Signing for the deaf using virtual humans

JA Bangham, SJ Cox, M Lincoln, I Marshall
University of East Anglia, Norwich
{ab,sjc,ml,im}@sys.uea.ac.uk
M Tutt, M Wells
TeleVirtual, Norwich
{marcus,mark}@televirtual.com

March, 2000

Abstract

Research at Televirtual (Norwich) and the University of East Anglia, funded predominantly by the Independent Television Commission and more recently by the UK Post Office also, has investigated the feasibility of using virtual signing as a communication medium for presenting information to the Deaf.

We describe and demonstrate the underlying virtual signer technology, and discuss the language processing techniques and discourse models which have been investigated for information communication in a transaction application in Post Offices, and for presentation of more general textual material in texts such as subtitles accompanying television programmes.

1 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, 10]. In particular, MPEG-4 (Version 2) provides two alternatives, "The Body" [4] and, through the adoption of VRML as a multimedia object, H-anim [8]. 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 signs clearly identifiable.

2 Motion capture and replay.

To achieve a fidelity appropriate for signing we capture the movements of a human signer directly and couple these with a virtual human [5, 7, 13], initially named *Simon-the-Signer* . Methods for capturing signing movements directly from video have been reported [3, 6, 9, 12, 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).

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

Figure 1: Sign motion capture

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 infra-red 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 *Simon-the-Signer* was developed using DirectX [2]. It is 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 the RIVA TNT chip by nVidia to render the resulting 5000 polygons at 50 frames s^{-1} using Direct-X on a Pentium class PC. As a full three dimensional model, *Simon-the-Signer* 's pose (see Figure 2) can be changed on-the-fly under user control.

3 Language processing.

Motion capture technology is used to generate data files of both individual signs and sign phrases. It can, of course, be used to motion capture entire static signed 'texts', however the prospect of providing access to information and services relies on being

Figure 2: *Simon-the-Signer* as 3D model

able to use shorter sign sequences from which appropriate information and responses can be synthesised dynamically.

However, there are a number of different variants of sign languages, ranging from natural sign languages (such as British Sign Language - BSL), 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 information about the local spoken language. A less extreme form of the latter is often used by signers who were not pre-lingually deaf and by inexperiences hearing signers (in the UK this is known as Sign Supported English - SSE) which retains the word order of the spoken language but omits many functional words.

Natural Language Processing techniques have been investigated to assess the longer term prospects of providing access to textual information [5, 7, 13]. A provisional system, constructed around the CMU link grammar parser [11], 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.

Initially our investigations (funded by the Independent Television Commission - ITC) were concerned with the feasibility of signing subtitle streams which accompany television programmes. This raised the additional problem that SSE signing still has too much content to be signed even once functional words are omitted. It is also of note that the original aim of these investigations was to consider the extent to which existing subtitle streams could be processed at a television receiver/set top box thereby providing the potential of access through signing to the large number of programmes which are subtitled. In addition, this approach potentially held the prospect of Deaf viewers controlling the speed of signing in a comparable way to which a hearing viewer controls the volume of sound.

The parsed analysis of sentences from the subtitle stream were analysed to prioritise modifiers and phrases so that low priority information could be identified. These could then be omitted from the signed presentation if the signing began to lag behind the accompanying television images The individual signs were motion-captured from a sign language interpreter and looked up in a word to sign dictionary for signing. In-

terpolation between the end position of one sign and the starting position of the next sign achieves 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. However, difficulties in maintaining a reasonably timely synchronisation to the accompanying television images 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 in this area (reported later) is to convert English textual information to national sign languages such as BSL.

In co-operation with the UK Post Office (PO), we have also been exploring the possibilities of increasing access to customer services through signing. The TESSA system aims to aid communication of 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. Any of these sentences and phrases spoken by the clerk is 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 which are recognised from the speech input. This system is to undergo user-evaluation by six members of the Deaf community (in conjunction with the RNID) in May 2000. Initial reaction from the smaller group of deaf people who have seen the system has been encouraging.

4 Conclusions

This work has demonstrated that virtual humans can be constructed with sufficient fidelity to deliver legible signing. The work on sign sequence generation from textual sources indicates that the most benefit to the Deaf community is through a BSL signed presentation, and that the additional issues of language translation and understanding have to be addressed to generate quality BSL signing. The speech to signing work has indicated the potential usefulness of this technology in restricted discourse domains to widen access to services. These form these the starting point for a recent European Union funded project: ViSiCAST.

References

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

[2] B. Bergen and P. Donnelly. *Inside DirectX*. Microsoft Press, Redmond, WA, 1998.

[3] 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.

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

[5] I. Marshall, F. Pezeshkpour, J. 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.

[6] I. Matthews, J. A. Bangham, R. Harvey, and S. Cox. A comparison of active shape model and scale decomposition based features for visual speech recognition. In *eccv*, pages 514–528, June 1998.

[7] F. Pezeshkpour, I. Marshall, R. Eliott, 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.

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

[9] 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.

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

[11] D. 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.

[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] 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.

[14] 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.