In: Proceedings of 18th International Symposium on Human Factors in Telecommunication (HFT 2001) in Bergen, Norway, 5 - 7 November, 2001

Sign language on the WWW

Margriet Verlinden, Corrie Tijsseling and Han Frowein,
Instituut voor Doven, Research/Development & Support, Theerestraat 42,
5271 GD Sint-Michielsgestel, the Netherlands
{m.verlinden, c.tijsseling, h.frowein}@ivd.nl

Abstract

The work presented is part of the EU-subsidised ViSiCAST project. The project aims to achieve semi-automatic translation from text to sign language. With the use of a computer, sentences in a written language are analysed and transposed to a sequence of signs. These signs are displayed using computer animation of a Virtual Human, i.e. an "avatar". One of the application areas for this technology is the world-wide-web. As a prototype application a web page has been developed which shows an avatar signing the weather forecast in sign language. Evaluation of this prototype application has given useful information about the quality of the signing avatar.

Key words: animation; avatars; sign language; semi-automatic translation; Internet; world-wide-web.

1 Sign Language and its Users

For people who are born deaf or have become deaf before learning a language, it is very difficult to speak and to read and write (King & Quigley, 1985; Paul & Quigley, 1994). Hearing children always learn their first language by hearing. Subsequently, they learn how to read, based on phonetics. This way one can imagine that very often the spoken language used around a deaf child, will not become its first language. Because of the limited perception of that language, learning all characteristics of that language in order to understand and use it to its full capacity is not possible. This is entirely different for Sign Language. Any well sighted child has full perception of this visual language, and can therefore learn this language in the same, natural way as hearing children learn a spoken language: by perceiving and imitating over and over.

Sign Language is a complete and genuine language in all its aspects. Everything that can be expressed in spoken language can be expressed in Sign Language as well. Sign Language also has its own grammar. There are very specific rules how a particular sign is performed, how signs are altered (=inflected) in a certain context, and how signs are combined to form full sentences. Speaking about Sign Language as one language is in fact not correct because there are many sign languages. There are about as many different sign languages as there are countries. Examples are American Sign Language, British Sign Language, Norwegian Sign Language and Sign Language of the Netherlands. Each of these sign languages have there own signs and there own grammar.

2 Application: Signed Weather Forecast on the Internet

In the EU-project ViSiCAST (www.visicast.co.uk), we aim to develop techniques and applications to present various information sources in animated sign language. The application areas focussed upon in the project are television, human-human contact in public services, and multimedia &

Internet. As opposed to video recordings, animated sign language can more easily be updated and requires less bandwidth to transport to the user. The sign language is performed by a virtual human, i.e. an avatar. The avatar designed for ViSiCAST is called 'Visia'. The first application that was developed for the Internet consists of a web page that can contain most simple weather forecasts in three sign languages, and that can easily be updated every day without recording new signs. Below is a picture of the Dutch version of this web page.

Figure 1. Web page with a weather forecast in text and in sign language, performed by the avatar 'Visia'

The design of the user interface was based on an inventory of user requirements by teachers of deaf computer users. The weather forecast is presented both in text and in sign language, for those users for whom both languages are complementary. The text and avatar Visia are presented in one page, in order to avoid having to switch between the two forms of the forecast. The buttons to control the avatar are kept to a minimum, and the look and feel corresponds to usual interfaces such as on video recorders and in movie-player software. In general, the interface was kept simple with the aim of making the system clear and pleasant to use for all users, including deaf people with limited literacy abilities.

The user needs a browser for the Internet, and Visia-avatar software that functions as a plug-in to the browser. When the user visits the page and requests Visia to sign the weather forecast, Visia subsequently performs the sign motions on the user's own PC by means of this plug-in. The list of motions to be performed is coded into a special mark-up language for signs so that it can be part of an HTML-web page.

The content for the web page is based on the forecast written by a Dutch meteorological institute. This free-form Dutch text is semi-automatically converted to sign language. First the sentences are manually converted to standard sentences. The standardised sentences are then automatically

converted to grammatically correct sign language by means of several lookup tables. The standard sentence forms are not entirely fixed; they contain up to 4 variables each. In combination with a large set of possible values for the variables these flexible sentences make it possible to express any weather forecast in standard sentences.

The set of standard sentences and the set of possible values for the variables determine which sign motions are needed for this avatar application. All these motions have been performed by a human signer and were recorded (hence 'captured'). The motions were captured in detail with tracking systems on the body for facial expressions, for posture, and for hand and finger shapes. The signer's movements were captured at a very high resolution, far higher than is usually used for animation or video games. This is important for the generation of animated sign language, where the slightest difference in speed, direction or gesture has a bearing upon meaning. Signs have been captured for Sign Language of the Netherlands, German Sign Language and British Sign Language, each with a native signer of that language.

The captured signs can be re-performed by Visia one by one separately, but also in a fluent sequence. Significantly, real-time visualisation software blends one captured sequence into another, creating a smooth signed performance – something that would be unachievable with joined-up clips of video. In the ViSiCAST weather forecast application, it allows sections of signing to be built-up which include variables, such as temperatures, weather types and wind directions. Another difference with regular video is that the avatar is really 3-dimensional. The user that is watching this avatar can enlarge, reduce and turn it. This is done with simple mouse movements, and can be done both when the avatar is moving and when she is standing still. This is an important advantage because forward movements in sign language are sometimes difficult to perceive when presented in 2 dimensions.

Prototypes of the application described in this section are now realised for the three captured sign languages and their corresponding written languages. The English and Dutch prototypes have been evaluated by users. The next section describes the Dutch evaluation.

3 Evaluation

The aims of evaluating the prototype weather forecast application are:

1. Assess the level of comprehension of animated weather forecasts.
2. Specify any animation features that need to be improved before proceeding the application on the Internet.

3.1 Method

The evaluation was carried out by a deaf researcher who is fluent in Dutch as well as in sign language. Subjects were approached by the experimenter during face-to-face or personal e-mail communication and informed about the ViSiCAST project and the experiment in sign language. If they were willing to participate they were asked whether they prefer to get information in sign language, in text, or whether they had no preference. Nine subjects with a preference for sign language or with no preference were selected. They consisted of 4 men and 5 women, and their age ranged from 20 to 53 years.

The material consisted of 2 weather reports for demonstration and 5 weather reports for the experiment. The reports represent a range of commonly occurring weather conditions and each report has at least two examples of the following information-containing features of sign language, which - on initial inspection - were selected as possible causes of comprehension errors:

- finger spelling
- mouthing
- facial expression
- movement
- location
- size
- handshape

The weather reports were always presented without text.

The subjects were tested individually. They were seated in a swivel chair facing a PC with the avatar presented in a window on the screen; the size of this window was adjusted to suit the subject's preference. During the experiment, the experimenter operated a video camera and was positioned on the side of the subject at a distance of about 2 meters.

Each weather report was broken up into sentences and presented as such to the subject. After a sentence was presented, the subject's task was to repeat that sentence in sign language, facing away from the screen and towards the experimenter with the video camera. After the two demonstration reports were presented to familiarise the subject with the avatar and the experimental task, the five experimental reports have been presented and the order of presentation was counter balanced between subjects.

After the last report, subjects were asked for general comments on the avatar and to rate the quality of the avatar's signing by putting a mark on continuous 3-point scale from very good to very bad. All of the subject's signing has been recorded on video, including spontaneous comments and the interview afterwards.

3.2 Comprehension scores

The evaluation resulted in 9 videotapes of approximately one hour each. The video tapes were analysed by the experimenter with the aid of score sheets. For each subject, each sign was scored as correct, partially correct (e.g. 'in the west' instead of 'from the west'), incorrect, or missed.

The overall scores averaged across subject were:

61 %	correct
7 %	partially correct
12 %	incorrect
20 %	missed

In order to get more insight in the causes of errors, the cumulative scores were calculated per report per person, and cumulative scores were calculated for each sign occurring in the test material. Based

on rough inspection of all scores, and on the behaviour and the remarks of the subjects, we arrived at the following hypotheses:

1. There was a learning effect, i.e. the scores improved over a session.
2. Signs that are mainly dependent on the avatar's mouthing were less well understood than signs that do not involve mouthing.
3. Signs that are mainly dependent on facial expression (mimicry) were less well understood than signs that do not involve facial expression.

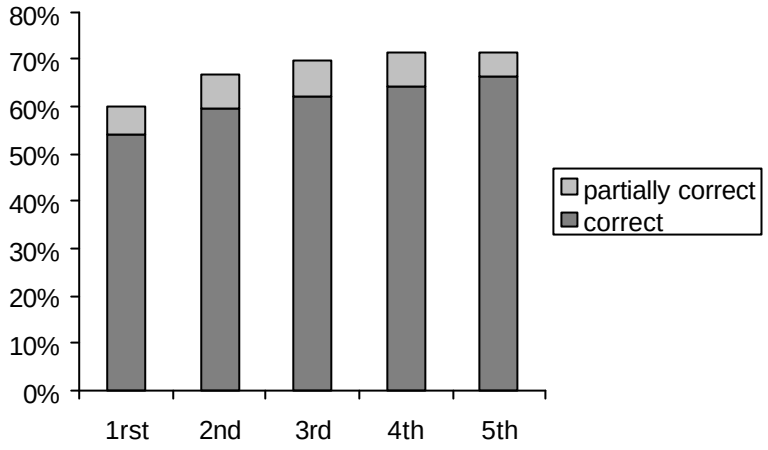

Figure 2. Scores for subsequently shown weather reports

Starting with the first hypothesis, Figure 2 suggests that the percentage of correct scores and the combined percentage of correct and partially correct scores improve across sessions. However, analyses of variance for both the percentage correct and the combined percentage showed that the p-values did not approach significance.

The avatar's mouthing

In order to test the second hypothesis, each sign was allocated to one of the following categories:

- **mainly dependent on mouthing**
  i.e. if the spoken component is left out, the sign either has no meaning, or has a completely different meaning, or can very easily be confused with signs for entirely different concepts.
- **partially dependent on mouthing**
  i.e. if the spoken component is left out, the sign looses part of its meaning. The information becomes less specific, but a substantial part remains intact, in other words, no wrong information is given, only less (e.g. difference between 'rain' and 'heavy rain').
- **independent of mouthing**
  i.e. if the spoken component is left out, the sign would unambiguously still have exactly the same meaning (includes signs without spoken component).

The percentage of correct and partially correct in each category is as shown in Figure 3. It indicates that for both the correct and the partially correct score the comprehension of signs that were mainly dependent on mouthing was considerably less than for signs that were only partially dependent or independent of mouthing. This was confirmed by analyses of variance tests which showed significant effects for percentage correct ($F=11.94$; $df=2,109$; $p<.0001$) and percentage of partially correct

(F=7.99; df=2,109; p<.001). The implication of this is that comprehension scores can be improved significantly by improving the mouthing components of the avatar.

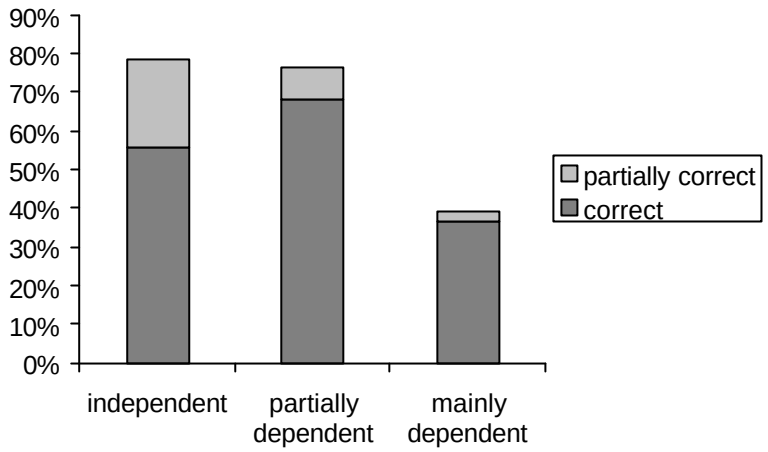

Figure 3. Scores for signs depending in various degrees on mouthing

The avatar's mimics

To test the third hypothesis, each sign was again allocated to one of the three categories:

- **mainly dependent on mimicry**
  i.e. if the mimics are left out, the sign either has no meaning, or has a completely different meaning, or can very easily be confused with signs for entirely different concepts.
- **partially dependent on mimicry**
  i.e. if the mimics are left out, the sign looses part of its meaning. The information becomes less specific, but a substantial part remains intact, in other words, no wrong information is given, only less (e.g. difference between 'rain' and 'heavy rain').
- **independent of mimicry**
  i.e. if the mimics are left out, the sign would unambiguously still have exactly the same meaning (includes signs without specific mimics).

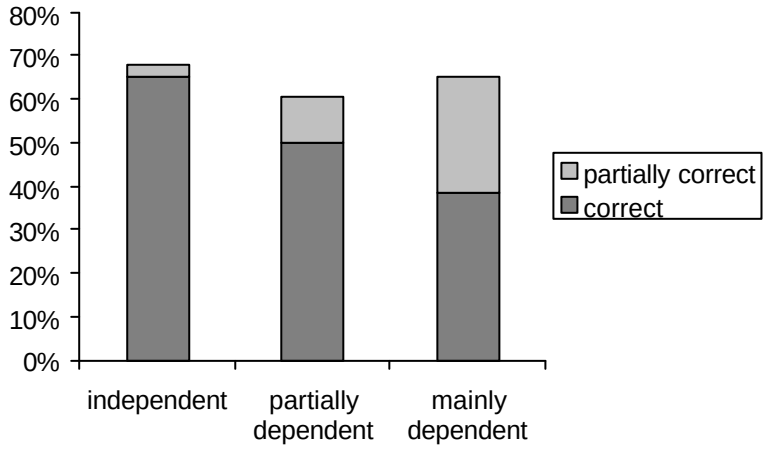

Figure 4. Scores for signs depending in various degrees on mimics

Figure 4 shows the percentage correct and partially correct as a function of dependence on mimicry. These data show that as dependence on mimicry increased, the percentage correct decreased but the percentage partially correct increased. This was confirmed by analyses of variance which showed significant effects for both the percentage correct ($F=8.57$; $df=2,109$; $p<.0001$) and the percentage partially correct ($F=14.32$; $df=2,109$; $p<.0001$). Interpreting these values it seems plausible that the more the meaning of the sign depend on the mimics, the more the subjects gave only part of the meaning in their repetition towards the camera ($F(2,109)=14.319$; $p<.001$). Therefore, we conclude that in order to reduce errors or missing values in the perception, improvement of the mimics of the avatar is less important than improvement of the mouthing. However, if in a certain application nuances need to be conveyed, improvement of the avatar's mimics is probably a good investment.

3.3 Subjective ratings

After carrying out the experimental task each subject was asked to rate the quality of the signs the way they were shown by the avatar by putting a mark on a continuous scale from very bad to very good. These markings are pictured below. They show that 7 of the 9 subjects rated the quality from reasonable to good. Only one person rated the animation as very bad.

X----------X----------XX-------XX----------XXX----------I

very bad reasonable very good

3.4 Results of interviews

After the ratings each subject was interviewed by the experimenter on the following topics:
- Did you have to get used to the avatar?
- How did you experience the signing avatar?
- Were there signs that you did not know?
- Which aspects of the avatar could make it difficult to understand the signing?

The following is a summary of the main points derived from these interviews.

Six of the nine subjects indicated that they had to get used to the avatar and three of the subjects thought the avatar was 'cold' (not spontaneous) or 'irritating'. It was noted by the experimenter that these subjects were also the only participants who were unfamiliar with using the Internet. This suggests that acceptance of the avatar is greater for web-surfers and that this acceptance may increase as a person becomes more familiar with the Internet. Two subjects were very enthusiastic about the avatar and said it was a pleasure to look at.

Seven subjects reported that they did not know all the signs. This is a problem of Sign Language of the Netherlands which is currently being standardised but not yet universally used by all deaf people, many of whom still use their own dialect of Sign Language of the Netherlands.

The main aspects of the avatar that were said to need improvement were:
- the mouthing, noted by six subjects
- the mimicry, noted by three subjects
- the speed of signing, noted by four subjects as being to slow

– the fact that the avatar's hands drop each time between signs—as seen by two subjects—interfere with the natural flow of signing

4 Conclusions

Sign language animation through an avatar is a potentially promising method for making web sites more accessible and user-friendly for deaf people.

The experimental evaluation described in this paper showed that 7 out of 9 deaf subjects regarded the quality of the signing of the current ViSiCAST prototype as reasonable or good. On the other hand, both the comprehension scores and the results of the interviews clearly indicate that the animation can still be significantly improved. The main aspects that need further attention are the mouthing and to a lesser extent the mimicry and the speed of the signing.

5 Acknowledgements

We are grateful for funding from the EU, under the Framework V IST Programme (Grant IST-1999-10500). We are also grateful to colleagues, both here at IvD, and at partner institutions, for their part in the work described above.

References

King, C.M. & Quigley, S.P. (1985). Reading and deafness. San Diego: College Hill Press.
Paul, P.V. & Quigley, S.P. (1994). Language and deafness (2nd ed.). San Diego: Singular Publishing Group