwards".

Synthesis

The ViSiCAST project involves the development of its signing avatar technology base in a number of ways. In the present context, the most important of these is the development of a more flexible sign generation system to support signing driven by the gesture-based lower levels of the SiGML notation described above. 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 gestural level of the SiGML 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.

APPLICATIONS

The work described above provides the underlying technology for the development of applications in ViSiCAST, as follows.

A major application area is to bring virtual human signing into the WWW environment, through 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 SiGML notation described below, the author of a WWW-page will be able to provide signed material as part of the page's content.

The Tessa system will be enhanced in a number of ways, for use in other contexts, such as health centres and hospitals, advice services, and shops. The first such enhancement 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?”. The final enhancement 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.

ViSiCAST technology is intended to be suitable for incorporation in evolving broadcast standards. This includes work on strategies for the deployment of the virtual signer in television set-top boxes, and on the transmission of signing represented in the SiGML notation. 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 SiGML notation into the *Multimedia Content Description Interface* framework of MPEG-7.

CONCLUSION

We have presented early work on the translation of text to sign language. Existing natural language translation technology is being applied to generate signing through a new notation, Signing Gesture Mark-up Language. This notation will drive a new generation of high-quality Virtual Humans.

It is clear that the technologies we have mentioned are undergoing rapid development for the benefit of hearing users. The work now being undertaken in ViSiCAST promises to extend these technologies to support deaf signing and hence to provide Deaf citizens with access to services currently available only to the hearing.

References

[1] Seamless-Solutions-Inc. *Signing Avatars*. Seamless Solutions Inc., FLA, 1998. http://www.seamless-solutions.com/html.

[2] Vcom3D. *Signing Avatars*. Vcom3D., FLA, 2000. http://www.signingavatar.com/.

[3] C. Babski. *Baxter : Virtual Humanoids In VRML*. Swiss Federal Institute of Technology, 1998. http://ligwww.epfl.ch/˜babski/StandardBody/.

[4] R. Koenen. *Overview of the MPEG-4 Standard*. ISO/IEC JTC1/SC29/WG11 N2725, 1999.

[5] B. Roehl. *Specification for a Standard VRML Humanoid*. H-ANIM WG, U.Waterloo, Canada, 1998. http://ece.uwaterloo.ca/˜h-anim/spec.html.

[6] T. OVeale, A. Conway, and B. Collins. The challenges of cross-modal translation: English-to-sign-language translation in the Zardoz system. *Machine Translation*, 13:81–106, 1998.

[7] I. Marshall, F. Pezeshkpour, J.A. Bangham, M. Wells, and R. Hughes. On the real time elision of text. In *RIFRA 98 - Proc. Int. Workshop on Extraction, Filtering and Automatic Summarization, Tunisia*. CNRS, November 1998.

[8] F. Pezeshkpour, I. Marshall, R. Elliott, and J. A. Bangham. Development of a legible deaf-signing virtual human. In *Proc. IEEE Conf. Multi-Media, Florence*, volume 1, pages pp333–338, 1999.

[9] M. Wells, F. Pezeshkpour, I. Marshall, M. Tutt, and J. A. Bangham. Simon: an innovative approach to signing on television. In *Proc. Int. Broadcasting Convention*, 1999.

[10] J. Schlenzig, E. Hunter, and R. Jain. Recursive indentification of gesture inputers using HMMs. In *IEEE Proc. Second Int. Conf. Computer Vision*, pages 187–194, 1994.

[11] J. Yamato, J. Ohya, and K. Ishii. Recognising human actions in time-sequential images using HMMs. In *IEEE Proc. Second Int. Conf. Computer Vision*, pages 379–385, 1992.

[12] T. Starner and A. Pentland. Real-time American sign language recognition from video using HMMs. In *Motion Based Recognition*, pages 227–243, 1997.

[13] R. Bowden, T. A. Mitchell, and M. Sarhadi. Reconstructing 3d pose and motion from a single camera view. In *British Machine Vision Conference 1998*, volume 2, pages 904–913, 1998.

[14] Iain Matthews, J. Andrew Bangham, Richard Harvey, and Stephen Cox. A comparison of active shape model and scale decomposition based features for visual speech recognition. In *eccv*, pages 514–528, June 1998.

[15] D.D.K. Sleator and D. Temperley. Parsing with a link grammar. Technical Report CMU-CS-91-196, School of CS, Carnegie Mellon University, Pittsburgh PA, October 1991.

[16] M. Lincoln, S.J. Cox, and M. Nakisa. The development and evaluation of a speech to sign translation system to assist transactions. *Int. Journal of Human-computer Studies*, 2000. In Preparation.

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

[18] J. van Eijck and H. Kamp. *From Discourse to Logic (Vol. 1 and 2)*. Kluwer, 1993.

[19] 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.

[20] Princeton University. *WordNet*. Princeton University, NJ, 2000. http://www.cogsci.princeton.edu/ wn/.

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

[22] 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.