AUTOMATIC PRODUCTION SYSTEM OF SIGN LANGUAGE CG ANIMATION FOR METEOROLOGICAL INFORMATION

Shuichi Umeda, Tsubasa Uchida, Makiko Azuma,

Taro Miyazaki, Naoto Kato, and Nobuyuki Hiruma

NHK, Japan

ABSTRACT

NHK has developed a system for computer generation of Japanese Sign Language (JSL) graphics for meteorological information. As JSL is a different language from Japanese, persons whose first language is JSL have been demanding more TV programmes with sign language in addition to the closed caption services.

The JSL CG system automatically generates animations from the telegrams distributed by the Japan Meteorological Agency so that the user can immediately see the latest meteorological information on the Internet. In addition, we are now working on development to adapt the system to the NHK Hybridcast, which is the integrated broadcast and broadband system in Japan.

A major issue for the CG animations to be accepted by persons who use JSL for their daily communication is that the automatically generated hand movements of the animated characters connecting the sign language words may seem unnatural. Therefore we have developed a new method for connecting and interpolating between sign language word motions.

INTRODUCTION

As the sole public broadcaster in Japan, NHK strives to rectify the so-called information gap that creates social barriers. In Japan, many people with inherent impaired hearing use Japanese Sign Language (JSL) as their first language. JSL has a distinctive grammatical structure that is linguistically different from spoken Japanese. Consequently, to assure accessibility to broadcast programmes for the deaf persons, the number of broadcast programmes with sign language should be increased in addition to increasing the number broadcasts with closed captions. However, in Japan, the ratio of sign language broadcasting time to all broadcasting time is still low (see Figure 1) (1). It is noteworthy that even as the ratio of closed captioned broadcasting time has been increasing over the last few years, the ratio of sign language broadcasting time has hardly increased.

However, increasing the ratio of sign language broadcasting time is difficult because it is necessary to ensure enough sign language casters who can handle programme production.

Technically speaking, as information and communication technology has advanced, multiple communication services have become available. Developing communication support technologies using sign language and making those technologies available to the public are thought to be linked to the social momentum demanding strengthening of information support for the hearing impaired.

While continuing broadcasts with sign language provided by human JSL casters, NHK is advancing research and development on sign language computer graphics (CG) at the Science & Technology Research Laboratories. If sign language CG can be available, for example, at times when human JSL casters are unavailable such as in the middle of the night, it will be possible to provide emergency broadcasts with sign language attached. Aiming to provide useful information correctly, NHK has set up an Internet site concerning meteorological information.

Figure 1 – Change of proportions to total broadcasting time over the last six years

Moreover, exploiting the characteristics of CG, which are different from those of a real life picture, would enable flexible practices such as utilization of sign language CG for guidance signs in public offices (2) and for sign language education (3).

The present report describes the target and current achievement of the sign language CG research by NHK. Several technologies are described, such as ones close to practical utilization (like the sign language CG automatic production system based on XML formalized information) as well as technologies at a stage of development. Linking these technologies will realize sign language CG for regular TV programmes.

ESSENTIAL FACTORS FOR ADVANCEMENT OF SIGN LANGUAGE CG

In Japan, although research on artificial sign language started as early as in 1992 (4), its applications have not been widely adopted. Considering this situation, the following three factors can be said to be essential in putting sign language CG to practical use.
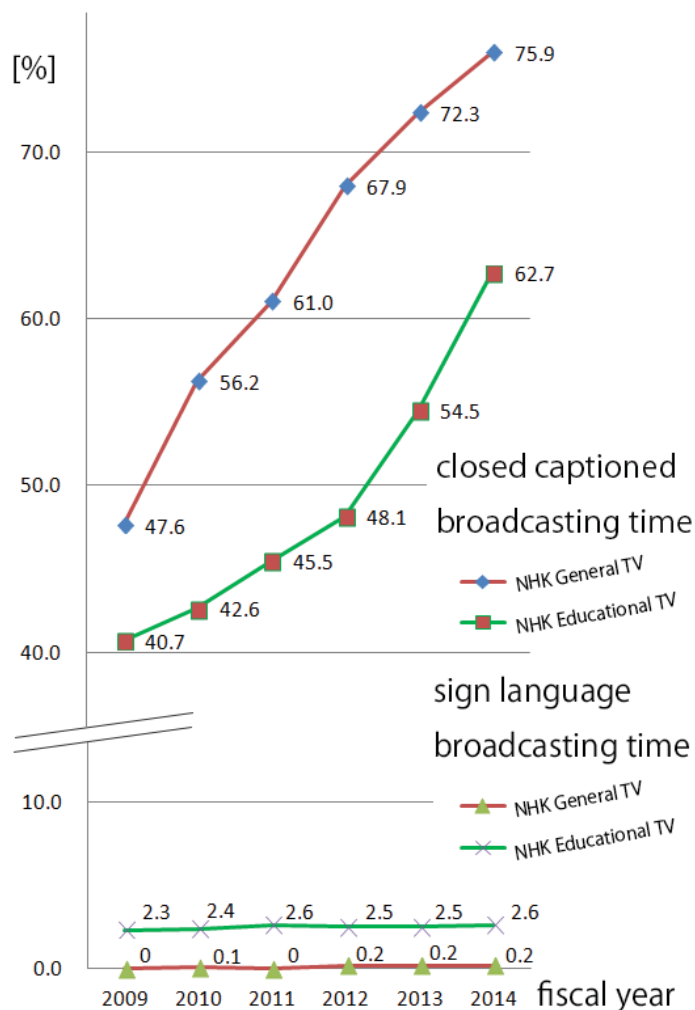
1: Presenting sign language CG to the public and gathering assessments

Numerous factors dictate the quality of sign language CG, including: a grammatical factor determining whether the sign language sentence is correct; a technological factor determining whether movements of generated animations are natural; and a subjective factor determining whether the external appearance of a CG character is to a person's liking. As for improving sign language CG, these quality determining factors need to be distinguished. To distinguish the subjective factor, the opinions of as many people as possible must be collected and evaluated.

Conventionally, to evaluate test CG images, subjects' opinions are gathered in a laboratory environment. However, it is not easy to invite enough deaf persons as subjects for evaluation, while experiments with insufficient subjects could result in inadequate accuracy.

Therefore, we have created a place for opening sign language CG to the public over the Internet for many people to see the CG and submit their assessments and opinions about them (refer to http://www.nhk.or.jp/signlanguage/).

2: Expanding the application areas of sign language CG

The range of sign language CG applications should be as diverse as possible. In addition to the terrestrial and satellite broadcasts, our eventual goal, we are working on the adaptation of sign language CG to NHK's "Hybridcast" which is a Japanese service (which corresponds to Europe's Hybrid Broadcast Broadband TV: HbbTV) that integrates TV broadcasting and communications. We are presently researching a service for delivering sign language CG by Hybridcast. The details of the service are described below in the section titled "Sign language CG system using Hybridcast".

3: Improving the quality of sign language CG

Sign language CG services that are acceptable to deaf persons can be achieved by improving the appearance and quality of movement of sign language CG characters. Since the appearance of a character is a subjective preference, production techniques are difficult to systemize. Therefore, we have asked various talented illustrators and animators to design more attractive characters for our system.

On the other hand, methods are introduced for improving the quality of movement of CG characters. In this report, they are described in detail in the section named "Method for controlling movement of sign language CG animation".

TECHNICAL FOUNDATIONS FOR AUTOMATIC SIGN LANGUAGE CG PRODUCTION

Template based methods

Although translating Japanese text into JSL on the basis of machine translation is being researched, general text is very difficult to translate from the standpoints of the accuracy of a translation. As for sign language CG that can be practically viable, a template based method is adopted for converting fixed phrases into strings of sign language animations. Although the text to be converted is limited, this method can quickly deliver error-free information.

3D model of characters for sign language CG

CG models for delivering sign language animation have a three-dimensional (3D) structure composed of a skeleton with surfaces depicted as polygons. The surface of the skeleton changes form in conjunction with the posture of the skeleton. The skeleton also has a face, and by defining the angle of rotation of the joints, utterances including facial expressions can be recreated by sign language.

This method is generally used for CG animated films, and if the skeleton structure is made consistently, various character models designed by various creators can be adopted for the sign language CG system. A specially created sign language CG model is designed to minimize processing costs, and when a sign language sentence with a speech duration of 80 seconds and containing 100 words is created on a single CPU, an utterances is started within 20 seconds and is completed in real time.

Sign language word motion data by optical motion capture

Movement data of sign language words are datasets that record the posture of the signer by means of optical motion capture. For optical motion capture, "markers" are attached to the joins (those of the fingers are shown in Figure 2) and images of those markers are captured by multiple cameras at 120 fps. Positional coordinates of multiple 2D images in 3D space are measured and recorded as rotational angles of joins. Since optical motion capture does not hinder the motion of the signer, it is suitable for recording sign language motions and therefore is thought to be superior to wearable recording devices such as the data glove as far as this application is concerned. In addition, by attaching markers to the face of the signer, the facial expressions can be recorded at the same time as the fingers motions. By using this method, we have already recorded more than 7000 sign language words derived from words frequent spoken during news programmes, etc.

Figure 2 – Markers attached to the fingers for an optical motion capture

When we have a CG character utter sign language, the facial expression is controlled by the joint angles, the same as in the case of fingers. However, the facial expression including the wrinkle and mouth form is difficult to portray even if we use 37 joints (Figure 3). In JSL, the facial expression is not only an emotional element but also a linguistic one, and portraying it faithfully is a future task.
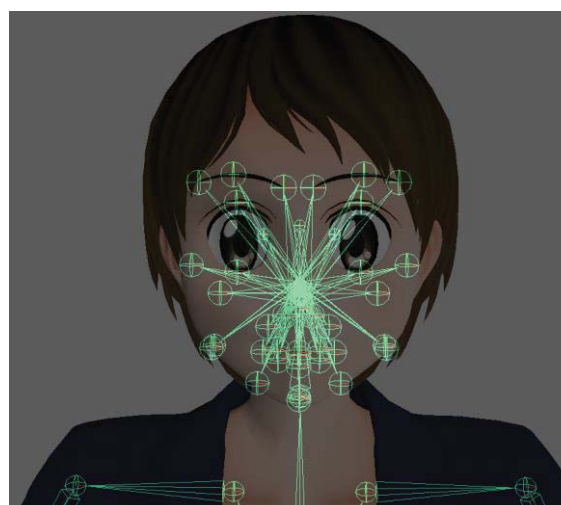

Figure 3 – Joints that control the facial expression

AUTOMATIC PRODUCTION SYSTEM FOR CREATING SIGN LANGUAGE CG ANIMATIONS FOR METEOROLOGICAL INFORMATION

For the abovementioned automatic production system for creating sign language CG animations for meteorological information, we have created an Internet site for making these motions of sign language words available to the public to ask for opinions from a wide audience.

In addition, we developed a production system which completes prefixed templates of sign language sentences with words or paragraphs created by motion capture and gets ready to show an automatically generated video clip. This system utilizes the template based generation method mentioned above. The system can analyze coded weather forecasts distributed by the Japan Meteorological Agency in XML format and convey them through animated characters using sign language [Figure 4]. It automatically generates and updates the sign language animation on the basis of the latest weather forecast as soon as it receives new information through the Internet. The system receives the code and inserts the relevant information into fixed phrases, such as "This is the weather information for [Tokyo]. Today's weather is [cloudy]." With the help of deaf persons and sign language interpreters, we are improving sign language expressions to be used as the templates.
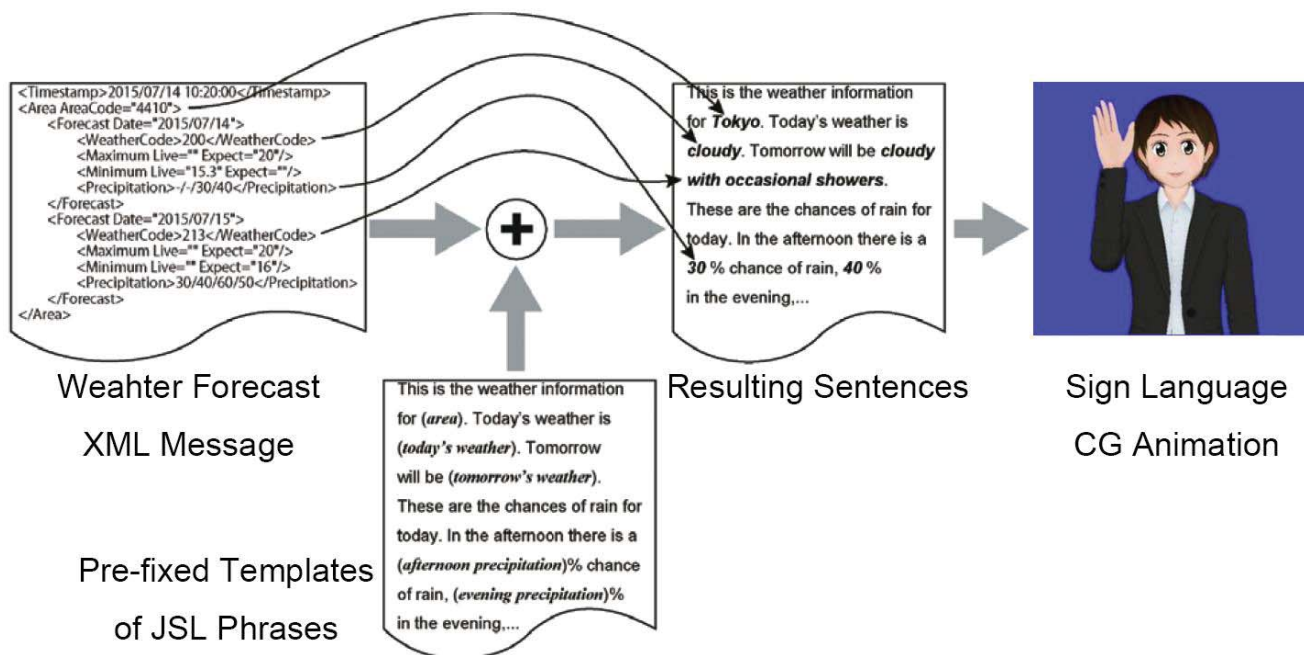

Figure 4 – Fixed phrases generator for weather forecast

The production system generates the weather forecasts for all 47 prefectural capitals of Japan whenever it receives an XML message so that users can see the latest weather forecast for any of these 47 cities in the form of the sign language CG via the Internet.

We confirmed that the system was able to perform the CG generation work of the quantity of making all prefectural information while it is difficult for human signers to perform the same task. Therefore, this is an example of the advantage of automatic sign language content generation.

SIGN LANGUAGE CG SYSTEM USING HYBRIDCAST

As a service for integrating broadcasts and communications, Hybridcast was launched in September, 2013 (5). It can be applied flexibly for a variety of services in forms such as contents independent of TV programmes and contents linked to TV programmes as well as bi-directional contents passing between both viewers and broadcasters. Consequently, it is considered to be the next place for providing sign language CG services after the stand alone Internet service.

As for the present standard for Hybridcast, the "Picture in Picture" function for showing sign language CG animation in a region set on the TV screen of a receiver is not realized in commercially available TV sets. First, thus we implemented a prototype Hybridcast service that is operated on a second screen called a companion device, such as the tablet.

Figure 5 shows the configuration of the prototype broadcasting system for emergency alerts in sign language CG. The system works as follows:

i) The distribution server receives a datagram from the Japan Meteorological Agency. However, in a different way to that used by the previous system, the distribution server does not perform rendering; instead,
ii) It delivers joint-rotation-angle data of generated sign language CG characters to the second screen at the user side.
iii) CG characters are then rendered in accordance with the received joint-rotation-angles on the second screen device.
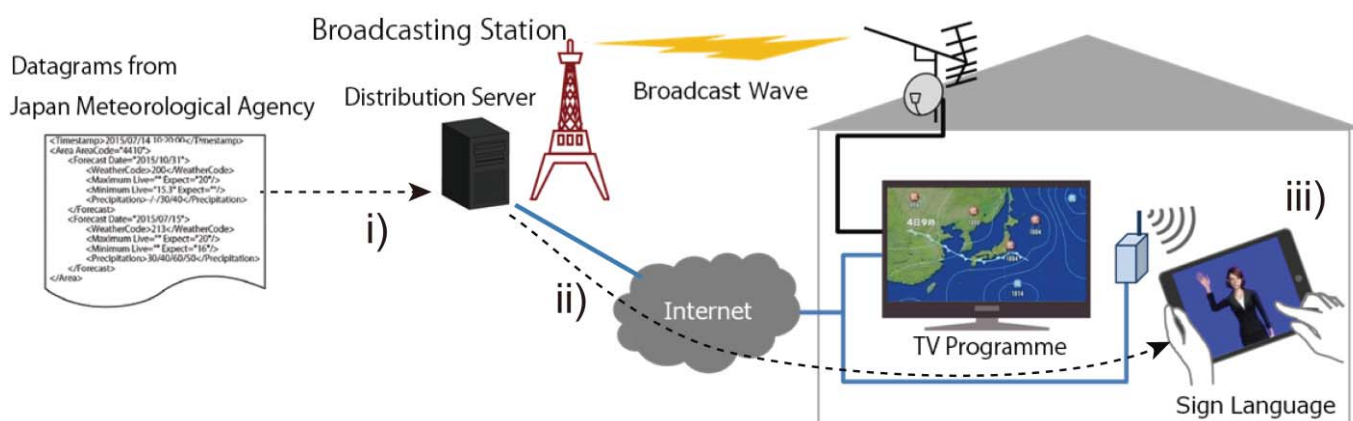

Figure 5 – Emergency alerts system with sign language CG using Hybridcast

In the experimental broadcasting environment, rendering is performed by WebGL function on the browser of the second screen, and the utterance of the sign language CG character is successfully executed. In this configuration, by completing downloads of large data such as CG character model's polygon and texture data at the moment that the power of the second screen device is switched on, emergency alerts can be presented with less delay. Moreover, TV receivers compatible with Hybridcast standard have a memory for storing information concerning, for example, the place of residence of users. By utilizing these data accordingly, users can possibly obtain guidance to evacuation areas near their place of residence in sign language CG.

METHOD FOR CONTROLLING MOVEMENT OF SIGN LANGUAGE CG ANIMATION

To generate a sequence of sign language, sentences from a multitude of sign language words or paragraphs recorded by motion capture are connected. The rotational angles of the joints of the connected sections are automatically interpolated. Unfortunately, however, the motions of these sections generated by our current prototype system are not satisfactorily human like and, a person viewing them might feel that they are unnatural. Furthermore, even the motion data accurately recorded by motion capture correction may be necessary because they can appear unnatural if seen as for reproduction of words in a context.

However, with a method for directly controlling the posture of a sign language CG character by means of rotational angles of joints, hand motions cannot be corrected as intended. To overcome this problem, we have developed a control method by taking advantage of a technique called "inverse kinematics: IK" [Figure 6].

A test program implementing the developed control method has the functions listed below.

i)   When a sequence of identification numbers of sign language words or paragraphs is input, motion data of a sign language sentence are temporarily synthesized with automatic interpolation between the words.
ii) Sign language CG characters and the trajectory of the signing hand are drawn on the generated animation screen.
iii) The shape of the drawn trajectory is freely changed, and the position of the hand is corrected manually.
iv) Rotational angles of the joints of the shoulders and elbows are calculated automatically from the position of the hands, and the image of the characters is rendered accordingly.
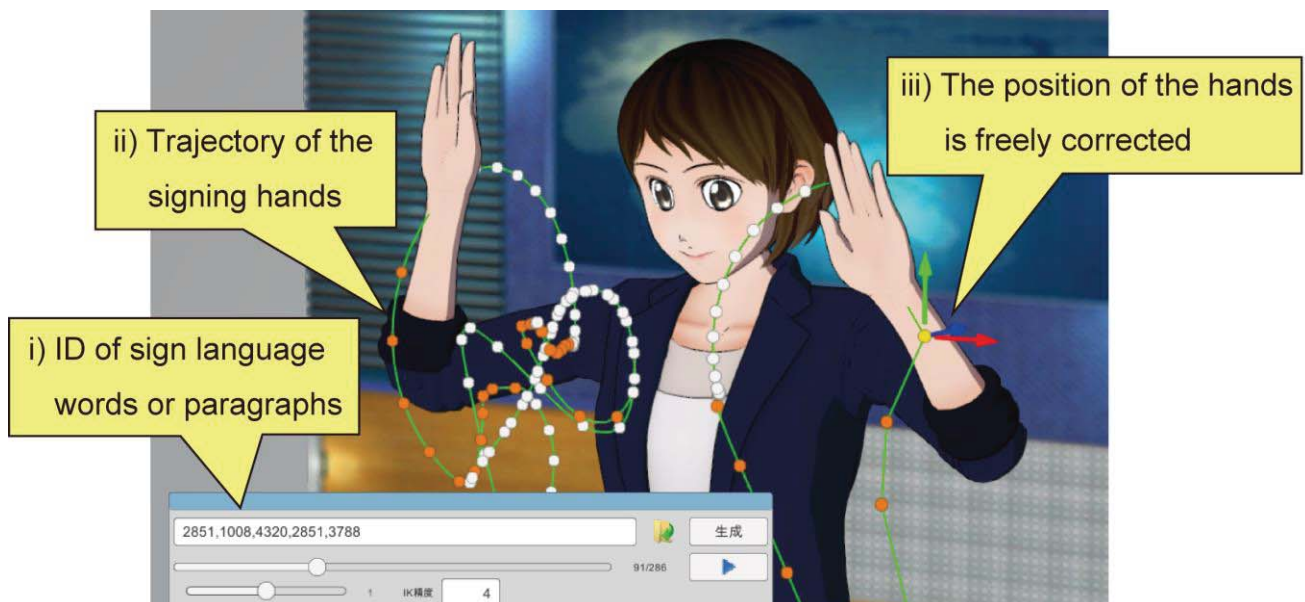

Figure 6 – Example of User Interfaces for Sign Language CG Editing

As a development platform, the game engine Unity was utilized. Although the rendering processing [function iv)] requires reiteration, the designed platform for real-time games enables rendering at more than 60 fps, so interactive modification is possible. By using this test program, the trajectory of the hand motion of the sign language CG character can theoretically be made exactly the same as that of a real person. However, it is not realistic to correct the trajectory of a hand manually every time a sign language sentence is synthesized. Therefore, the technology is needed for automatically performing correction [function iii)], and this is our next target.

Although our immediate goal is to make the motion of the CG character approximate that of a real person, our concern is that the sign language CG movement that mechanically imitates human movement exactly may make people feel uneasy. For example, motions of CG animated characters and sign language CG animation (6) (which were highly rated by a hearing impaired Japanese person) were not felt to be unnatural despite them being obviously different from motions of a real person. Making the motions of sign language CG just same as the recorded motions of a human signer is not necessarily appropriate, but it is important to match the movement to the appearance of the CG character.

CONCLUSIONS

The current status and future plans of sign language CG at NHK were described. The automatic production system for creating sign language CG animations for meteorological information can only deliver fixed sign language phrases prepared in advance. Even so, relying on a computer, it can automatically render sentence patterns that cannot be uttered in sign language by a person, thereby reducing operating costs. Moreover, by having amiable CG characters utter sign language, the popularity of sign language can be enhanced. From now onwards, while expanding the places in which Hybridcast can be applied, we intend to enhance the usefulness of sign language as well as improve the naturalness of characters' motions.

REFERENCES

1. Ministry of Internal Affairs and Communications, The results such as the closed-captioned broadcasting 2009-2014. MIC Press Document. April, 2011 to 2016.

2. Sarah Ebling and John Glauert, 2013. Exploiting the Full Potential of JASigning to Build an Avatar Signing Train Announcements. Proceedings of Sign Language Translation and Avatar Technology 2013 .

3. Georg Tschare, 2015. Learning sign language in a playful way with SiGame – App. Proceedings of Sign Language Translation and Avatar Technology 2015 .

4. H. Sagawa, H. Sakou, and M. Abe, 1992. Sign Language Translation System using Continuous DP Matching, Transactions of Information Processing Society of Japan, September, 1992. 69. pp.87 to 94.

5. H. Fujisawa, A. Baba, K. Matsumura, S. Mitsuya, M. Takechi, Y. Kanatsugu, H. Hamada, and H. Katoh, 2011. Hybridcast; Advanced Hybrid Broadcast and Broadband System, IBC Conference 2011.

6. World Federation of the Deaf, 2007. Advertisement for the World Federation of the Deaf in Spain , https://www.youtube.com/watch?v=wW2KBXrPEdM. July, 2007.