Virtual Human Signing as Expressive Animation

John Glauert* Richard Kennaway* Ralph Elliott*
Barry-John Theobald*
*School of Computing Sciences
UEA Norwich, UK
{jrwg,re,jrk,bjt}@cmp.uea.ac.uk

Abstract

We present an overview of research at UEA into the animation of sign language using a gesture notation, outlining applications that have been developed and key aspects of the implementation. We argue that the requirements for virtual human signing involve the development of expressive characters. Although the principal focus of work has been on sign language, we believe that the work can be generalised easily and has a strong contribution to make to future research on expressive characters.

1 Signing Research at UEA

1 in 1000 people become deaf before they have acquired speech and may always have a low reading age for written English. Sign is their natural language. British Sign Language (BSL) has its own grammar and linguistic structure that is not based on English.

Sign language is expressive in its own right, and is multimodal, combining manual gestures, other bodily movements, and facial expressions. Facial information is especially important, conveying key semantic information. Just as intonation can affect the meaning of a sentence, for instance, turning a statement into a question, or indicating irony, so, facial gestures modify manual gestures in crucial ways. In addition, certain signs use the same manual content combined with different mouthings, often related to speech, to distinguish closely related concepts.

Research at UEA addresses the linguistics of sign language, where little is documented about grammar and semantics, and explores generation of signing, using gesture notation. We have developed SiGML (Elliott et al., 2001) (Signing Gesture Markup Language) for representing sign language utterances.

SiGML is used to generate realistic animation of signing using Virtual Human Avatars. The Animgen system (Kennaway, 2001) employs advanced techniques for skeletal animation to realise precise hand shapes and movements, leading to accurate bodily contacts. In addition, a range of facial gestures is animated by weighting morph targets giving appropriate displacements for facial mesh points.

In collaboration with Televirtual Ltd, a local multimedia company we have developed description formats for specifying avatars and their streams of animation parameters.

Complete systems have been produced allowing control of content animation using a range of avatars embedded in a range of applications including support on web pages. The framework integrates SiGML processing through Animgen, and supports a number of avatars developed separately at UEA and by Televirtual.

Early work was based on signing captured via motion sensors, using blending techniques to concatenate motion sequences. Further work is based on capture via video, especially for facial expressions, providing a basis for recognising signs from motion data.

2 Virtual Signing Applications

Since deaf people do not necessarily find information easy to absorb in text, their access to services is restricted, despite the requirements of recent legislation. There is little support for digital services in sign.

Recent projects by colleagues at UEA include Simon the Signer (Pezeshkpour et al., 1999), winner of two Royal Television Society Awards, and TESSA (Cox et al., 2002), winner of the top BCS IT Award, undertaken within the EU ViSiCAST project (ViSiCAST, 2000). Both Simon the Signer and TESSA (see Figure 1) used motion captured signs that are blended into sequences on demand.

2.1 Simon the Signer

Simon the Signer took words from a television subtitle stream and rendered a sequence of signs in Sign Supported English (SSE) to appear as an optional commentary on screen. SSE is widely used in education of deaf people, using a subset of BSL signs presented in English word order. Although technically successful, the use of SSE rather than true BSL was not fully accepted by the deaf community since it does not provide the required cultural richness.

There are obvious benefits for broadcasters if signing can be generated from an existing low-bandwidth data

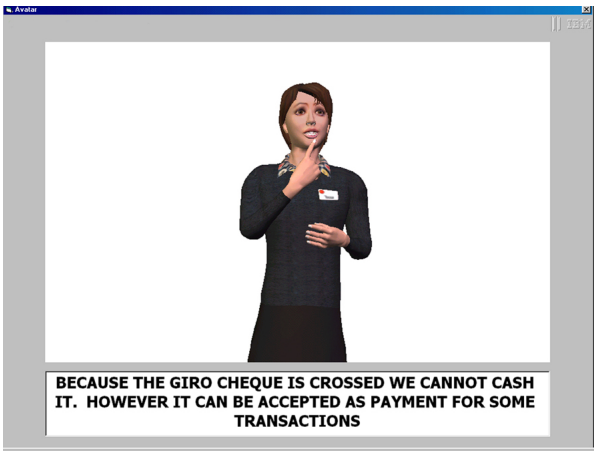

Figure 1: TESSA

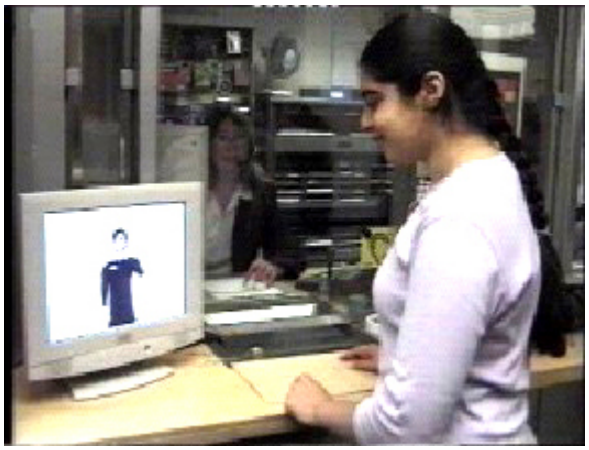

Figure 2: TESSA in use in a Post Office

stream such as subtitles. However, this seems a distant prospect. The use of video of sign interpreters is established. However, open captioned signing is not acceptable to hearing audiences and is only broadcast for a limited range of programmes at unsocial times. The bandwidth requirements are too high to broadcast a separate video stream for every channel that can be composited with the standard data streams in a set top box.

An alternative approach being explored in a current project is to capture the performance of a sign language interpreter, and transmit motion data parameters to drive an avatar in the set top box. Experiments show that the bandwidth requirement would be of the same order as for a speech channel.

2.2 TESSA

TESSA enables a Post Office clerk to communicate with a deaf customer by the use of speech recognition and avatar animation. Phrases used in standard transactions at Post Offices are recognised automatically and trigger animation of the corresponding sign language phrase in BSL. In addition to recognising fixed phrases, TESSA will handle phrases containing variable values such as days of the week or amounts of money and will substitute the corresponding BSL signs.

TESSA is an example of the use of an expressive character for communication. As the system is interactive and covers an extensive, though finite, domain, it provides genuine mediation between a hearing clerk and a deaf customer.

In addition to exploring applications in broadcast and to support face to face transactions, the ViSiCAST project also developed tools for providing low-bandwidth signing on the Web through a plugin for Internet Explorer.

2.3 Signed Weather Forecasts

Although the weather varies hour by hour, summary weather forecasts conform to a fixed pattern. The domain can be fully described for a number of natural spoken languages and natural sign languages. A system has been developed that enables forecasts to be presented by seamless blending of captured sign phrases using the web plugin.

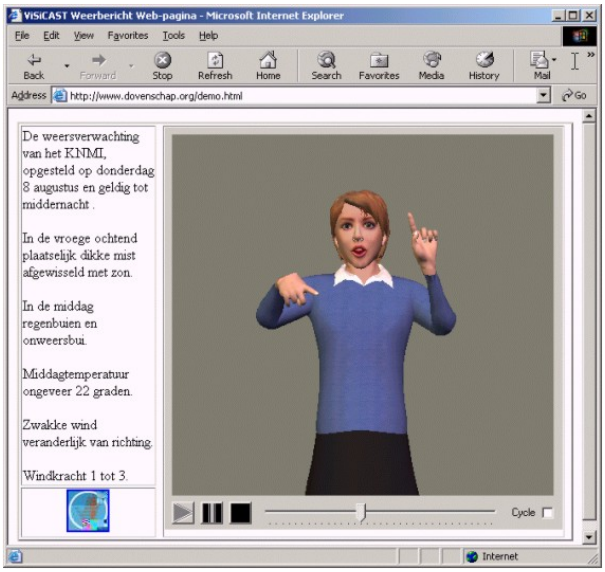

Figure 3: Weather Forecast on the Web

A tool has been developed which allows a non-signer to build forecasts, using standard weather phrases, for conversion into text and sign for a number of languages. Our implementation covers English, BSL, Dutch, SLN (Sign Language of the Netherlands) (see Figure 3), and DGS (German Sign Language). The Weather Forecast Creator, illustrated in Figure 4, may be used with a user interface in English, German, or Dutch and may be used to generate signing and text for all three countries. Hence it is not necessary for the content creator to know signing, or even the national language to be provided as text.

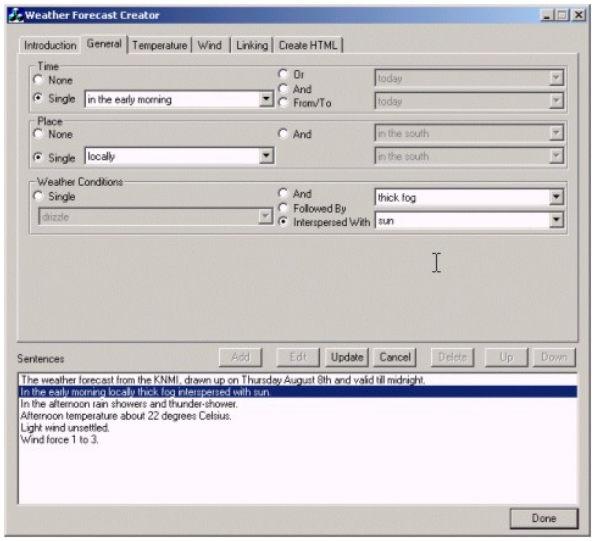

Figure 4: Weather Forecast Creator Application

Figure 6: German Website

2.4 Signing on eGovernment Websites

While earlier projects have been based on seamless concatenation of motion captured signs, the eSIGN project focuses on content created by synthesis from notation. As a result, information can easily be updated without the need for an expensive capture session. Information of an ephemeral nature can be generated automatically and interactively.

To enhance the usefulness of the internet for sign language users, the eSIGN project is developing signed commentary to accompany eGovernment forms. Figures 5 and 6 show web content under development.

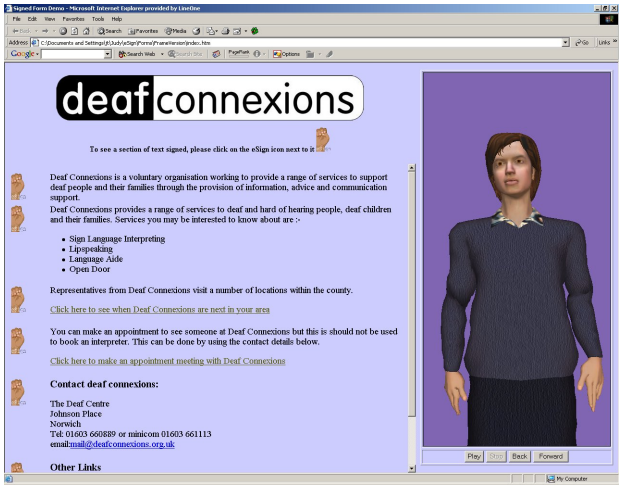

Figure 5: British Website

3 SiGML Notation

SiGML – Signing Gesture Markup Language (Elliott et al., 2001) – was initially developed in the ViSiCAST project as a key component of a prototype "natural-language-to-signed-animation" system developed in that project. Thus the primary purpose of SiGML is to support the definition of signing gestures in a manner allowing them to be animated in real-time using a computer-generated virtual human character, or avatar.

SiGML is an XML application language. We focus attention here on the major component of SiGML, referred to as "gestural" SiGML, which is used to drive the synthetic signing system. However, it should be noted that SiGML also allows the incorporation into the definition of a signed performance of data obtained by other means, including "motion capture" data, that is, motion parameters obtained by recording the actions of a human signer. Gestural SiGML is based on HamNoSys, the long-established Hamburg Notation System (Prillwitz et al., 1989) developed at the Institute for Deutsche Gebärdensprache (IDGS) at the University of Hamburg. Although it is based on HamNoSys, SiGML allows some physical features of the signer's posture to be specified with a greater degree of precision than HamNoSys. However, the semantic relation between the two notations is close: HamNoSys can be (and is) translated into SiGML; no significant information is lost in this process, and so any SiGML sign thus generated can generally be translated back into HamNoSys.

The purpose of HamNoSys is to support the transcription and analysis of human signing in a manner that is independent of the particular sign language used by the signer. Hence HamNoSys supports the transcription of signs at the phonetic level, providing special symbols and

structuring devices for representing phonetically significant features of signing such as hand shape and positions in “signing space”. HamNoSys effectively embodies a model of sign language phonetics, a model which is retained, largely unmodified, in SiGML.

A distinctive feature of sign language, in contrast with speech, is that is allows several distinct articulators to be in play concurrently. The most important articulators are the signer’s hands, but other bodily movements and various forms of facial gesture such as eye gaze and mouthing also have significance at the phonetic level in signing. Thus Gestural SiGML, like HamNoSys, has both a “manual” and a “non-manual” component. The manual component has a richer structure and is more fully specified than the non-manual, reflecting the fact that some non-manual aspects of signing, and their phonetic status, are less well-defined than are manual aspects.

3.1 Manual Signing

The manual component of SiGML allows a sign to be defined in terms of transitions between static postures, each of which may involve either or both of the signer’s hands. A hand posture is determined by the location of the hand in signing space, its shape, and its spatial orientation.

There is a core set of commonly occurring hand-shapes, such as “flat hand”,“fist” and“cee” (the shape of the hand when it is wrapped round a cylindrical object like a cup). A much larger repertoire of hand-shapes can be defined by applying modifications to these basic hand-shapes, for example, bending of individual fingers or the thumb, splaying of fingers, and various forms of contact between fingers. For two-handed signs, the notation allows precise specification of the relative configuration of the hands with respect to each other: this is achieved through the concept (taken from HamNoSys) of a “hand constellation”.

Various forms of hand motion may be specified: straight line, circular, or zig-zag. Each of these motions can be modified or refined in a wide variety of ways, of which the following is a small sample: a straight-line motion may be arced; the number of quarter-turns may be specified for a circular motion, whose radius may be varied dynamically to give a spiral effect; the plane in which a zig-zag movement is performed can be explicitly specified. Several more specialised forms of motion such as finger fluttering and wrist rotation are also supported. Another form of motion consists of a change of hand-shape or of hand-orientation. Motions may be combined in sequence and in parallel. There are modifiers which control the manner in which a motion is performed

3.2 Non-manual Signing

The definition of non-manual signing features in SiGML is based on the corresponding definitions for HamNoSys 4 (Hanke et al., 2000). A hierarchy of independent tiers, corresponds to distinct articulators. These may specify shoulder, body and head movements and eye gaze. Facial expressions control eye-brows, eye-lids, and nose. A repertoire of mouthings covers visemes for speech, along with other mouth gestures.

Here, “facial expression” refers to expressive uses of the face which form part of the linguistic performance, rather than those which communicate the signer’s attitude or emotional response.

3.3 SiGML Examples

```
<sigml>
<hamgestural_sign gloss="film">
  <sign_manual>
    <split_handconfig>
      <handconfig handshape="flat" extfidir="u"
            palmor="d"/>
      <handconfig handshape="finger2" thumbpos="across"
            extfidir="r" palmor="r"/>
    </split_handconfig>
    <split_location>
      <location_hand digits="2" contact="touch"/>
      <location_hand location="wristback"
            side="palmar" contact="touch"/>
    </split_location>
    <wristmotion motion="swinging"/>
  </sign_manual>
</hamgestural_sign>
</sigml>
```

Figure 7: SiGML for BSL sign “film”

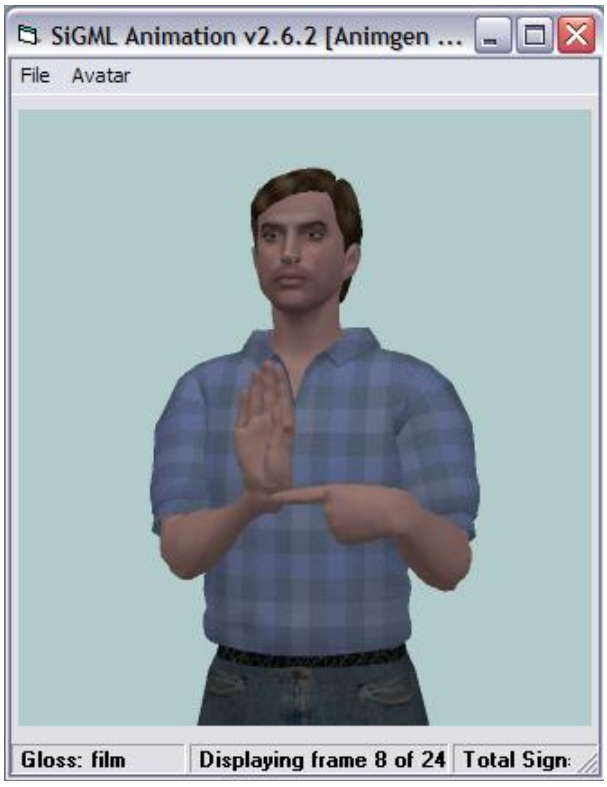

Figure 8: Initial configuration for BSL “film” sign

Figures 7 and 9 show two examples of individual SiGML signs. The first of these is the SiGML definition for the sign "film" in British Sign Language (BSL). This sign has a manual component but no non-manual component. Most of the former is devoted to the definition of the sign's initial configuration, shown in Figure 8. Both hands are involved in this configuration, and so for both hands there are specifications of their shape and orientation, followed by a specification of their locations with respect to each other: here the back of the wrist of the dominant hand is in contact with the extended index finger of the non-dominant hand. The motion for this sign is expressed comparatively succinctly in the `<wristmotion ...>` element, which specifies a swinging motion of the dominant (raised) hand.

```
<sigml>
<hamgestural_sign gloss="tell_story">
  <sign_manual both_hands="true"
               lr_symm="true" outofphase="true">
    <handconfig handshape="flat" thumbpos="out"/>
    <handconfig extfidir="ul"/>
    <handconfig palmor="ul"/>
    <handconstellation contact="close">
      <location_hand location="tip" digits="3"/>
      <location_hand location="palm"/>
      <location_bodyarm location="shoulders"/>
    </handconstellation>
    <rpt_motion repetition="fromstart">
      <tgt_motion>
        <circularmotion axis="l"/>
        <handconstellation contact="close">
          <location_hand location="tip" digits="3"/>
          <location_hand location="palm"/>
        </handconstellation>
      </tgt_motion>
    </rpt_motion>
  </sign_manual>
</hamgestural_sign>
</sigml>
```

Figure 9: SiGML for BSL Sign "Tell-story"

Figure 9 shows the SiGML definition for the BSL sign "tell-(the-)story". Here, in contrast to the previous example, both hands are involved not only in the sign's initial configuration (shown in Figure 10), but also in the subsequent motion. The attributes in the main `<sign_manual ...>` element specify that both hands participate in the motion, that there is left-right symmetry in this motion, and that the motions of the two hands are to be performed out-of-phase with each other. In this sign the "location" part of the intitial configuration consists of a hand-constellation which specifies the precise configuration of the hands with respect to each other, as well as the position in signing space of the two hands thus combined. Given the symmetry characteristics already specified for the two-handed motion as described above, an explicit movement specification is required for the dominant hand only. In this case, the required motion has significant internal structure explicitly defined in the notation: the motion has an explicit target, and is repeated once from its starting configuration.

Recently, as described in (Elliott et al., 2004), we have been developing support in our synthetic animation framework for the non-manual features of SiGML. The SiGML example shown in Figure 11, illustrates the fact that the system described there can be adapted to synthesise expressions of emotion, as well as the linguistically significant facial expressions required for signing. Figure 12 shows a pair of frames from the resulting animation: in the first of these frames the avatar's face is still in its neutral posture, in the second it has reached a much more expressive one.

Figure 10: Frame from BSL "tell-story" sign

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE sigml SYSTEM
  "http://www.visicast.cmp.uea.ac.uk/sigml/sigml.dtd">
<sigml>
<hamgestural_sign gloss="vguido-sad">
  <sign_nonmanual>
    <extra_tier>
      <extra_par>
        <extra_movement movement="X14"/>
        <extra_movement movement="X15"/>
        <extra_movement movement="X24"/>
        <extra_movement movement="X38"/>
        <extra_movement movement="X41"/>
      </extra_par>
    </extra_tier>
  </sign_nonmanual>
  <sign_manual/>
</hamgestural_sign>
</sigml>
```

Figure 11: SiGML for a tearful facial expression

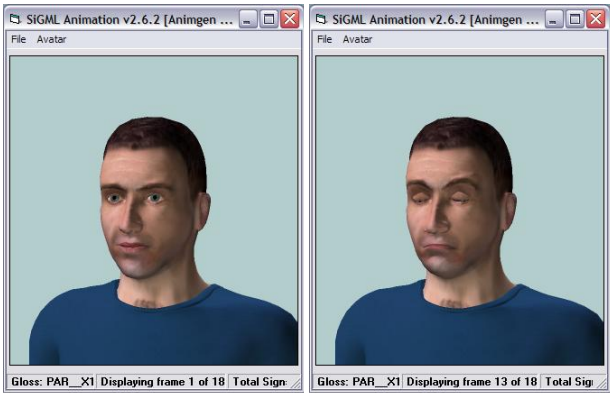

Figure 12: Two frames from the tearful facial animation

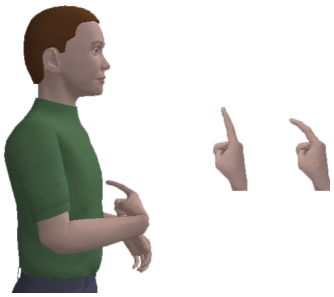

Figure 13: Pointing inwards

4 Animation of Signing

The ViSiCAST and eSign projects have achieved significant results in creating signing animations from HamNoSys (Kennaway, 2001, 2003). Although this notation was originally developed for researchers into signing to communicate with each other about signs, it has proved a suitable basis for synthetic animation.

4.1 Manual Animation

To translate the human-meaningful notations of HamNoSys into numerical animation data, several problems must be solved. The fuzzy categories of HamNoSys—the "chest", a "large" movement, a "close" proximity, a "fist" handshape etc.—must be replaced by numerical locations, distances and joint rotations. The locations named by HamNoSys, of which there are a few hundred, must be provided as part of the definition of the avatar. In general there is no way to automatically determine these locations by calculation from the surface mesh or the animation bones (which do not always closely correspond to the physiological bones). Given these, the various sizes of movements and proximities can be defined as multiples of measurements of the avatar. For example, we define near and far distances from the torso as a certain proportion of the length of the arms, since in signing these are primarily used as locations at which to place the hands.

HamNoSys transcriptions often leave out information which is obvious to the human reader and writer of the notation. Sometimes it simply takes a standard default value: absence of any explicit location for a sign means that it happens in the middle of the signing space. Sometimes it is dependent on the context: when the hands touch each other, the location at which they touch is often not specified.

HamNoSys specifies various types of repetition: repeating a movement from its starting position, repeating a movement from where its previous occurrence finished, repeating it several times getting larger and larger, etc. Various modes of repetition can be combined, and it can be quite complicated to determine exactly what an arbitrary repetition specification really means.

There are other features of the posture and movement which HamNoSys does not record at all. For example, it mostly describes what the hands do; what the rest of the arm must do to place the hands in the positions specified is not described. The animation software must be programmed with rules to decide how high the elbows are raised, and whether the collarbone joint moves. Physical objects are prevented by their nature from penetrating each other. The avatar's body parts are under no such constraint, except for whatever has been explicitly programmed.

To synthesise a lifelike movement from one posture to another, we use a semi-abstract biocontrol model to determine the accelerations and decelerations, parameterised in a way that lets us animate the various manners of movement which HamNoSys can specify: fast, slow, tense, with a sudden stop, etc.

Sometimes, the simplest way to resolve the problem of what a given piece of HamNoSys means is to make it more detailed, explicitly specifying information that is impractical to calculate: for example, specifying which points on the hands are in contact instead of merely saying that the hands contact each other. We are currently moving towards a version of SiGML that will allow the specification of more detail of this sort, and thus separate the problem of filling in the missing information from that of animating the gesture. This allows the trade-off between the effort of the transcriber and the effort of the animator to be made in different ways.

In some instances, HamNoSys transcriptions have been found to be incorrect, even when made by experienced users of the notation. There is a tendency for people to write down not the actual motion, but an idea of the motion that sometimes does not closely match it. An example occurring frequently in the HamNoSys corpus of over 3,000 signs of German Sign Language is that of an inward pointing finger (see Figure 13).

Often, the hand shape has been transcribed as if it were the first of the two hands in that figure, with the wrist bent sharply so as to point the whole hand at the signer. In reality, the hand shape will be more like the second shape

shown, with the fingertip pointing towards the signer, and the back of the hand pointing in a direction above, left, and behind the signer. This is perhaps an indication that a signing notation should transcribe these signs by recording the direction in which the finger points, rather than the direction in which the back of the hand points. In general, we can say that in any gesture, some geometric properties of the posture and movement are significant and some are not. A possible definition of the significant aspects is: those which would remain the same even when the sign was performed by a different avatar, with different body proportions. We are currently considering a revised version of the notation which would attempt to record signs in terms of such significant properties, and would be intended from the beginning for computer animation. Extending HamNoSys or SiGML to other classes of movement, such as those required by interactive characters in virtual environments, will be the subject of future research.

4.2 Facial Expressions

As mentioned above, non-manual signing includes a range of bodily movements, of head and shoulders, that can be animated by controlling the articulation of appropriate joints. In addition, there are facial expressions that are animated by controlling the vertices of the facial mesh.

Some expressions, denoted *mouth gestures*, come from a set of gestures used in signing, such as puffing out a cheek or raising the eyebrows. Other expressions, denoted *mouth pictures* consist of the visemes corresponding to an arbitrary phonetic (IPA) string. For convenience, this viseme string is expressed using the SAMPA (Wells, 2003) conventions for transcription of the IPA.

As reported in (Elliott et al., 2004), Animgen assumes that each avatar comes with a set of facial deformations, which are named morphs, which can be applied in combination to animation frames. Animgen has no detailed model of morphs but specifies a weighting for each morph for each animation frame.

Facial non-manuals used in SiGML are encoded as *morph trajectories*. A trajectory consists of a morph name, the maximum weighting of that morph to be applied, and an envelope describing the attack, sustain, and release for the morph.

Morph trajectories can be combined in series and in parallel to build up an arbitrarily complex definitions that are specified in a configuration file specific to each avatar. The creator of the avatar creates the avatar's morph set and the mapping of SiGML facial elements.

Figure 14 gives such a specification for a mouth gesture, which is defined as a mouthing of "bEm". It is realised by a sequence of three morphs corresponding to the three phonemes, where the first and third have been given identical visual representations.

The timings (slow, medium, fast, zero, or sustain to end

```
<mouth_gesture sigmlName="L09">
    <morph name="mbp" amount="1" timing="mt-ft"/>
    <morph name="aaa" timing="ft-mt"/>
    <morph name="mbp" timing="mteml"/>
</mouth_gesture>
```

Figure 14: Definition for Mouth Gesture L09

of sign) can be given symbolically (s, f, m, −, or e). The symbolic tokens are mapped to times in another configuration file.

Additionally, "manner" components determine how the morph approaches its full value during the attack, and how it tails off during the release. The possible values for this are "t" (tense) and "l" (lax). They are mapped to sets of parameters for a general model of accelerations and decelerations.

An extension of this format, illustrated in Figure 15, is used to define morph trajectories for viseme sequences derived from SAMPA strings.

```
<sampa phonemes="EIszi">
    <morph name="cgng" amount="0.7" timing="m t - m t"/>
</sampa>

<sampa phonemes="a_I">
    <morph name="aaa" timing="m t - m t"/>
    <morph name="cgng" amount="0.7" timing="m t - m t"/>
</sampa>
```

Figure 15: Defining SAMPA codes E, I, s, z, i, and aI

Several phonemes may correspond to the same viseme, for example *E, I, s, z,* and *i*. Hence a single specification is used to animate any of the phonemes in a list. Diphthongs are often required, but they vary from language to language. In order that a single set of definitions can be used for all languages, we require that diphthongs are tied together with an underscore. Hence the diphthong *aI* is encoded as *a_I*.

In order to handle coarticulation in a viseme sequence, the release of one trajectory is overlapped with the attack of the next. However, this is a largely untested approach and we are not confident that it will provide mouth movements suitable for lip-reading, for example. Instead, we are working to incorporate leading work on audio-visual speech synthesis based on appearance models that has been undertaken at UEA. This work is discussed below.

4.3 Animation System Architecture

As stated earlier the synthetic animation system we describe here was developed as part of a complete prototype system in the ViSiCAST project, in which the input is a natural language (English) text for which a signed animation is generated. This system divides into a front-end, which applies natural language processing techniques based on DRT and HPSG (Safar and Marshall, 2002), and a back-end — the system described here. The interface between these two subsystems is a

phonetic-level definition of the required signing sequence expressed in SiGML. The main data flow in this back-end SiGML-to-Animation system can be viewed as a pipeline as shown in Figure 16. Of the three processing stages shown in this figure, the most significant (because the most novel) is the central one which converts SiGML sign definitions into the corresponding animation frame definitions, as described earlier in this section. One important feature of the architecture not explicitly represented in the diagram is the fact that the synthetic animation module has an additional input, namely the description of the avatar's geometry, which is supplied with each avatar as an essential part of its definition.

The first stage in the pipeline decomposes the input stream into individual "signing units". A signing unit is typically an individual sign expressed in gestural SiGML, but it may instead consist of motion capture data for a sign (in which case it by-passes the synthetic animation stage). The first stage can also perform translation of a sign definition from HamNoSys to SiGML. The final stage in the pipeline consists of rendering software, which applies conventional 3-D animation techniques to each packet of frame data, first to determine the configuration of the avatar's surface mesh corresponding to the given configuration for its virtual skeleton (and morph weights), and then to render this mesh with the appropriate colouring and texture on-screen. A separate controlling module manages the scheduling of the necessary data transfers between the individual stages shown in Figure 16.

Figure 16: Processing Sequence for Synthetic Animation of SiGML

4.4 Appearance Models for Faces

Appearance models (Cootes et al., 1998) are statistical models of the shape and appearance variation of the face, which are learnt from hand-labelled facial images. Traditionally these have been used in the computer vision community to track and recognise faces (and other objects) in video sequences. Analysis is done by synthesis, i.e. the model is able to synthesise realistic example images by applying the appropriate parameters and an optimiser is used to update an estimate of the parameters such that the original and model generated images coincide. The face in an image is then encoded in terms of the parameters of the model, or is mapped to a point in a face-space spanned by the model.

Work at UEA on modelling talking faces has focussed on the use of shape and appearance models. A talker first recites a series of training sentences and the video analysed using the shape and appearance model. Since the face in a single frame forms a point in the model-space, a sentence forms a trajectory in this space. These trajectories are segmented according to their phoneme boundaries, derived from the corresponding acoustic signal.

To synthesise a novel utterance, a sequence of phoneme symbols is required. The synthesiser then selects a sub-trajectory from the original data that corresponds to the desired phoneme in the closest context to that in which it appears in the new utterance. These sub-trajectories are concatenated to form a new trajectory of model parameters, which are then applied to the model to create a realistic synthetic talking face.

This approach provides the flexibility and efficiency of traditional graphics-based talking faces with the realism of traditional image-based talking faces. A further advantage is that a complete avatar can be animated using this technique, so the talking head can be coupled with signing and other manual gestures (Theobald et al., 2003). Here the geometry of the face of the avatar is updated using the shape component of the appearance model, while the appearance component provides a texture update that significantly improves the realism when, for example, only a single texture is used.

5 Future for Signing and Expressive Characters

To develop virtual human signing it has been essential to address issues of both human animation and content creation. Animation only becomes acceptable once it achieves good visual realism with relatively natural motion. To support useful quantities of signed content it was necessary to develop scripting techniques soundly based in signing linguistics.

A benefit of using notation is that semantically unimportant information can be left implicit. An example is the position of elbows during signing. During animation, such implicit information is reconstructed using inverse kinematics. For representing more general gestures it is likely to be necessary to provide the option of being more explicit about aspects of gesture that do not matter for signing, but the principle of minimising the amount of explicit information is crucial.

The choice of a high-level representation is important if animation is to be scripted without knowledge of the physical dimensions of the avatar. A crucial part of our work has been an extended avatar definition format that enables the Animgen software to generate acceptable animation for any compliant avatar. A number of different avatars have been used in illustrations in this paper, but the software does is generic.

The notation concentrates on gestures for signing and only addresses upper body movement. There are few features relating to interaction with the environment, al-

though contacts between parts of the body are addressed in detail. We intend to develop SiGML to encompass a wider range of movement and gesture including conversational gesturing, whole-body actions such as walking and running, and interaction with physical objects.

An immediate application will be through the EPOCH Network of Excellence (Arnold, 2003), using avatars to help the user visit virtual cultural heritage sites constructed using the CHARISMATIC UEA/TU Braunschweig modeller (Day et al., 2003). A scenario would be that the user follows a walking, talking, multi-lingual virtual guide to places of interest in the scene. Ideally the visitor should be able to interact (via speech) with the virtual guide as well as the rest of the model.

We have introduced the leading work on audio-visual speech synthesis that is undertaken at UEA. To date, this work has focussed on the synthesis of the visible articulators associated with speech production only, i.e. the lips, teeth and tongue. It is well known that realistic conversational characters require expressive speech, which is lacking in the current system. To determine whether the model is able to re-synthesise the range of expressions required by a conversional character, it is currently being used to analyse the face of a signer and re-synthesise the facial movements on a virtual signer.

Much of our experience with signing appears to have wider application to work on expressive characters. The repertoire of techniques is clearly applicable to animation of more general gestures, although it remains to be seen how much extension is necessary to notations for signing and to animation techniques to achieve this purpose.

Acknowledgements

We acknowledge with thanks financial support from the European Union, and assistance from our partners in the ViSiCAST and eSIGN projects, in particular Televirtual Ltd. who supplied one of the avatars and the supporting rendering software.

References

D.B. Arnold. Plans for the EPOCH Network. In *VAST2003 Symposium*, Brighton, 2003.

T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. In H. Burkhardt and B. Neumann, editors, *Proc. European Conference on Computer Vision 1998, Vol 2*, pages 484–498. Springer-Verlag, 1998.

S.J. Cox, M. Lincoln, J Tryggvason, M Nakisa, M Wells, M. Tutt, and S Abbott. TESSA, a system to aid communication with deaf people. In *ASSETS 2002, Fifth International ACM SIGCAPH Conference on Assistive Technologies*, pages 205–212, Edinburgh, Scotland, 2002.

A.M. Day, D.B. Arnold, D. Fellner, and S. Havemann. Combining polygonal and subdivision surface approaches to modelling of urban environments. In *Proceedings of Cyberworld 2003*, Singapore, December 2003 2003.

R. Elliott, J.R.W. Glauert, and J.R. Kennaway. A Framework for Non-Manual Gestures in a Synthetic Signing System. In *Proc. Cambridge Workshop Series on Universal Access and Assistive Technology (CWUAAT)*, 2004.

R Elliott, JRW Glauert, JR Kennaway, and KJ Parsons. D5-2: SiGML Definition. working document, ViSiCAST Project, 2001.

T Hanke, G Langer, C Metzger, and C Schmaling. D5-1: Interface Definitions. working document, ViSiCAST Project, 2000.

J.R. Kennaway. Experience with and requirements for a gesture description language for synthetic animation. In *5th International Workshop on Gesture and Sign Language Based Human-Computer Interaction*, LNAI, to appear. Springer-Verlag, 2003.

R. Kennaway. Synthetic animation of deaf signing gestures. In *4th International Workshop on Gesture and Sign Language Based Human-Computer Interaction*, LNAI, pages 146–157. Springer-Verlag, 2001.

F. Pezeshkpour, I. Marshall, R. Elliott, and J.A. Bangham. Development of a legible deaf signing virtual human. In *IEEE Multimedia Systems '99 (IEEE ICMCS '99)*, 1999.

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.

E. Safar and I. Marshall. Sign Language Translation via DRT and HPSG. In A. Gelbukh, editor, *Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, Lecture Notes in Computer Science (LNCS), pages pp58–68, Mexico City, Mexico, 2002. Springer-Verlag.

B.-J. Theobald, A. Bangham, I. Matthews, J.R.W. Glauert, and C.C. Cawley. 2.5D Visual Speech Synthesis Using Apprearance Models. In *British Machine Vision Conference (BMVC 2003)*, Norwich, UK, 2003. BMVA.

ViSiCAST. Virtual Signing: Capture, Animation,and Storage. http://www.visicast.cmp.uea.ac.uk, 2000.

J Wells. SAMPA computer readable alphabet. http://www.phon.ucl.ac.uk/, 2003.