SYNTHETIC SIGNING FOR THE DEAF: eSIGN

Inge Zwitserlood, Margriet Verlinden, Johan Ros, Sanny van der Schoot

Department of Research and Development,
Viataal, Theerestraat 42, NL-5272 GD Sint-Michielsgestel, The Netherlands
Tel. +73 55 88 489, Fax +73 55 17 879, Text phone +73 55 88 328, E-mail
{i.zwitserlood,m.verlinden,j.ros,s.vdschoot}@viataal.nl

Abstract: This paper illustrates the use of synthesis in providing information in sign language to those who are deaf or hearing impaired, particularly on the Internet. It gives a detailed discussion of the techniques involved and their development. The software will be illustrated during the presentation.

Keywords: sign language, synthesis, animation, deaf, hearing impaired, information

1 Background

1.1 Deaf people's language skills

Hearing children have full access to the language(s) spoken around and to them. Through a combination of spoken input and children's innate language capacities, they gradually acquire these languages, starting right from the beginning. For many people who are born deaf or become deaf at a very early age, however, it is impossible to fully acquire a spoken language since the input is extremely small; i.e. the visual information of a spoken language is limited to what one can see on the outside of the mouth. These people need explicit instruction in order to learn spoken language. Obviously, this cannot start as early as the language acquisition of hearing people. Consequently, the spoken language skills of many deaf people are usually (far) less than that of hearing people. In contrast, acquisition of a signed language can take place in the same, natural way as acquisition of a spoken language by a hearing child, provided that it has sufficient input, of good quality.

Sign languages are genuine, natural languages. In contrast to what is widely believed, there is not one universal sign language, but many. Sign languages are not derived from spoken languages, neither are they invented. Sign languages develop in interaction of deaf people, and they have their own grammar and lexicon. They are not mutually comprehensible (although related sign languages can have similar structures).

1.2 Information flow

Much information (notably information on the internet) is provided in written form. The written form of most spoken languages is based on their phonology and phonetics. For hearing people, it is fairly easy to learn to code the spoken language (i.e. writing) and to decode the written form (i.e. reading). In contrast, this is much more difficult for many deaf people. Not having access to the phonetics, and (consequently) not having full spoken language skills, their command of the written form is often rather poor, even though it is clearly visible. Many deaf people are functionally analphabetic and thus miss large amounts of information. The only way to give these people accessible information is to provide it in a sign language. If there were a generally accepted notation system for sign languages, signed information could be provided in written form alongside the written form of the spoken language. Such a system lacks, however. In the few cases where information in sign language is given, this is done by a movie of a person signing, for instance on the internet or on a CD-ROM.

1.3 Limitations of presenting sign language information on video

Movies of signed content have four serious disadvantages. First, if there are changes in the content, it is impossible to simply erase and/or add parts of the video; the entire video has to be recorded anew. Second, a person signing information for a movie (especially long stretches of content) has to be very alert during the recording; every time s/he makes a mistake the particular part of the movie has to be resigned and rerecorded. Third, sign languages are three-dimensional languages. Since movies are merely two-dimensional, information may be lost in movies of signed text. These problems are obviated with the animation technique, that will be explained below. Finally, movies take a tremendous amount of memory capacity. Downloading movies from the internet takes a long time, and a CD-ROM can only contain a quite limited span of signed content without losing too much quality.

2. Animation driven by captured motions

The technology reported on here also can be used to provide information in sign language; not by a video, but by a signing computer animation: an avatar. The first project in which means for the provision of animated sign language were developed by seven companies and universities in France, Great Britain, Germany, and the Netherlands, was the EU-subsidised ViSiCAST project (Virtual Signing: Capture, Animation, Storage and Transmission). In this project the technique of motion capturing was used. A real signer makes separate signs, wearing special equipment, viz. a motion capture suit. This suit consists of i) polhemus magnetic sensors on the body (including the head and arms), that record the wrist, upper arm, head and upper torso positions in three dimensional space relative to a magnetic field source; ii) two cybergloves (each having 18 resistive elements), that record the finger and thumb positions relative to the hand; and iii) a video face tracker, that records facial expressions. This trackers consists of a helmet mounted camera with infra-red filters surrounded by infra-red light emitting diodes that illuminate 18 reflectors positioned on the face (notably near the mouth and eyebrows). The suit is illustrated in Figure 1.

Figure 1, Signer in motion capture suit

The three separate data streams are synthesised into a single motion-data stream. This stream can animate a virtual human (an 'avatar') directly. (For more details, see Elliot *et al.* 2000; Kennaway 2001.)

2.1 Animation

A special avatar was developed in ViSiCAST, called 'Visia'. As all avatars, she consists of a skeleton and skin. The skeleton consists of a set of 'bones'. An avatar bone consists of two markers that mark the beginning and end of a bone and a third marker that defines the orientation of the bone; as distinct from a human bone. The bone structure of an avatar and the outer appearance of Visia are illustrated in Figure 2.

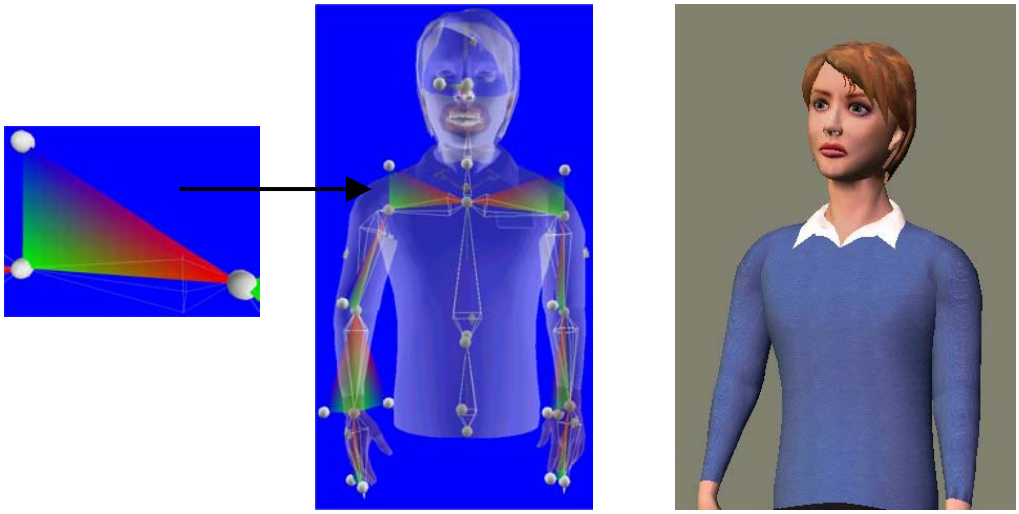


Figure 2, Bone - in skeleton of avatar - outer appearance of avatar 'Visia'

Playing a motion-data file of a sign animates the skeleton of an avatar into making the motions in the sign, whereby the skin of the avatar is constantly and automatically adjusted to fit the bone constellations.

Signing avatars have several advantages over movies of signed content. First, the HTML pages needed for the avatar (containing merely play lists for the data files) are significantly smaller than movies files and thus the download time is merely a fraction of the download time of a movie. Second, the avatar, although presented on a flat screen, is three-dimensional: it is not fixed to a particular position. The avatar can be turned, enlarged and reduced by the onlooker, even while it is signing.

2.2 Creation and evaluation of signed content

The motion capture process was used in ViSiCAST for capturing signs in three different sign languages (British Sign Language [BSL], German Sign Language [DGS], and Sign Language of the Netherlands [NGT]). A limited domain was chosen, viz. the weather. For each sign language approximately 350 signs in this domain were recorded and stored in data files. In real signing, more than single signs are needed: phrases, sentences and longer stretches of signed text. Thus, the recorded signs needed to be combined to form grammatical sentences in each of the sign languages. The ViSiCAST partners developed a semi-automatic translation tool, with which a restricted set of phrases and sentences in Dutch, English and German can be translated into the restricted set of phrases and sentences in NGT, BSL and DGS, respectively, that are needed for the weather forecasts. A complete translation of a weather forecast is actually a play list of data files (recorded signs). It is stored as an HTML page. A users needs to install the avatar software and the motion-data files of the signs on his computer in order to see the animation-signed content of that HTML page.

In the Netherlands, the animated weather reports were evaluated with nine deaf subjects. The understandability was tested objectively (60%) and subjectively (reasonable - good) (Verlinden *et al.* 2001). Nevertheless, it was clear that improvements were needed with respect to the appearance of the avatar, the mouthing and mimicry, and the fluency within the sign sequences.

2.3 Necessary improvements

Notwithstanding the preliminary positive reactions of users, further technical improvement were needed, since the motion capture technique itself turns out to have serious drawbacks for use with signs. First, the equipment used for capturing signs is not accurate enough for the very specific and sometimes very small movements of the hands and fingers. As a consequence, most signs need to be modified after capturing, a time consuming process. Second, setting up and calibrating the equipment for every recording session is difficult and very time consuming. Third, signers are not comfortable wearing the necessary equipment. Finally, extension and adaptation of the signed lexicon is difficult: since the equipment is very costly it is not readily available, and once stored, a sign cannot be adapted with respect to linguistic variation (inflection) or for prosody. Therefore, it was concluded that it was necessary to develop the technology for genuine synthetic signing.

3. The second stage: animation driven by notation

The synthetic signing technique was further developed by seven companies and universities in Germany, Great-Britain and the Netherlands, starting in ViSiCAST en subsequently in the - also EU funded - project eSIGN (Essential Sign Language Information on Government Networks). The new avatar developed in eSIGN by Televirtual Ltd is called 'virtual Guido'. Signs are no longer captured. Instead, they are transcribed in a phonetic transcription system developed for sign language at the University of Hamburg: HamNoSys (Hamburg Notation System, Hanke 2002) and a font was designed for use on the computer. With this system all of the sign components can be given specific values. It is not possible for computers to process the transcriptions as such and these, therefore, need to be translated in a language that is amenable to computer processing. To this aim, an XML encoding of HamNoSys was designed at the University of East Anglia: Signing Gesture Markup Language or SiGML, and a translator was written from HamNoSys to SiGML. Together with a description of the geometry of the avatar, animation data can be sent to the avatar, who then makes the requested sign(s) (see Kennaway 2003 for more details). A great advantage with respect to motion capture is that users only need to install the avatar software on their computers; since avatar signing is driven by the combination of SiGML (on particular web pages) and the avatar descriptions (in the avatar software), it is not necessary to install a database of signs and reinstall a new one every time new animated signs become available. An example of the HamNoSys transcription of an NGT sign and its translation in SiGML is given in Table 1.

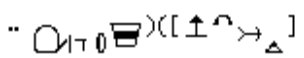
HamNoSys		**SiGML**
	→	<hamgestural_sign gloss="Ham2HPSG"> <sign_manual both_hands="true" lr_symm="true"> <handconfig handshape="flat" thumbpos="out"/> <handconfig extfidir="uo"/> <handconfig palmor="l"/> <location_bodyarm location="chest" contact="close"/> <par_motion> <directedmotion direction="o" curve="u"/> <tgt_motion> <changeposture/> <handconfig extfidir="o"/> </tgt_motion> </par_motion> </sign_manual> </hamgestural_sign>

Table 1 HamNoSys transcription and SiGML translation of the NGT sign for 'to visit'

Handshape (flat hand with spread thumb) is expressed by  in HamNoSys, orientation of fingers and thumb by the respective symbols  and . The initial hand position is near the chest () and the movement consists of a curved forward movement (), with a simultaneous change in finger orientation (). The diaeresis (¨) indicates that the non-dominant hand mirrors the dominant hand. The sign, as performed with the eSIGN avatar ('virtual Guido'), is illustrated in Figure 3.

Figure 3, Avatar 'vGuido' makes the citation form of the NGT sign for 'to visit'

The HamNoSys transcription system originally focused on the manual part of signs. Additional transcription symbols for the nonmanual part was developed in ViSiCAST, since the nonmanual part of signing is as

important in signing as the manual part. The transcription system for the nonmanuals contains codes for several positions and movements of body parts (shoulders, back, chest), head, and face (eyebrows, eyes, nose, lips, tongue and lower jaw). These are currently implemented in eSIGN.

3.1 Creation of synthetic signed content

The main aim of the eSIGN project is to provide Deaf citizens in the partner countries with information of the local government on the internet, in sign. This information is usually already provided in writing and thus needs to be translated into the respective target sign languages. For various reasons, development of automatic translation programmes goes far beyond the scope of a single project. First, the current knowledge of the grammar of the sign languages involved is insufficient. Second, even automatic translation texts from written form of one language into the written form of another language is problematic still. Third, translation here is not from text to text, but from text to sign synthesis (comparable to translating written English into spoken Russian). Prosody thus needs to be implemented as well, and in a different (signed) modality.

Instead of automatic translation, two ways are developed to create the signed content. First, tools are developed in which simple translations are contained, in the same manner as the weather forecast creation tool in ViSiCAST. For obvious reasons these concern limited domains. Such tools can be used by anybody who commands the spoken language: structured signed content is made on the basis of a lexicon and a set of preconstructed phrase and sentence models of the spoken language, that are linked to a lexicon and a set of preconstructed phrase and sentence models of the sign language.

Second, sign language users can construct content on the basis of the knowledge they have of their own language. Content can be either translation of written text, or new signed text. The person creating content uses an editing programme that is developed by the University of Hamburg, in which s/he can retrieve signs stored in a database, put them in the right order and provide the sign strings with the correct prosody. This is comparable to writing a text on a word processor in that all that is needed are the language skills, the necessary software and the skills to use the software. Advantages of construction of synthetic signed content to construction of movies of signed content are that the person creating the signed content can do this at his/her own pace. S/he is not dependent on the availability of a studio or a camera person. Repairing mistakes and adapting signed text is rather easy in comparison to making and editing movies of signed text.

3.2 Synthetic signed information in the Netherlands

In the Netherlands, information content is created with regard to legislation about labour and job support for the deaf population. Three applications are made. First, the explanation of the Dutch law on reintegration of the labour handicapped is translated. Second, summaries of a set of job vacancies (viz. the vacancies at Viataal) will be presented in animated NGT on the Viataal web site. Third, translations or explanations of forms that deaf people often need to fill out have been made, such as an application form for interpreting services, a form for reimbursement of travel expenses, a form for empowerment of attorney and requests for additional schooling or a trial job. All forms but one need to be filled out by hand. A deaf person will be able to download the web page with the particular form he needs to fill out and, by clicking on a hyperlink or on a particular icon, let the avatar sign the required information. This is illustrated in Figure 4.

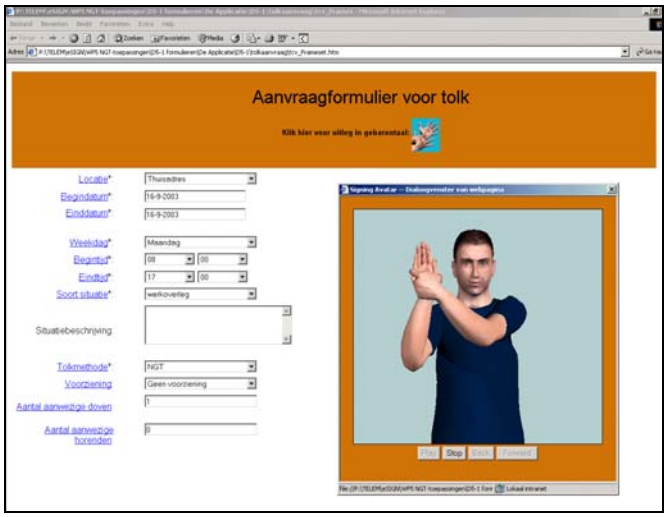


Figure 4, Example of web page with form containing information in Dutch and NGT

A first evaluation showed that implementation of the non-manual part needed much improvement still, and adaptations needed to be made to the appearance of the avatar.

So far, content creation is done by experts, at Viataal. In a follow-up project, Gebarennet ('Sign web') (funded by the Dutch ministry of Health, Welfare and Sport) we plan to make the technique available to deaf users in the near future. The avatar software and (a simplified version of) the editor software will become available to the deaf public and lessons will be developed in which sign language users will be taught how to create signed content on the basis of a given lexicon of approximately 4.000 NGT signs. Furthermore, we will encourage companies and clubs to provide the information on their web pages in synthetic NGT and we intend to provide translation services.

4. Conclusion

The technology developed in ViSiCAST and eSIGN will enable a wide use of synthetic signing. The applications developed in the project, aiming to help deaf people to gain better access to government information and to improve their employment situation, are a first step to this overall aim. As soon as the technology is ready for use by non-specialists and made available to both information providers and the target group of deaf people, this will result in better access to all kinds of information for deaf people.

Acknowledgements

We are grateful for funding from the EU, under the Framework V IST Programme (Grants IST-1999-10500 and the eContent Programme (EDC-22124 ESIGN/27960). We are also grateful to our collegues at the partner institutions (Televirtual Ltd., University of East Anglia, Institut für Deutsche Gebärdensprache at the University of Hamburg, Norfolk County Council, Royal National Institute for the Deaf, and Systematics Integrations) for their support.

References

Elliot, R. J.R.W. Glauert, J.R. Kennaway, I. Marshall (2000) The Development of Language Processing Support for the ViSiCAST Project. In: *4th International ACM SIGCAPH Conference on Assistive Technologies (ASSETS 2000)*, pp. 101-108

Hanke, T. (2002) *ViSiCAST Deliverable D5-1: Interface Definitions*. Manuscript

Kennaway, R. (2002) Synthetic Animation of Deaf Signing Gestures. In: *Lecture Notes in Computer Science* Vol. 2298, pp. 146-157

Kennaway, R. (2004). Experience with and requirements for a gesture description language for synthetic animation. In: *Lecture Notes in Computer Science* Vol. 2915, pp. 300-311

Verlinden, M., C. Tijsseling & H. Frowein (2001). Sign Language on the WWW. In: *Proceedings of the 18th International Symposium on Human Factors in Telecommunications (HFT2001) in Bergen, Norway, 5-7 November 2001*, pp. 197-2004