Animating Sign Language:

The eSIGN Approach

Animating Sign Language: The eSIGN Approach

Written information, especially essential eGovernment information, is often inaccessible to Deaf Sign Language users. The eSIGN technologies have been developed as a direct response to the need to create sign language content efficiently. This document outlines the results of the work done by the eSIGN project and will help key stakeholders to learn how they can include signing on web pages and other screen-based technologies.

eSIGN (Essential Sign Language Information on Government Networks) was originally funded under the Information Society Technologies (IST) programme of the European Union's Fifth Framework and supported by the European eContent programme. The eSIGN partners, and contributors to this document, are:

University of Hamburg – Institute of German Sign Language and Communication of the Deaf (DE): Thomas Hanke, Constanze Schmaling and Hortensia Popescu

University of East Anglia – School of Computing Sciences (UK): John Glauert, Ralph Elliott, Richard Kennaway and Judy Tryggvason

Viataal (NL): Han Frowein, Margriet Verlinden and Inge Zwitserlood

Royal National Institute for Deaf People (UK): Guido Gybels, Neil Thomas and Mary Henley

Norfolk County Council (UK): Tim Anderson and Bryn Davies

Televirtual (UK): Mark Wells

EDS Business Solutions GmbH (DE): Uwe Ehrhardt

Table of Contents

1 Introduction

Attitudes to accessibility have improved. New laws, such as the Disability Discrimination Act in the UK, mean that organisations supplying information to the general public are likely to want to make their services more accessible. For sign language users written text can be hard, or impossible, to understand. eSIGN technology offers tools which can provide information in sign language, using virtual humans (avatars). This document outlines the eSIGN approach and explains how signed content can be added to web pages and other screen-based media.

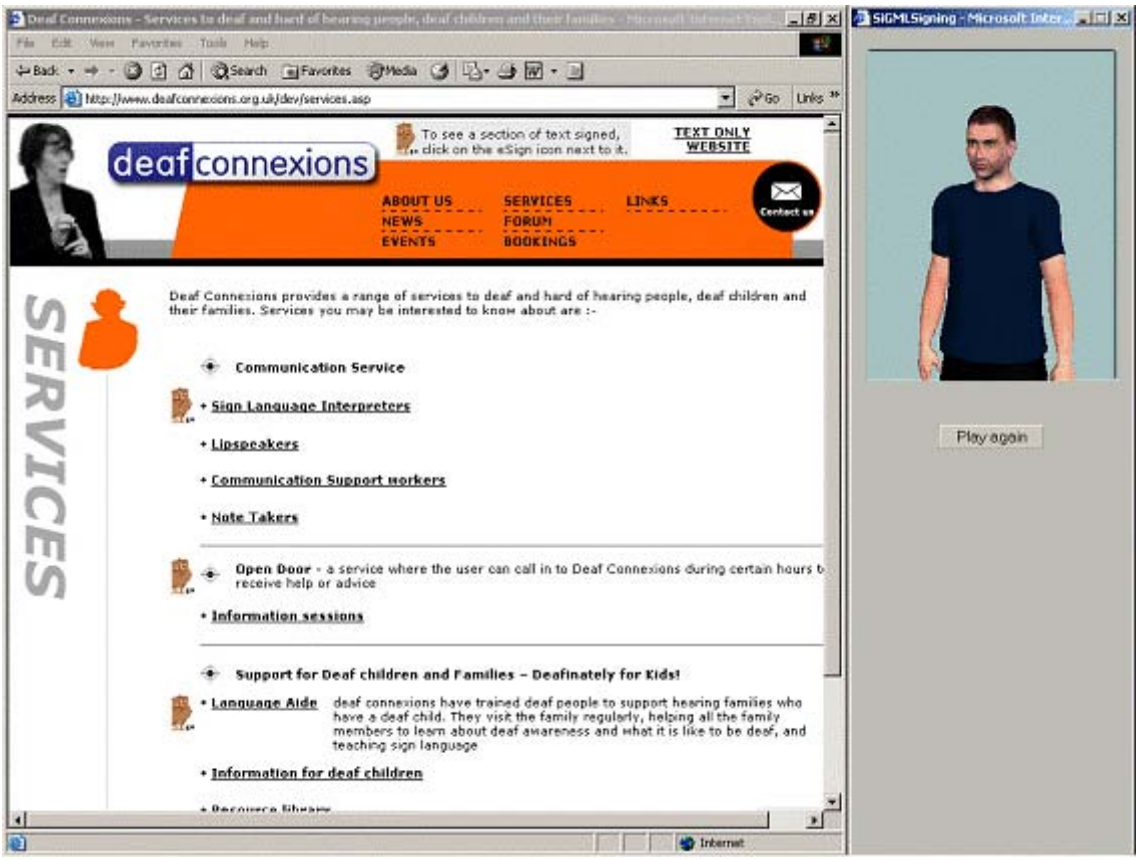


Figure 1: Example of a web site providing sign language information using an avatar.

1.1 Who will find this document useful?

This document is aimed at a range of potential stakeholders, many of whom will have little previous experience of sign language, or its value to the Deaf community. Service providers, technology implementers, sign language translators and computer software engineers all play a part. By using any of the eSIGN technologies, both service providers and end users will become aware of its advantages, and the demand for signed information will increase.

Providers of Services for Deaf People

EVERYONE who makes information available to the general public needs to consider accessibility. For sign language users, eSIGN systems can help. eSIGN technology provides a flexible means by which any organisation can provide signed information on

screen-based media (web sites, kiosks, information points, displays etc.). eSIGN software would be suitable for use by a variety of public-facing organisations:

- banks
- post offices
- museums and galleries
- company receptions and help desks
- internet mail order companies
- eGovernment organisations
- services specifically for the Deaf community

are just a few examples where the addition of sign language would be a real asset.

The signed material can be easily updated by anyone with sign language translation skills. So this method of providing signed information can be implemented economically. In some situations creating signed content can be partially automated. In such cases, site maintenance could be undertaken by people with no signing skills at all.

Technology Implementers

Organisations that produce web sites, information kiosks and other screen-based public display units will be able to host signed content using eSIGN technology. Simple interactive translation systems are possible.

Signed versions of printed matter, such as information brochures, leaflets, etc. on mediums such as CD-ROM, could also be produced.

Sign Linguists

Translators and Sign Linguists play a key role in the development of virtual signing technology.

At present there are three ‘dictionaries’, or lexicons, one for each of the sign languages involved in the eSIGN project:

- British Sign Language (BSL),
- German Sign Language (DGS)
- Sign Language of the Netherlands (NGT).

These lexicons are an open resource and are used to build signed content. Anyone developing eSIGN-style material can use the lexicons as they wish, and also modify and update them. In this way, users of the tools always have the best possible database of signs available to them.

The expertise of Translators and Sign Linguists is essential to build the lexicons and in using them to make accurate sign language translations of any written texts. Their contribution to monitoring the quality of any signed output is also vital.

As the technology becomes more widely used, the demand for Sign Linguists and trained Signers to help extend these lexicons will increase too. New ones for sign languages from other countries could also be created in the future.

Computer Scientists

Web page developers

A Cookbook is available as guide for web page developers. It gives technical details of the various ways in which signing can be added to new or existing web pages with only minimal adjustment to the HTML code.

Content Management

The virtual signing technology can be integrated into web site content management systems. Simple and straightforward updates of signed information can be made in parallel with routine updates of text information.

In some cases this process could be entirely automated. For example, a site displaying opening times for council offices may draw information from a lookup table. It is possible to build a tool so that when information in the table is adjusted, the signed sequence is updated as well.

Using this approach content can be (semi-)automatically created for a range of purposes. Automatic creators for signed online weather forecasts and job advertisement web pages have already been developed. In both cases, a tool enables non-technical staff with little or no signing skill to prepare updated signed pages. The scope of this type of tool is broad. It can be used for any type of content with a regular structure. Future applications could include:

- display boards listing delays to trains or planes along with the reason for the delay
- deadlines for the return of various pieces of council paperwork
- leisure; community events of different types at a range of locations
- automatic ticket kiosks for travel or other purposes.

1.2 What would Deaf People Like?

Why don't they just read the text?

As with spoken languages, most countries have their own sign language. Within the eSIGN project three different sign languages are represented – BSL, DGS and NGT.

Many people who are born deaf learn sign language as their primary language, and it remains their preferred, or first, language. There is no written form of sign language, so Deaf people have to rely on reading and writing in their second or less-preferred language (e.g. English, German or Dutch). A significant proportion of Deaf people therefore have a strong preference for accessing information in sign language rather than as written text.

At present it is not usually possible to access signed information on the Internet. A small number of sites offer video clips, but usually Deaf site users can only access information by reading the text. This can present great challenges for many Deaf people, and is a particular problem in the context of eGovernment sites. The information on such sites may be complex. In addition it is important that people understand it correctly in order that they can successfully access these important services as and when they need to. The access that Deaf people have to this type of essential information could be greatly improved by the provision of sign language information online.

Why don't they get a Sign Language Interpreter?

Human aids to communication, for example sign language interpreters, are regularly used by Deaf people to support communication. This may be during a face to face meeting, or to help with tasks that require reading and writing. However, there is a shortage of sign language interpreters. So it is often not possible to gain access to these services, especially at short notice. Interpreter services are also expensive.

Although virtual signing can never be expected to replace human interpreters, or to significantly reduce demand for them, this technology can provide a readily available alternative, increasing access to information in situations where an interpreter would not usually be an option.

Why can't you just use videos of signing?

Sign language can be displayed on web sites using video, and this method is usually satisfactory to Deaf people, especially those who have access to fast internet connections. However, there are disadvantages to this means of providing information:

- Videos must be well produced and of high quality, which is expensive for the content provider.
- Each time content changes new videos must be made, increasing the costs further.
- There are continuity issues. Making videos consistent, i.e using the same signer, in the same clothing and with the same background, so that signed phrases may be joined together, complicates the content maintenance process.
- Storage and download of videos can also be problematic as they are large files. For home users on dial up connections the time and cost involved in download of video sequences may be prohibitive.

Virtual Human Signing as an alternative to video

Virtual human signing thus provides a better alternative to videos as a means of presenting signed information on the Internet. The avatar software needs to be installed by the user before virtual signing can be accessed. This is offered either as a free download or CD-ROM. The signed content is placed on the server of the provider. Since it consists of SiGML files (a form of XML), which are very much smaller than videos, the amount of storage space required on the server is minimal, and download time for the users is very fast.

Another advantage of virtual human signing over video is that it is more user-friendly. A user can browse more quickly through information, control the speed of signing and change the view angle of the virtual signer, things which are not possible with videos.

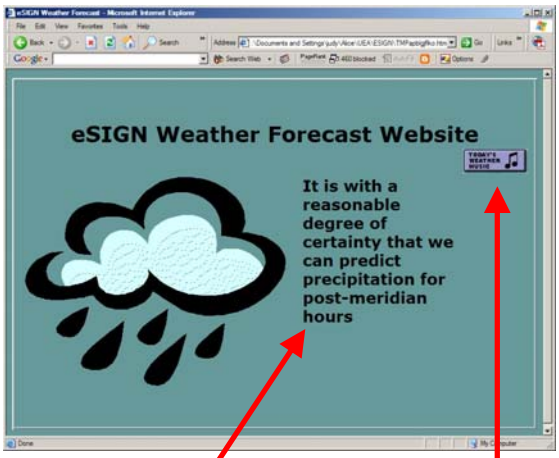
Is automatic translation possible?

Sadly no. It would be nice to have text written in English automatically translated into, say, British Sign Language. Unfortunately, the grammar and vocabulary used in sign languages is quite different from European spoken languages. Even within the domain of spoken languages, translation is an extremely complex matter, and only very basic automatic translation is currently possible. For sign languages the problems are further compounded by things such as spatial information (where should the sign "over there" point to? etc.). However, Section 2.2 describes an automated solution that eSIGN paratners have developed for restricted language domains.

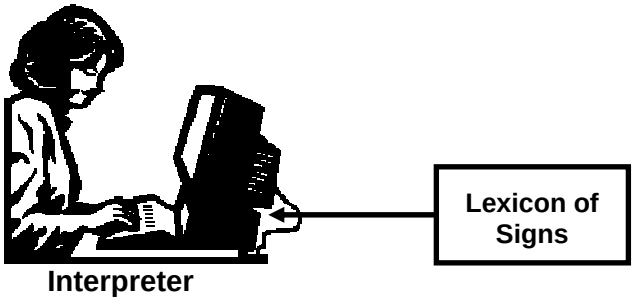
1.3 How do I get signing on a web page?

Let's assume you already have a simple web page you want to modify. These are the five basic steps you would follow. More detailed information is contained in following sections.

1. Prepare your text content. Just converting to Plain English, German, Dutch etc. may not be enough. Deaf people may not be clear about some information that hearing people take for granted. The text may need extending or modifying. Interpreters and other experts should be consulted.



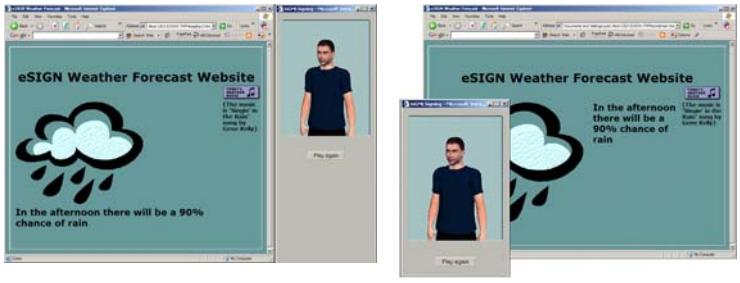
2. As with any with any foreign language, get trained interpreters (or signers) to do the translation. This involves using the **eSIGN Editor** software, which creates or uses a lexicon of signing descriptions and computer animation instructions (SiGML).



3. Modify your web page layout:

Either adjust your web pages to make room for the avatar:

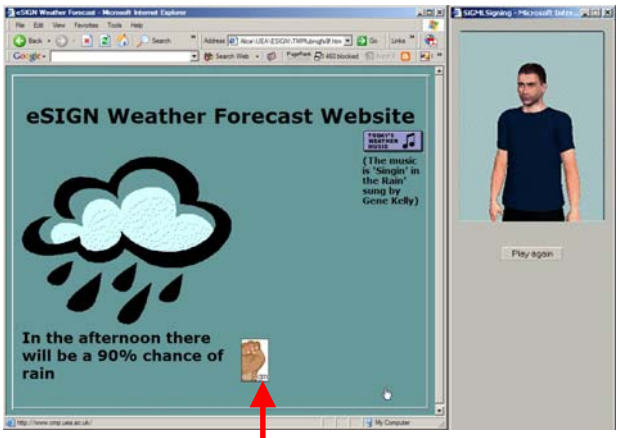
Or leave it as it is and use the 'floating avatar' model.



Either adjust webpage layout so it's ready to make room for the Avatar …

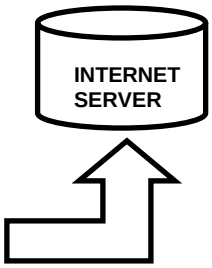
… or leave it as it is so the 'floating avatar' can be added later.

4. Hyperlink the text to be signed, and add a small standard script which displays the Avatar and makes it sign the animation instructions created in Step 2. The links are usually small icons, but the text itself could be linked instead.



Small '*sign this*' icon. When clicked, the Avatar signs the requested phrase. You could also hyperlink the text itself instead.

5. Upload your modified web page and SiGML (animation instructions) to the internet server.



+

```
<sigml>
   <hamgestural_sign gloss="HAUS3">
     <sign_nonmanual>
       <mouthing_tier>
         <mouth_picture picture="haUs"/>
       </mouthing_tier>
     </sign_nonmanual>
     <sign_manualboth_hands="true" lr_symm="true">
       <handconfig handshape="flat"/>
           <handconfig extfidir="o"/>
       <handconfigpalmor="dl" rel_palmor="true"/>
       <handconstellation contact="touch"/>
             <directedmotion direction="dr"/>
             <directedmotion direction="d"/>
         </sign_manual>
     </hamgestural_sign>
</sigml>
```

Updating your pages is done in a similar way. However, if your content has a regular format, and the content covers a limited domain, it may be possible to automate page updates. No knowledge of sign language is required for this.

2 Content Creation the eSIGN Way

This section describes the stages of content creation. Starting from the creation of a sign lexicon, signed sequences are constructed and then added to web browsers or tailored applications.

2.1 Manual Editing

Steps in the Translation Process: eSIGN Editor

The translation process needs to be done by specialists who are trained in translation and/or the creation of signed content, for example, sign language interpreters or relay interpreters. Although this may look complicated to the layman, a number of interpreters have already been sucessfully trained and find it easy to use the Editor.

The original text is entered into the eSIGN editor and divided into sentences or phrases[1]. The corresponding sign language translation is entered alongside the original text, gloss by gloss[2].

Figure 3: Original text and gloss transcription in the eSIGN editor

Glosses with their HamNoSys strings[3] are taken directly from the lexicon database for the appropriate language. Due to the diversity of sign language (dialects, individual preferences, etc.) there is often more than one entry for a given gloss. In order to choose the correct sign, it is possible to either

- have a look at one or more video clips that can accompany each sign in the database;
- pick a sign and let the avatar sign it; or
- read the HamNoSys string that describes the manual components of the sign.

Signs that are not in the database are transcribed in HamNoSys and may be entered into the database.

Mouth pictures can be retrieved from a pronunciation table in the database by simply writing the mouthing in standard orthography. It will then be automatically transferred

[1] A full description of the eSIGN editor and its structure is available on request. The eSIGN editor is freely available under licence to all content creators.

[2] Glosses are uninflected spoken words that are used as labels for the corresponding signs.

[3] HamNoSys is a way of describing sign language in the form of special characters. See section 4.3 for more details

into SAMPA[4]. It is also possible to enter SAMPA manually. There is also a database for the mouth gestures with various examples, graphics and video clips.

It is possible to alter signs that are retrieved from the database. For example it may be necessary to change the HamNoSys string so that a sign performance suits the given context. In some instances, mouthing instructions can be edited, or added if they were not stored alongside the sign in the database.

In the eSIGN editor it is also possible to add any necessary facial expressions, body movements and pauses, or to alter the position or location of a sign.

In addition, it is possible to add URLs and link the text with illustrations that will be displayed on the webpage simultaneously with the signing.

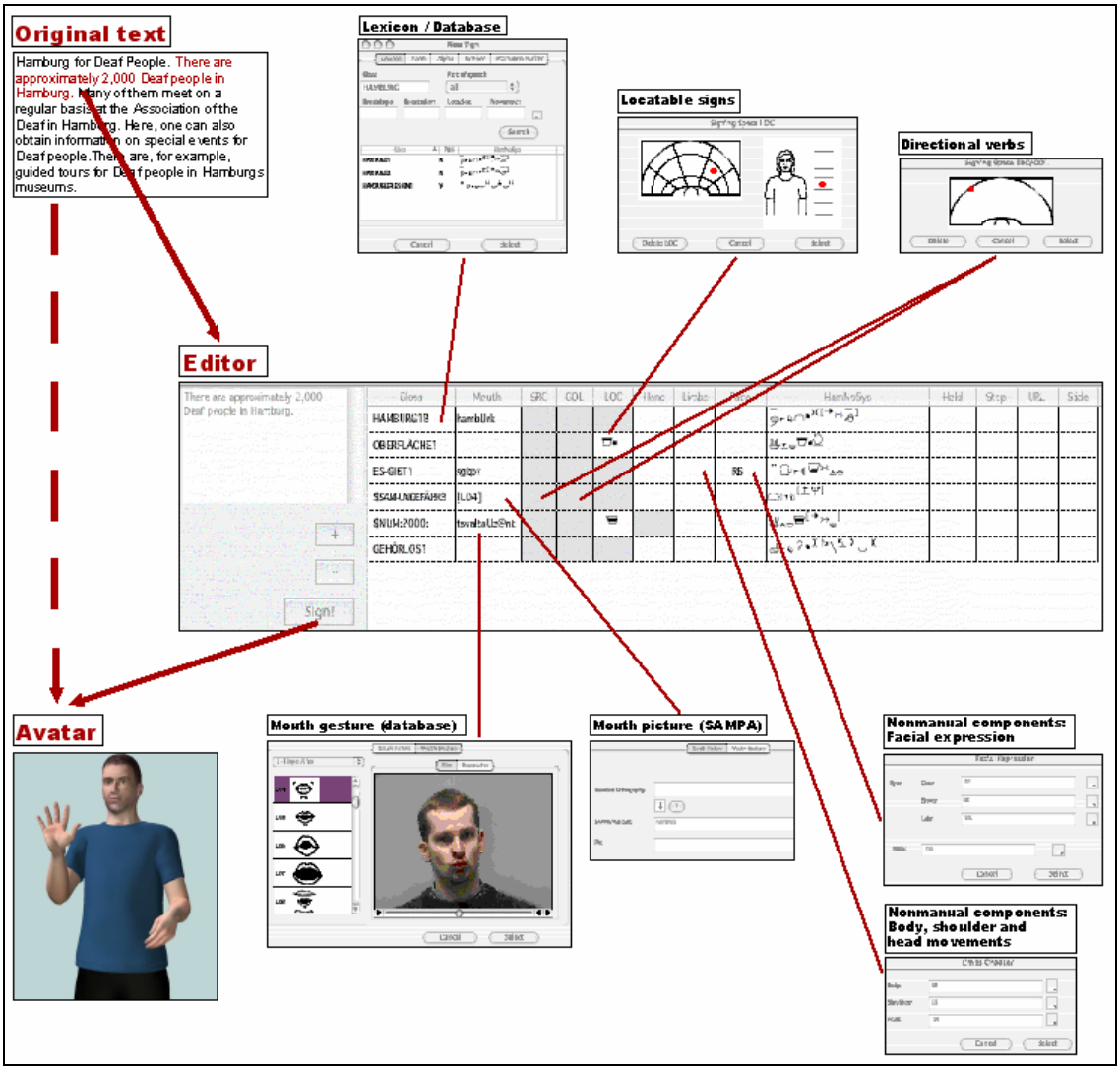


Figure 4: Overview of the structure of the eSIGN editor

In a last step, the avatar's performance is examined, and any tweaks or edits can be made. If a sign looks different from what is expected, HamNoSys experts will check the HamNoSys string in the database. If necessary, the transcription of a sign is corrected.

[4] SAMPA is used for encoding phonetic names for the mouth pictures by means of an alphabetic character set. The phonetic names represent the gestalt performed when producing the corresponding sound. For German pronunciation data supplied by the IKP at the University of Bonn is used.

Text Amendments and Changes: Translation vs. Explanation

When providing information in a visual language, literal translations do not always offer the best solution. Where it seems appropriate and/or necessary for linguistic reasons, chunks of information may have to be rearranged. This may be necessary so that the translations resembles a more spontaneous sign language and reflects the needs of a language that does not have a written equivalent.

Some texts may require some additional explanation. This is true for example with legal texts. Because of the complexity of much legal wording, an exact translation can be a difficult task. It would be possible to translate the texts into sign language verbatim, but it would probably be of little use for the average Deaf user. A hearing person has numerous opportunities to find explanatory texts on the Internet or in books that can help with comprehension or clarification. As these additional books and texts are not available in sign language, a direct translation of the text is of little use to a Deaf person.

In such cases, rewriting may be necessary in order to go into more detail. Explanations of difficult legal terminology or complex government regulations are examples. To make sure that the content retains its precise original meaning it may be necessary to get professional advice.

2.2 Structured Content

eGovernment portals are likely to have a certain amount information to present in a consistently structured way. Times, dates and amounts of money, for example, may change frequently but the format in which they are presented on web pages is likely to remain the same for long periods of time. Since web page content might need to be altered regularly, a tool which does not require great technical expertise on the part of the user would make frequent amendments much easier to make, ensuring a viable overall solution. Moreover, it is not realistic, nor practical in the long term, to develop a new tool for every web page or group of web pages.

eSIGN provides a general software tool for such content creation – known as the Structured Content Creation Tool (SGCT). This tool takes as input:

- an HTML template, the basic page including code for the signing avatar
- a set of XML files defining:
    - the sign sequences available as building-blocks
    - the ways in which these sequences may be combined
    - the GUI which the content creator uses to generate a complete HTML page.

Setting up the XML currently requires computer software expertise, although a more user-friendly version is being developed.

Given these inputs the tool displays the GUI containing the content creation options for the (content creating) user to use. Output is in the form of HTML pages.

The tool also supports multi-sign-lingual operation, both in terms of the language used in the GUI, and in terms of the sign language required in the generated sign sequences. The latter is implemented as XSLT transforms.

The basic sequence of events for the Structured Content Creation Tool can be described as shown in the figure below.

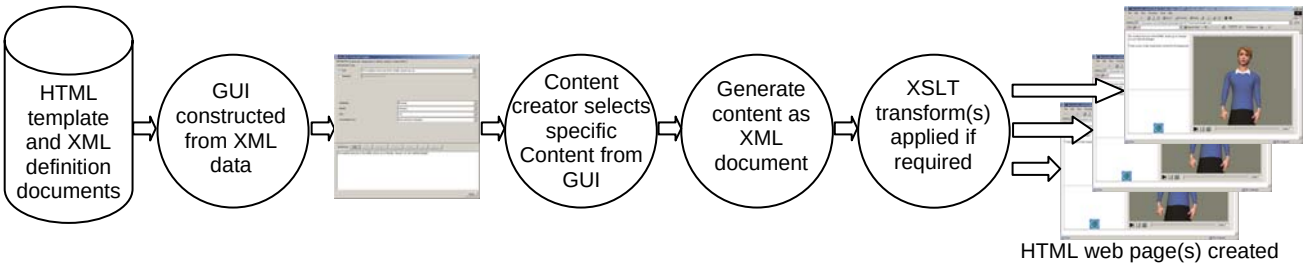


Figure 5: Content Creation Pipeline

Two demonstration applications have already been produced, one for a Weather Forecast Creator, and another which supports the creation of HTML pages for a Flight Booking system. The screen shots in Figures 6 and 7 show the results. Tabs at the top of each GUI allow the easy selection of the required information by subject area. Deletions, additions and amendments are possible. The final tab on each GUI lets the user choose an HTML filename, and then to create a signed web page from the selected data.

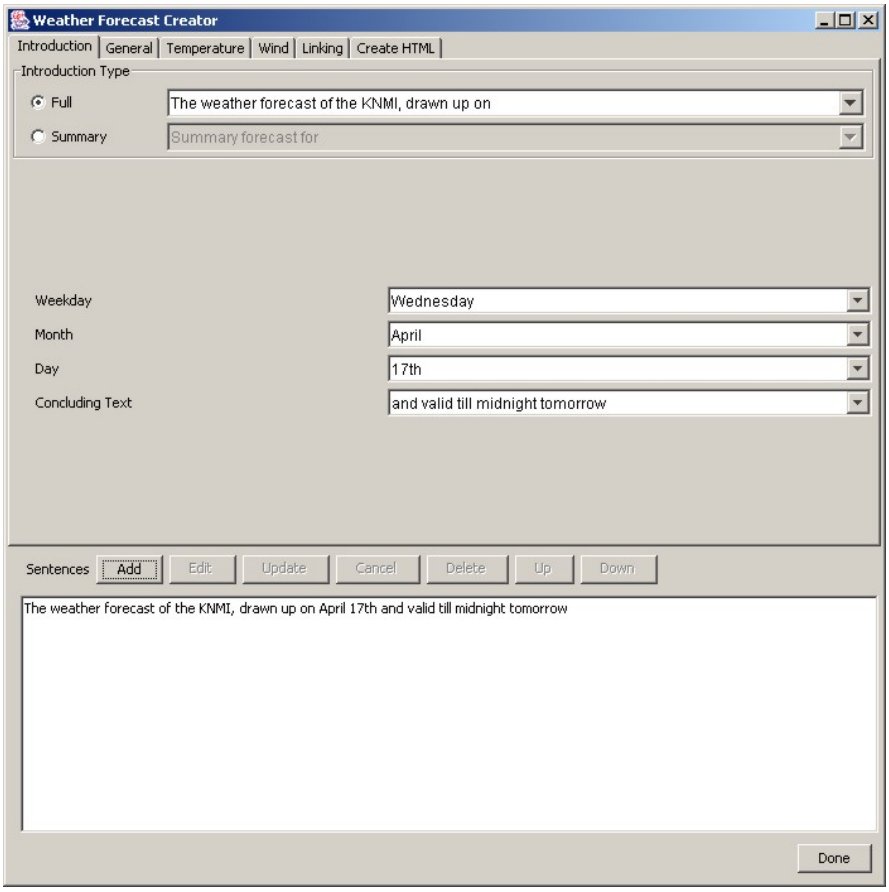


Figure 6: Example application of the structured content creation tool

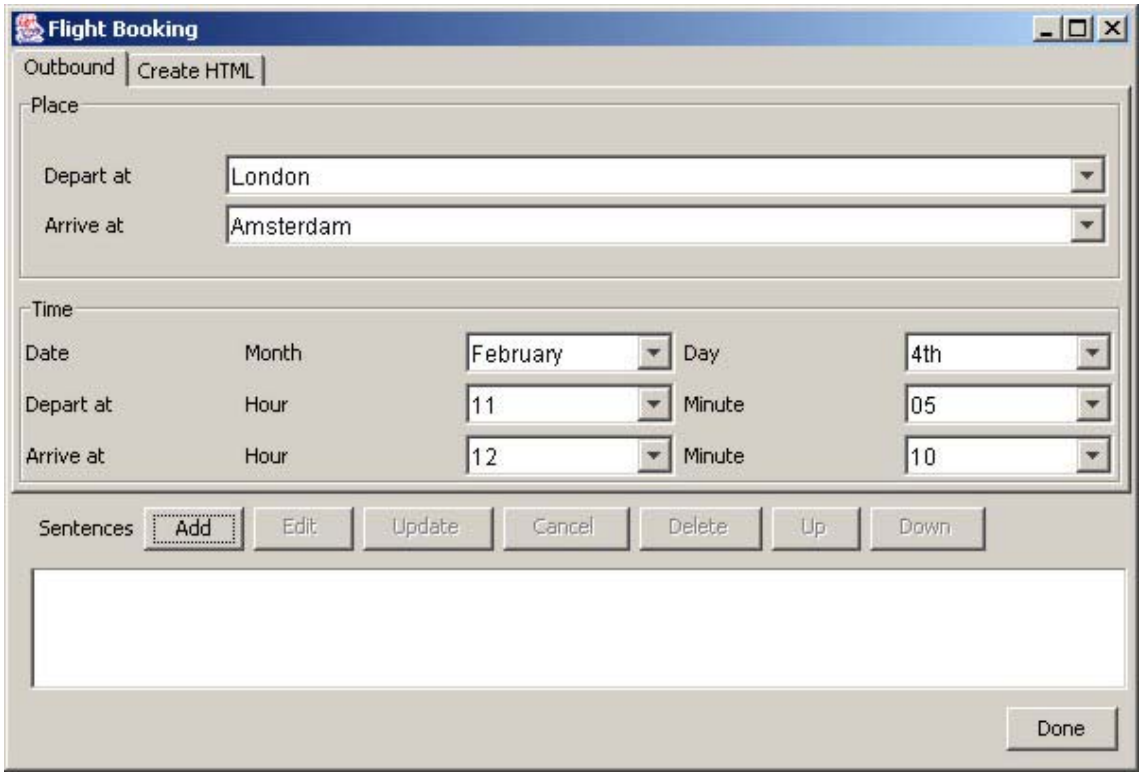


Figure 7: Example application of the structured content creation tool

Figure 8: Two further demonstration GUIs of the Structured Content Creation Tool

Further demonstrations have been made by eSIGN partners Viataal for the creation of job vacancy web pages. This is another typical example where the content changes frequently but where the structure remains the same. Sample part-screen shots from this application, *Vacature opstellen in NGT*, are shown in Figures 8 and 9.

Figure 9: Sample part-screen shots from Viataal's Vacature opstellen in NGT

2.3 Lexicon Creation

Before starting the translation of a text it is advisable to create a basic lexicon for the chosen content area. It may be possible to retrieve signs from an already existing corpus/database. This is the approach taken at the University of Hamburg where the database is fed from various projects.

New signs are entered into the database with a gloss name. Each sign is transcribed with HamNoSys[5] and all other relevant information is added. This includes information on mouth pictures or mouth gestures and on "inflection": the latter is necessary for the so-called directional verbs[6] and locatable signs[7].

Having set up a basic lexicon, the content creation process is much faster. If signs that are not yet in the database are needed in the translation process, e.g. because a given sign/gloss doesn't fit in a given context, these may be added to the database at a later stage.

2.4 Deployment – eSIGN examples

To be fully inclusive, eGovernment must offer Deaf citizens the opportunity to access information, and for communication and transactions to be incorporated into solutions. eSIGN provides deployment examples in the form of signed web pages. eSIGN technology has also been used in an application developed for non-signing assistants to provide government information to deaf clients in a face-to-face environment.

Web Pages

To develop web pages which include signed content, the developer will need to create signed versions of the web page text, by taking signs from a lexicon (see section 2.3), and creating new sign sequences (section 2.1). Once completed, the content may be added to web pages through the following steps:

- Script HTML page to include the avatar control

[5] See section 4.3 for details.

[6] Directional verbs are signs whose direction of path movement changes according to actor and patient of an action. In the HPSG-approach taken in the ViSiCAST project, the direction of movement of these verbs is not specified in the lexicon but only information on size and type of the path movement. The actual movement can be deducted from this lexical information plus source and goal of the verb.

[7] Locatable signs are signs that can be performed in different locations in space/on the body depending on the context. These signs have no location specified in the lexicon.

- Script HTML page to include the relevant sign sequences, or references to the appropriate SiGML[8] files.
- Add appropriate web page mechanisms to trigger the avatar signing of sections of text.

Web Page Design Issues

All eSIGN partners encountered challenges when it came to web page design. Their experiences, described here, serve to illustrate those which most web page developers will encounter.

Areas selected for deployment were primarily be those of interest to deaf people. But they also needed to satisfy additional criteria. Web pages should be of an appropriate size and simplicity and they should be suitable for the avatar to sign. The original content for which a signed version is to be developed had to be carefully chosen.

Naturally, all created pages should adopt the style of the parent site, and any modifications to enable signed versions should be small. For eSIGN, these were mostly achieved by linking from a “sign this” logo next to the text, or a hyperlink from the actual text to be signed. Initially the concept of pages based entirely on sign language were considered, but user research in all three countries showed a preference on having the original web site text visible, with sign language translations available on demand.

There were two solutions adopted to the introduction of the avatar to websites. Both entailed using frames, which meant that the avatar did not have to be re-loaded for every new page visited.

1. Reduce the space taken up by the regular web page to make room for the avatar
2. Make the avatar appear in the form of a pop-up window which could be moved freely over the web page being viewed.

In both cases, one frame corresponds to the avatar, together with either the SiGML it might be required to sign, or a file reference to it. The remaining frame(s) contains the original content of the page.

Using frames means that signed versions would not be available from all browsers, but as an ActiveX control, the avatar is currently only available for use in Internet Explorer 6 which is frame-compliant. Cross-browser and cross-platform issues are currently only addressed through the development of a server rendering platform which can create videos of the virtual signing output. Further modifications to make applications available to all platforms and browsers are still under development.

The initial eSIGN pages were developed individually, but a cookbook on how to create webpages incorporating the avatar is now available as part of the eSIGN software distribution.

Different examples offered by partners as solutions to the above issues are detailed below.

Germany

The aim was to provide deaf people with signed version of sections of Hamburg’s website (http://gebaerden.hamburg.de/). Pages that were translated for the project include:

[8] A form of XML. See Section 4.3 for details

- A “Welcome” page comprising two signed parts: first, a general introduction to the city, and second, a part called “Hamburg für Gehörlose” (Hamburg for Deaf people). This links to:
- The “Integrationsamt” (Integration Office) introduction page. The Integration Office is the authority dealing with all issues that have to do with people with disabilities in workplaces. This in turn links to:
    - A page with information about equalisation claims, which are paid by enterprises that do not meet a certain quote of people with disabilities among their employees.
    - A section from the Social Welfare Legislation (“SGB IX”) that explains the rights of disabled people.
- A group of pages about the Hamburg Lost Property Office:
    - The Lost Property Office: introduction
    - General information
    - Auctions
    - Lost something?
    - Found something?
    - Tracing lost property online
- Selected pages from the Hamburg City Parliament website:
    - Welcome to the HCP website
    - Question time on current affairs
    - HCP press information (archive)
    - Speeches of the president and vice-presidents (archive)
    - Who is working at the parliament?
    - HCP meetings – debates and votings
    - The HCP board

One significant aspect of the German pages was that where text was particularly complex (for example legal text), the text itself was rewritten to make it more comprehensible before a signed version was created.

Netherlands

The aim was to help deaf people improve their employment prospects and gain improved access to social services and facilities that are offered by government bodies.

Firstly, signed assistance for the completion of three e-forms was produced:

- Machtiging tot rechtstreekse betaling aan de jobcoach(-organisatie) *(Authorisation for direct payment to the job coach (/organisation)* An explanation of this legally-worded paper form, and the implications of signing, or not signing it.
- Declaratie reiskosten *(Reclaim travel expenses).*

  Explicit instructions for this paper form which is frequently completed incorrectly.
- Aanvraagformulier Tolk *(Application form for interpreting services)* An online application form where brief signed explanations are provided for each section which needs to be completed.

Secondly, a web page was created with signed summaries of job vacancies at Viataal.

Thirdly, a translation of a web page that contains information about the rights of handicapped people, provided by UWV, (Uitvoeringsinstantie Werknemersverzekeringen) the Dutch body that executes social insurance for employees and helps people to find or keep a job (www.uwv.nl).

The Dutch Gebarennet site can be seen at http://www.gebarennet.nl

UK

The UK deployment also provides sign language support for Deaf people on web pages, and signed assistance to help with the completion of online transactions. Pages from the newly created Deaf Connexions website were selected to demonstrate use of the eSIGN technology. Deaf Connexions is a voluntary sector organisation offering support and services to Deaf and hard of hearing people in Norfolk. The four pages used were:

- 'Contact Deaf Connexions' (signed information about the organisation).
- 'Arranging a meeting with Norfolk Deaf Connexions' (signed help to fill in a complex online form to arrange a meeting with a Deaf Connexions representative).
- 'When are Deaf Connexions next in your area?' (signed assistance for a simple online form, to find out when and where Deaf Connexions representatives will be in a particular areas so that a meeting can be arranged).
- 'Open Door' (information listing all facilities Deaf Connexions provide).

As a 'taster' the Deaf Connexions site plays and initial Flash movie of some signing for anyone who wants to see what the signing avatar looks like, before they download the avatar software.

The Deaf Connextions site is at http://www.deafconnexions.org.uk/

Other Deployment Examples - VANESSA

An additional UK solution recognises that not all deaf citizens have access to computers and internet connections in the home, and that not all Government information is available online at present. In the Norfolk area Council Information Centres (CICs) have been provided as a first point of contact for members of the public wanting access to Government information and services. As a result of this VANESSA (Voice Activated Network Enabled Speech and Sign Assistant) was developed and made available through one of the newly launched CICs. VANESSA aims to help CIC assistants communicate with their deaf clients, by providing Signed assistance for:

- Filling in paper forms such as those for Housing Benefit
- Arranging an interpreter booking for very complex matters
- More general issues such as "I will get you a form to fill in" or redirecting clients to more appropriate offices.

To develop this service into similar systems for installation elsewhere VANESSA could easily be modified by the following steps:

- Identification of extra phrases required for each new application
- Listing alternative ways of saying those phrases

- Creating signed versions of the phrases (see sections 2.1 and 2.3)
- Modifying a given text file to include the phrases, the alternative lists, references to the associated SiGML files(s) and (optionally) a list of likely responses the deaf client can choose from.

Screen shots of the VANESSA system are shown below in figures 10 and 11.

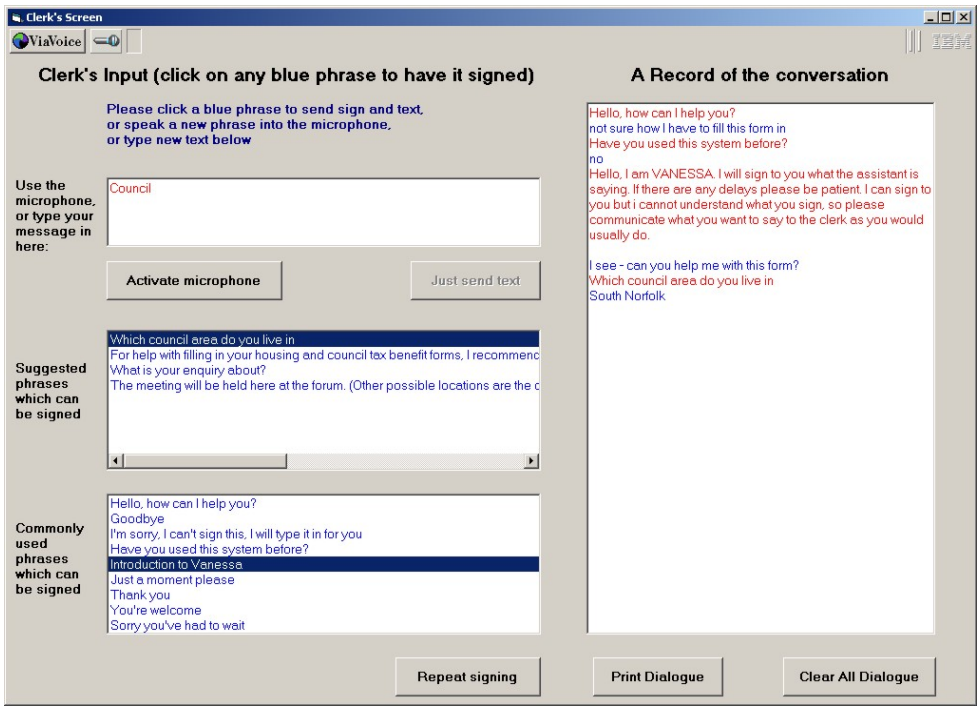


Figure 10 VANESSA interface, as seen by the Council Information Centre staff

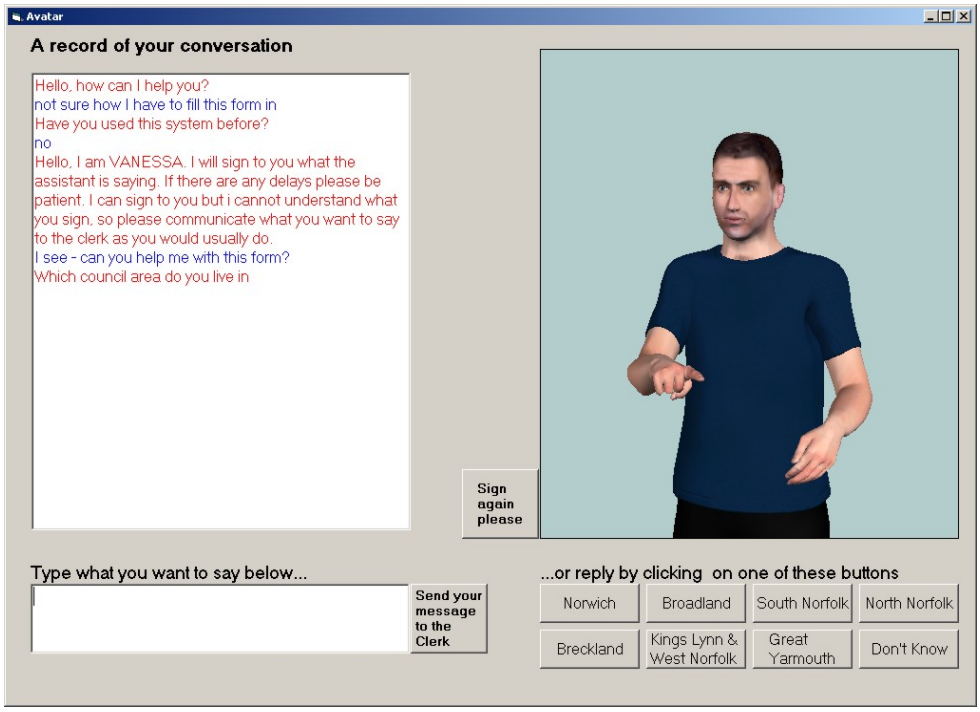


Figure 11: VANESSA Interface as seen by the Deaf Client

3 Content Creation Principles

This is about some of the practical aspects of successful content creation using eSIGN tools. It is essential to develop content in close collaboration with Deaf people. There is also a section the respective bandwidth requirements for eSIGN technogies and video.

3.1 User Involvement

User Involvement During the eSIGN Project

Products that are developed for specific end users will always benefit from involvement from those users themselves. This is especially the case in projects involving languages, specifically when these languages are not the native languages of the (other) project executors. For this reason, Deaf signers have participated in the eSIGN project in each partner country. Moreover, input was obtained from many additional Deaf people through focus groups and various evaluations of the project outputs.

The project and the end results benefited from this input in several ways.

- comments from the end users in this project were valuable with respect to the comprehensibly of the (synthetic) signing, the pleasantness of the virtual signing person and any strain in perceiving the signing.
- some of the people in the focus groups and evaluations volunteered ideas for future applications.
- users gave interesting views on the designs of the web pages presented to them
- the Deaf eSIGN staff played a major role in the creation of the signed lexicons and phrases.

Besides this, within the project the Deaf staff have been key to raising awareness of the technology and educating end users as to it's purpose and potential. The idea of synthetic signing is new to most Deaf people, many of whom are not used to any abstract forms of representation of their language. Therefore, signing avatars from both the ViSiCAST and the eSIGN project have been shown at many of the focus groups, presentations at Deaf clubs and evaluations. This was mainly done by members of the Deaf staff. At such sessions it was explained that the synthetic signing they saw was a technique that is still under development and the assurance that this technique is not intended to and will not replace real interpreters (a concern raised by some), but is aimed at increasing the amount of information in sign language that is available, especially on the internet. This resulted in quite a few (especially young) end users seeing the advantage of synthetic signing.

User Involvement in New Applications

In the application of synthetic signing it will remain necessary to involve end users. There are several ways in which they can be included. In the first place, (after training) they can help create new content and further expand the lexicons. Secondly, they can edit or comment on signed translations in terms of fluency and comprehensibility. (Unfortunately, most Deaf people are not fluent enough in a spoken language to do translations themselves; translations are best done by professional interpreters (see 1.1.3). The exception to this is relay interpreters; bi-lingual Deaf people employed in providing sign language for TV or videos). As described in 2.1.2, a mere translation of a

written text will not always lead to better understanding if the terminology used is difficult and structure of the text is complex.

Furthermore, it may be helpful to involve end users in the design of the web page that will contain signed information, since Deaf people can have special needs or desires that may be overlooked or not taken on board by hearing people. (for example a design could be made with the fixed design of text to the left side and an avatar to the right side, where a Deaf person might prefer a flexible avatar window that he can drag to the centre of the screen.)

While this project has mainly looked at the translation of written texts into sign language it is also possible for a sign language user to use this technology to create signed content directly, a process that does not require detailed knowledge of written language.

Finally, acceptance of synthetic signing will have to be further investigated and promoted, especially in the near future. This can be best done by, or with help of, Deaf people.

3.2 Avatar Compared to Digital Video

Why should a website use avatar technology if digital video still provides better quality? There is no question that current avatar technology provides lower quality than digital video with respect to smoothness of movement, expressivity of the face and many more details, and there are definitely many applications where digital video should be the technology of choice. However, many restrictions apply to digital video, such as the bandwidth requirements for digital video and the production cost.

Bandwidth: The amount of data used to drive the avatar is very low, in the order of size of corresponding written text, and it is independent of the resolution of the on-screen display.

Even the most recent codecs for compressing digital video deliver data rates exceeding those for avatar data by several orders of magnitudes:

1 second of signing	Avatar	MPEG-4
320x240	1 KB	68 KB
640x480	1 KB	250 KB

(Data taken from one eSIGN example text available both as digital video with a human signer and as an avatar performance.)

Experience with digital video in sign language on the WWW showed that many people did not make use of it since downloading the video to their home computers simply took too long or was to expensive. This is true for many potential users even today as broadband access with flat rate charges is not yet available everywhere in Europe.

Of course, digital video can be compressed to lower bandwidths than stated here by reducing resolution, temporal resolution or colour, but then video quality deteriorates so that both comprehensibility, and acceptability, of the video decrease significantly, losing the quality advantage of digital video signing.

Finally, the avatar 3D model also allows the viewer to compensate for some of the disadvantages of "canned" productions of a language that is mostly used in face-to-

face communication as he or she can zoom in, e.g. to have a closer look at the mouthing, or rotate the avatar to have a look from a different perspective.

Flexibility: Depending on the quality to be reached, the production of a digital video clip can be quite expensive as good quality requires studio technology either to be purchased or rented. Longer texts cannot be filmed in one shot and therefore require post-production. However, this initial capture is only a fraction of the total cost of many sign language texts to be expected on eGovernment sites: Once the text signed needs to be changed to meet changing legal regulations, to update information or whatever changes need to be made, you have to start from scratch width digital video: It is practically impossible to later film patches to change a signed text. Instead, at least whole sentences need to be redone, but realistically one will need to redo a whole text. This is where the avatar technology excels: Simply replace the phrases that need to be updated, export to website, and your new version of the signed text is ready.

For regularly changing text, or database query systems with limited flexibility of the language output, eSIGN technology also allows automatic generation of signed text. Certainly the creation of the language model needed exceeds that of producing canned text for the first couple of iterations, but then the model can be used to update the text without further translation cost. This approach is close-to-impossible to implement with digital video as stitching together pieces of digital video, e.g. sign by sign, results in rather unacceptable videos.

4 More about how we do it

How do we get a virtual human to sign? How is the avatar built? A lot of the eSIGN research has been devoted to creating avatar technology, and researching the best way to give the avatar instructions on how to move. One of the first steps is to describe the signing in the HamNoSys language. There is a final section here describing the principles of HamNoSys.

4.1 Virtual Human Avatar Technology

A computer-generated virtual human or avatar is modelled as a 3-dimensional deformable surface mesh which is attached to an underlying skeleton. The surface mesh is a connected network of coloured textured polygons. Standard 3-D rendering techniques, exploiting both graphics hardware and software, allow the surface mesh to be displayed on a computer screen. The surface mesh is notionally attached to the avatar's skeleton, a hierarchically structured set of bones. This notional attachment means that a change to the skeleton configuration determines a corresponding change in the spatial configuration of the surface mesh. Thus one way of changing the posture of the avatar is to make a change in the configuration of its skeleton. This method is the one generally used to move the avatar's limbs and fingers.

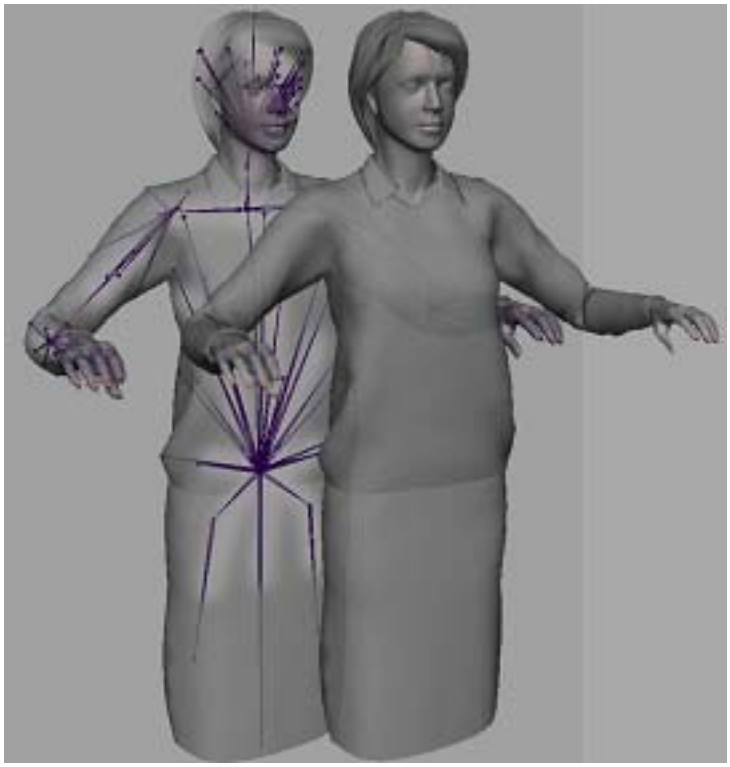


Figure 2: Avatar 'skeleton' and surface mesh.

Another way of changing the posture of the avatar is through the application of "morphs", a morph being a directly defined distortion of the surface mesh. Given a set of these morphs, one or more of them may be applied to the avatar at the same time in varying degrees, that is, with varying "weights". This technique provides a relatively flexible mechanism for manipulating the avatar's face. An animation of the avatar

consists of a temporal sequence of frames, each of which defines a static posture of the avatar at the appropriate moment. Each of these postures in turn can be defined, as has just been explained, by specifying the configuration of the avatar's skeleton, together possibly with some morphs which define additional distortions to be applied to the avatar.

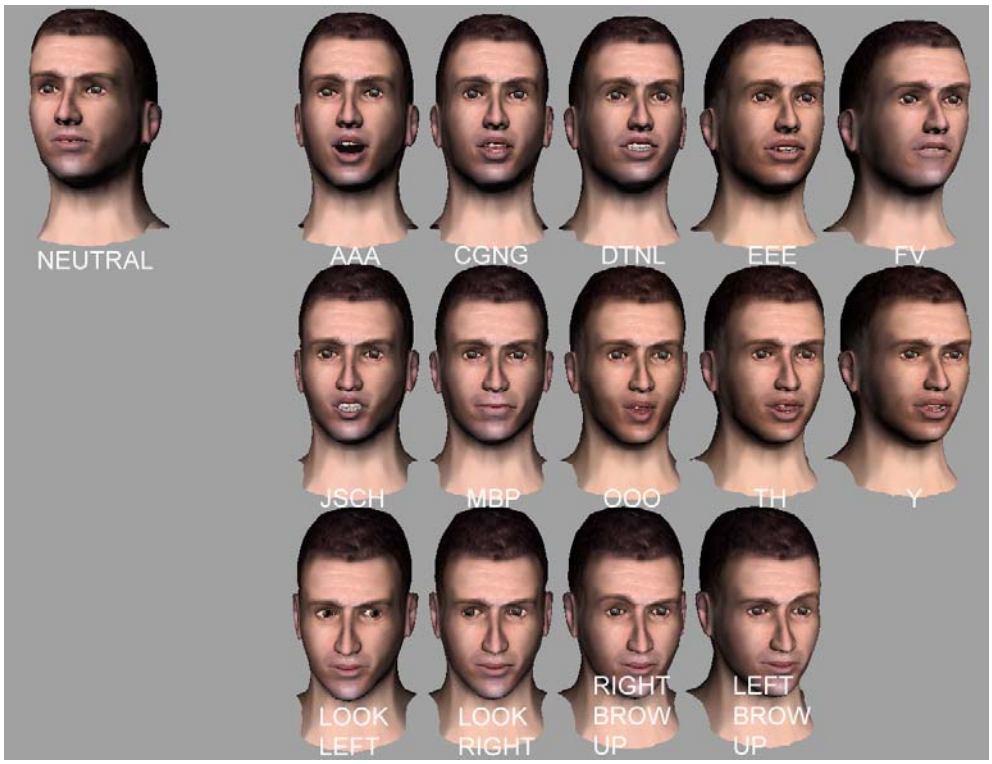


Figure 3: Avatar facial morphs.

In order to make an avatar sign, pre-specified animation sequences must be sent to the avatar. There are two possible approaches to the generation of such animation sequences; motion capture or synthetic generation. At the start of the earlier ViSiCAST project, motion capture was used exclusively to generate signed content. This involves the use of magnetic and video-based tracking techniques to record the motion of a real human signer. The recorded data can be segmented, resulting in a set of motion data files constituting a lexicon of signs. These can then be replayed through the signing avatar on demand. The TESSA system, developed in the ViSiCAST project to support transactions between a counter clerk and a deaf customer in a Post Office was based on this technology.

However, the resource costs involved in the use of motion capture technology in this way are relatively high and the flexibility of reusing signs from existing sequences is limited. These factors motivated the development in the ViSiCAST project of a radical alternative approach in which a signed animation is generated *synthetically* from an input script in the SiGML notation.

SiGML (Signing Gesture Markup Language) is an XML application which supports the definition of sign sequences. The synthetic signing system is driven by "Gestural-

SiGML", a subset of the full SiGML notation, which is based on the HamNoSys notation for sign languages[9].

The basic avatar animation and rendering technology used in eSIGN is supplied by Televirtual. This consists of the animation and rendering software embodied in Televirtual's Mask-2 system, together with the character definitions of the VGuido and Visia4 avatar characters. Each of these avatar definitions incorporates a detailed description of the relation between the avatar's skeleton and significant sites on the surface mesh: this description is essential to the generation of precise signing data. VGuido has been developed specifically for eSIGN, after informal consultation with sign language users in Germany and the UK. It would be possible, but expensive, to create different avatars, with different clothing, hair, etc.

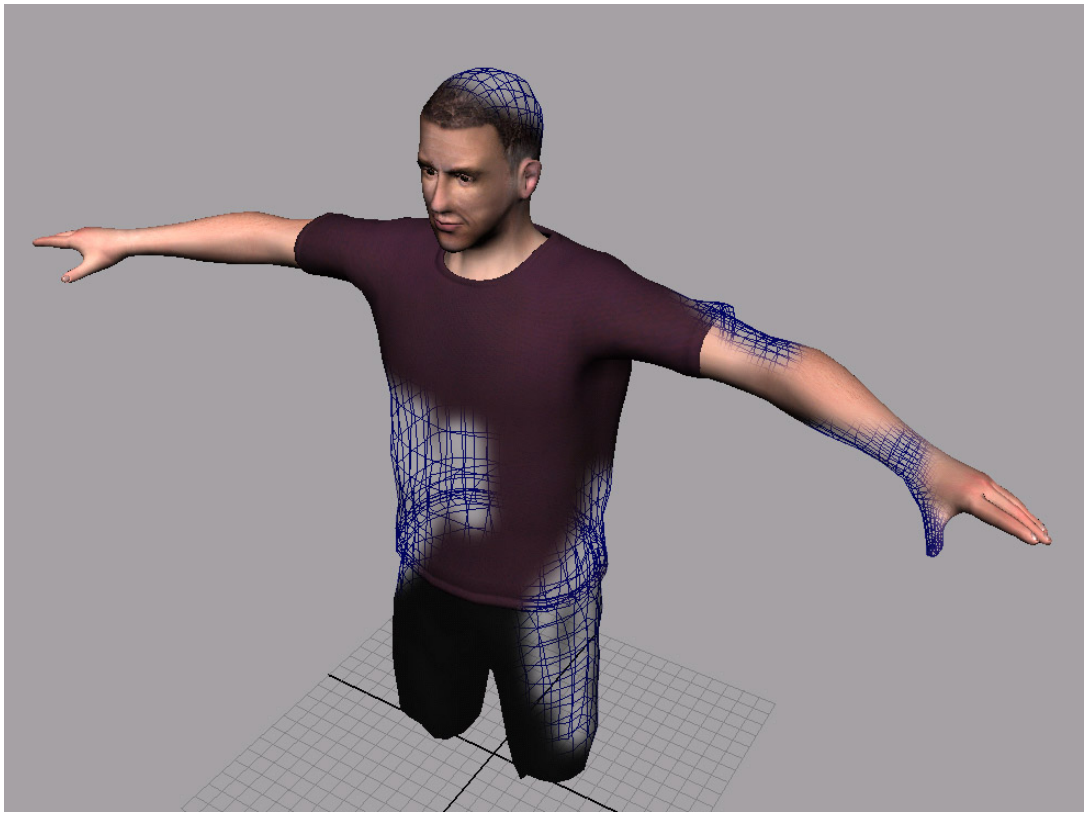


Figure 4: The VGuido avatar, developed by Televirtual for the eSIGN project.Some of the surface mesh has been cut away to reveal the underlying structure.

The basic animation and rendering technology supplied by Televirtual has been augmented by the SiGMLSigning software developed at UEA. This provides an ActiveX control and SiGMLToSigningAvatar, which may be loaded into an HTML page. These provide an interface whose main functions are to: support the input of a signing sequence defined in SiGML; support the generation of the corresponding sequence of animation frames; schedule the rendering of these frames using the Mask-2 ActiveX controls provided by Televirtual. (In fact SiGMLSigning can be used to drive any avatar which supports a simple rendering interface similar to that provided by the Mask-2 system.)

[9] 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 Gernam Sign Language and Communication of the Deaf, University of Hamburg, 1989.

In addition to the fundamental animation capabilities just described, the SiGMLSigning and Mask-2 systems provide the HTML developer with scriptable control of system features such as the virtual camera position and viewing parameters, the lighting, and the avatar's ambient motions.

4.2 Animation for Signing Avatars

This section describes the central concepts that are necessary for animating avatar technology in the production of sign language. There are three main options available, motion capture, key frame animation and script driven generative animation (the one chosen for eSIGN).

Motion Capture

Motion Capture or Gesture recognition has been used for the development of Signing Avatars since the mid 90's. Broadly speaking it falls into two areas:

(a) Magnetic Tracking or Capture.

(b) Optical tracking or Capture.

(a) Magnetic: Sensors are attached to various joints locations in the upper torso, arms and head. Typically between 15 and 18 sensors would be employed. Such an assembly of sensors and wires is commonly termed a Capture Suit, even though it does not resemble a garment. A large magnetic field generator is then employed to create a 'capture area', generally of 4 X 4 metres, although multiple field generators may produce larger areas. To address the complexity of signing, Data gloves must also be deployed in combination with a magnetic capture suit, and a separate facial tracking system must also be worn in order to recognize and record augmented facial expressions. The resulting positional and rotational movements are then recorded and stored in a database for replay. together with data from the gloves and face tracker.

- Advantages: The process is real-time and may be monitored on the fly. Capture process is loyal. Resultant movement is high fidelity and natural.
- Disagdvantages: Multiple capture devices. Larger amounts of data defy both processing and storage. Asymmetric (unique) moves demand complex blending algorithms to produce a usable lexicon. Specialised operators needed. Data is often noisy and post processing can therefore be necessary. Each avatar requires a mapping for each location on the body to a joint in the avatar's skeleton.

(b) Optical: Retro reflective markers are attached to joints locations on the arms and torso. A subset of markers are attached to key points on the face and hands. A system of multiple static cameras, operating in the infra-red spectrum with ultra fast shutter speeds, tracks all markers by pulse firing light at them. The latest optical systems operate in real time, but the complexity of capturing human signing tends to dictate an off-line process.

- Advantages: Very high fidelity results, even though occlusion issues make hand capture problematical.
- Disadvantages: Large amounts of data, big post processing overheads, high Equipment costs, high equipment operation costs, specialised staff needed.

Key Frame Animation

A deformable mesh and bones system is manipulated to a variety of target positions to mimic a human signer using procedural modelling tools such as Alias Wavefront MAYA.

- Advantages: Relatively small file sizes result in a more efficient database for storage and recall.
- Disadvantages: Heavy modelling task overheads, medium Fidelity (may seem mechanical) results.

NB: Interpolation between keyframes is a non trivial exercise for naturalistic continuous motion. Large quantities of keyframed animation would require a number of skilled staff or a long lead time and is therefore expensive.

Script Driven Generative Animation

Generally known within eSIGN as synthetic animation this is the creation of sign language from scripted descriptions of signing motions. This is the approach adopted for eSIGN.

- Advantages: Auto-generation, small database, fast retrieval, web distributable, streamable datasets, graphics computing task only, text-to-sign system viability.
- Against: Low - Medium Fidelity, heavy computing task, creating naturalistic motion is a non-trivial task.

It is also worth mentioning the portability of the various techniques. Motion capture and key frame animation are avatar specific. While it is probably possible to generate new data sets for new avatars this would take time, skill and of course double the data storage requirements. It may be possible to convert data on the fly, but this could prove problematic. One of the big advantages of script-driven (synthetic) animation is that converting the motion for use with a new avatar requires a small amount of configuration data, making it easily the most avatar independent.

As described above, the “gestural” subset of the SiGML notation allows a sign to be defined in a form suitable for input to the synthetic signing animation software. However SiGML also allows a sign to be defined directly as animation data in Character Animation Stream (CAS) format. Tools are available which will convert motion capture data in Televirtual’s BAF format (used in the ViSiCAST project) to CAS format. Thus the eSIGN project’s SiGMLSigning software also supports the use of motion captured data in the generation of signing sequences.

4.3 Synthetic Animation of Signing

The motion capture approach has traditionally been used in the production of sign language. However, as discussed above this approach still requires that each and every sign occurring in a signed text to be recorded, a time-consuming process when building a lexicon of signs for a given sign language.

Synthetic animation, on the other hand, constructs human-like motion from scripted descriptions of signing motions, in our case a superset of HamNoSys, a well-defined notation system for sign language transcription. Due to sophisticated models of human

motion used, the resulting performance of synthetically generated sign can come close to the quality of motion-capture play-back in terms of recognisability of the signs.

The concept of synthetic animation used in eSIGN is to create scripted descriptions for individual signs and store them in a database. While populating this database may take some time signed phrases can be made quite quickly by selecting the required signs from the database and assembling them in the correct order. Contextual issues may mean that some modification of the signs used is required, but generally phrases can be created quite quickly.

The major advantage of this approach lies in its flexibility: The lexicon-building task does not require special equipment, but only a database. The morphological richness of sign languages can be modelled in a sign language editing environment without the need to manually describe each inflected form.

How does such a formal description of a sign look like? Let us take an example from German Sign Language meaning “house”:

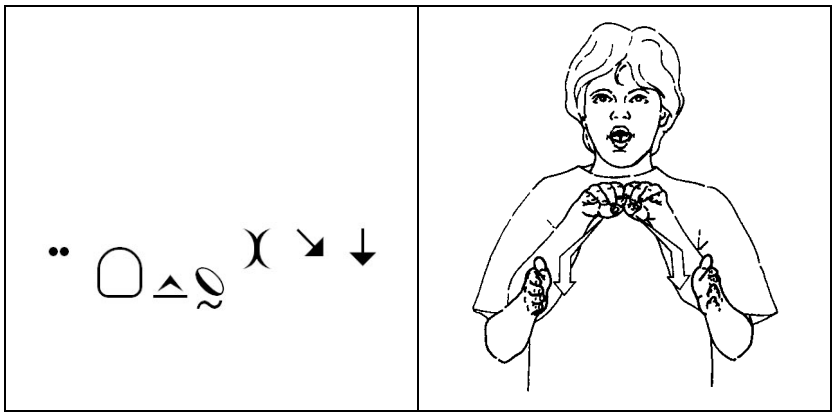


Figure: 2 HamNoSys notation and illustration of the DGS sign for ‘house ’ (The two flat hands indicate a roof and a wall each).

The first symbol signals that the sign is two-handed symmetrical. The second symbol means a flat handshape, followed by two symbols (plus a subscript) for the orientation of the right hand at the beginning of the sign. The description of the initial posture concludes with a contact symbol, meaning that the two hands are in contact with each other. Two arrows finally describe the movements to be performed. This sign is often accompanied by a mouthing that is derived from the German word “Haus” (house), so a description of the sequence of visemes (visible phonemes) for “Haus” is stored in the lexical entry together with the manual activity. (Other lexical entries have mouth activities unrelated to spoken language or none at all. Other classes of nonmanual activities occur in signing, but are not typically parts of lexemes.)

If this lexeme becomes part of a sentence, it may take inflections, in this case e.g. modifying the location in signing space or the size of the object referenced. The HamNoSys notation for some types of modification can be computed on the fly, so there is no need for separate lexicon entries.

SiGML, the avatar's “native language” is an XML encoding of all the information needed to describe signs. For our example entry, the SiGML looks as follows:

```
<sigml>
   <hamgestural_sign gloss="HAUS3">
     <sign_nonmanual>
       <mouthing_tier>
         <mouth_picture picture="haUs"/>
       </mouthing_tier>
     </sign_nonmanual>
     <sign_manual both_hands="true" lr_symm="true">
       <handconfig handshape="flat"/>
       <handconfig extfidir="o"/>
       <handconfig palmor="dl" rel_palmor="true"/>
       <handconstellation contact="touch"/>
       <directedmotion direction="dr"/>
       <directedmotion direction="d"/>
     </sign_manual>
   </hamgestural_sign>
</sigml>
```

HamNoSys and other components of SiGML mix primitives for static gestalts (such as parts of the initial posture of a sign) with dynamics (such as movement directions) by intention. This allows the transcriber to focus on essential characteristics of the signs when describing a sign. This information, together with knowledge about common aspects of human motion as used in signing such as speed, size of movement, etc., is also used by the movement generation process to compute the avatar's movements from the scripted instructions.

4.4 Why not Automatic Translation?

Once we have a machine-readable lexicon and some formal means to apply morphological operations, it would be nice to have text written in English automatically translated into, say, British Sign Language. Unfortunately, sign languages are quite different in their grammar and lexicon from most European spoken languages. So if you consider how restricted a domain must be even for translation from English into German to work reliably, it is clear how difficult this task would be for translating into sign language. Moreover, research on sign language only started in the 1960s, and cannot answer all the questions necessary to create useful computational linguistics models of sign languages. (Due to the lack of written forms of sign languages, a statistical approach on translation is also far from reality.)

In eSIGN, we have therefore taken a split approach:

- For information sources built on a very restricted domain, we have designed a language model for the required (small) subset of the target sign language, allowing automatic sign language production.
- For all other kinds of information, spoken language text is translated into sign language by a human translator. eSIGN tools are used to make this translation task as easy as possible.

5 Next Steps

The Objective of this document has been to describe the potential for the eSIGN technology and the wider picture of its use in providing sign language content. Details of the licensing and typical costs are set out below:

5.1 Licensing Model

The end user software, i.e. the avatar player, is freely licensed. The editing tools require a licence for use, however, this is free of charge for uses such as research, education and non for profit application. For commercial use license arrangements must be made, details can be found on the eSIGN webpage.

5.2 Content Creation Costs

As shown in this document content creation has two basic elements, the creation of individual signs (to be placed in the lexicon) and use of signs from the lexicon to produce signed phrases. Creating a sign takes considerably longer than extracting a sign for use in a particular phrase (hence the lexicon approach).

As such, the cost of creating a sequence of signed content depends on how many of the signs required are already available to use, and how many must be created from scratch. The size of the available lexicon, and the subject areas that it covers, have the greatest effect on this. Over time, as more content is created and the size of the lexicons is increased, the cost of content creation will fall.

The tables below set out the resources required to create new signs, and those required to create signed phrases using available signs. These figures are for trained staff with experience of the tools.

Creating a new sign	• Production: approximately 4 signs/hour (average figure, varies according to the complexity of the sign)
Content creation from existing signs	• Two methods – direct: text to synthetic signs – indirect: text via SL-video to synthetic signs • Costs about the same: approximately 20 signs per hour
Cost to translate between spoken languages	• Approximately one Euro per line German to Chinese • 21 - 47 Euro per page English to Chinese • 28 - 47 Euro per page English to Swahili The translation for eSIGN content will therefore be a little bit higher.

5.3 Further information

For more information, and references to sites containing eSIGN content, please refer to the project website at www.sign-lang.uni-hamburg.de/eSIGN. There you will also find the links to download the eSIGN player software and the eSIGN authoring kit.

5.4 Acknowledgements

eSIGN: Essential sign language information on government networks, was funded by the European Union eContent programme project EDC-22124 ESIGN 27960.

This document is based on eSIGN Deliverable D7-2 written in August 2004 and contains contributions from *Uwe Ehrhardt, Bryn Davies, Neil Thomas, Mary Sheard, John Glauert, Ralph Elliott, Judy Tryggvason, Thomas Hanke, Constanze Schmaling, Mark Wells, and Inge Zwitserlood.*