1. Introduction

Japanese Sign Language (JSL) is the first language for Japanese deaf and hard of hearing people and is easier for them to understand than Japanese. However, there is not enough information provided in JSL for users because the number of JSL interpreters is too small to translate much information from Japanese.

Recently, avatar-based sign language systems have been developed to increase the amount of sign language information. In these systems, signing avatars are made by using computer animation techniques, such as motion capture. Several systems are proposed for creating sign language animation for avatars that can be practically used within specific domains. In the domain of transactions made at a counter, Stephen et al. [1] proposed an English-British sign language system to aid in transactions between a deaf person and a clerk in a post office. Ruben et al. [2] also proposed a Spanish Sign Language system to help deaf people when they want to renew their driver's license. In the weather forecast domain, Margriet et al. [3] proposed multi-sign language systems that included Dutch Sign Language, German Sign Language and British Sign Language, while Azuma et al. [4] proposed a JSL synthesis system based on weather forecast data distributed by the Japan Meteorological Agency. As for the effectiveness of JSL CG animation in sports content, it has been stated that conveying what is happening in sign language or subtitles is important because many sports use sounds as signals [5]. However, there is no avatar-based sign language system for the real-time sports domain. In sports games, information presented without audio tends to be insufficient for deaf and hard of hearing people because games progress quickly and there are many complex rules. If sign language information could be automatically generated by using an avatar-based system, the number of sports programs presented with sign language in real time will increase. According to the results obtained from a preliminary interview of deaf and hard of hearing people, for content such as general news programs, all audio information needs to be interpreted sequentially in JSL in order to be sufficient for these viewers. The results showed that there are times that these viewers can understand the situation without audio information for sports programs. Therefore, what is also important is choosing effective information to be displayed in JSL at an effective timing and avoiding displaying unnecessary information that interferes with viewing, specifically for sports.
We propose JSL display systems for this domain. In the systems, JSL CG animation is generated by using avatar-based sign language synthesis. For application to the sports domain, we have to overcome two major problems as follows.

1. Collocation synthesis problem
Sign words are usually expressed one by one in the signer's front space, but they are done from right to left in their space when gathered into a collocation. In our application, there are many collocations peculiar to sports, for example, "0 (points) to 3 (points)" or "Japan versus USA". These collocations are difficult to transfer to CG animation because they cannot be made just by connecting the motion data of each word.

2. Problems with displaying two types of video
In the sports domain, two types of video have to be displayed, the JSL CG animation and the live sports. Displaying the two causes deaf and hard of hearing people to become fatigued because they are forced to move their eyes frequently between them.

We solved these two problems as follows.

To solve the first problem, we replace motion data for the body parts used for each word in collocations peculiar to sports with motion data borrowed from other words. For the second problem, we developed three prototype display systems to display the two types of video simultaneously: (a) a web browser-based system, (b) a tablet application based system and (c) a tablet & TV system.

The rest of this paper is organized as follows. In Section 2, we give an overview of a system we developed for automatically generating JSL CG. In Section 3, we present our three proposed display systems for solving the two problems in the sports domain. In Sections 4 and 5, we share the experimental results for our systems and discuss the results. Then, we conclude our paper in Section 6 by describing future work.

2. Automatic JSL CG Generation System

2.1 System Overview
The system for automatically generating JSL CG consists of a data analysis module and a JSL CG synthesis module, as shown in Figure 1.

The data analysis module receives game metadata in the XML format distributed during sports events and extracts the parameters needed for the JSL CG synthesis module from the data. The parameters include various pieces of information about the event, such as the competing countries, current score, highest scorer, type of penalty and player information. The data analysis module creates the JSL sentences by filling in the blanks of JSL templates with the extracted parameters. The templates are filled out during a game by corresponding motion data is referenced. The motion data are explained in the next section.

2.2 Sign Language Animation Synthesis
Sign language animation is produced by connecting the corresponding motion data of the JSL sentences, and then rendering the data in Unity, a cross-platform game engine, to automate the motion blending and video-file output process.

The motion data consist of body-joint rotation and...
facial displacement amounts recorded by using an optical motion capture system and stored in the Biovision Hierarchy (BVH) format. The capture system recorded the movement of a sign language expert as joint data of 112 points at 60 fps by using 42 cameras installed around a signer. These data included not only hand and body movements but also movements of the face that represent grammatically important mouth shapes and facial expressions. The original sign language motion was performed by a JSL expert, and a 3D skeletal model was then created in accordance with the body and face shapes of that expert. The optical motion capture and the bone structure of the CG character for JSL are shown in Figure 2. At the time of writing, we captured motions for more than 7000 words and phrases, and we continue to expand the database.

3. Proposed Display Systems

3.1 Collocation Synthesis

Words of sign language are usually expressed one by one in the signer’s front space. For example, "0 (points) to 3 (points)", which is a collocation peculiar to sports, is shown in Figure 3. However, the CG animation is not natural regarding sports collocations. With a native human signer, this collocation is expressed by representing the "0" sign on the right hand and "3" sign on the left hand at the same time.

We improved collocation synthesis as follows.

Since motion is converted into data for each joint, it is easy to extract only specified joint data for a certain JSL motion. Therefore, it becomes possible to handle the data for each body part independently, such as the face, right arm, right wrist and each finger. Consequently, it is possible to generate new sign language motion data by decomposing pre-recorded motion data and reconstructing it. Figure 4 shows an example of the result of manually converting motion data and representing it as a JSL CG animation (0 - 3).

In the case of the expression "0 - 3" shown in Figure 3, to represent the "0" motion on the right side of the face, we used the data for the "a" motion to position the wrist joint to the right side of the face. By replacing the wrist joint data from the "a" motion with that of the "0" motion, the new "0" motion is expressed naturally on the right side of the face. Similarly, the "3" motion, which is normally represented with the right hand, is reversed...
horizontally so that it is represented with the left hand. Finally, by combining the expressions to represent the "0" motion on the right hand and "3" motion on the left hand at the same time, "0 - 3" can be reproduced.

Although collocations do not occur only in sports, they are frequently required in sports in particular, such as to express scores or competing countries. Our previous version of the system for automatically generating JSL CG had no function for representing collocation rules; therefore, we implemented and verified only the representation of scores, which can easily be implemented. In sign language, there are linguistic characteristics expressing a time series and the relationship between a subject and objects in front of, behind and to the left and right of the signer. However, all collocations expressing these characteristics are unsupported in the present system. We have considered functions for collocations in sports as an intermediate step to creating a generalized JSL CG animation system in the future.

On the collocations used for expressing scores, when synthesizing a number with multiple digits, such as two or three digits, the time required for JSL movements is extended compared with one-digit numbers. Therefore, we prepared several patterns of base motions that included different combinations of digits of numbers, for example, a combination of motions for two digits on the right hand and one digit on the left hand or motions for two digits on both hands. Finally, a score expressed with a number with multiple digits can be synthesized by replacing the wrist joint data from prepared base motions with that for the motions of other numbers. As for the change in the order of expressions according to the subject, generally, the expression of a topicalized subject, such as a team's points, tends to be represented first. In addition to this, whether the left hand or right hand is used first also depends on the dominant hand of the signer. We decided to unify expressions made with the left hand first by referencing the JSL expert of our original sign language motion data.

3.2 Three Types of Display Systems

We developed novel display systems for providing information in consideration of opinion survey results on a prototype system [6]. We designed and implemented three types of display systems, shown in Figure 5: a web browser (a), tablet app (b) and tablet & TV (c). We intend for the systems to be used in conjunction with main game video provided by broadcasting or live streaming services. The systems display JSL CG video clips generated in real-time through a web browser or tablet application, and the video clips are synchronized with the events in the game video. Subtitles corresponding to these clips and score information are also updated automatically.

In addition, a function was implemented that allows viewers to watch videos that explain rules, which was highly rated in a preliminary survey.

4. Evaluation

We applied the automatic JSL CG generation system to actual winter sports games. To confirm the difference in the performance of the systems with regard to sports, we chose two different types of sports, ice hockey and curling, as target sports. Ice hockey is a fast-paced and physical sport, and penalties occur at random points in the game. However, when a game is interrupted due to a penalty, deaf and hard of hearing people can have difficulty understanding the situation. Curling is a slow-paced and non-physical sport, and the game progresses in a fixed fashion such that events occur regularly. The purpose was to verify whether users understand the information provided by the JSL CG animation, to ascertain the most desirable display method, to acquire opinions from deaf users and to verify the performance of real-time content generation. The generated JSL CG animation indicated the current score, the elapsed time, types of penalties (for ice hockey), types of shots (for curling) and so on. Other than the JSL video clips, the generated content included subtitles in Japanese, visualized audio content, tablet vibration or a whistle icon when a penalty occurred and score information.

The purpose of the automatic JSL CG generation system is supporting not only native speakers of sign language but also non-native people with hearing loss,
so we implemented a function for creating Japanese subtitles as well as JSL CG animations. Japanese subtitles are created by using the same method for generating JSL CG animation, that is, filling in the blanks of Japanese subtitle templates with extracted parameters. From the results of the previous opinion survey, which involved four native sign language speakers and one sign language interpreter who had been born to deaf parents, most participants suggested that it is desirable not only to display sign language but also that Japanese subtitles complement the displayed sign language to provide information. All of the participants in the previous survey could read Japanese without problem.

4.1 Participants

The participants were three deaf signers (one male and two females) ranging from 20 to 30 years in age. Two participants had used JSL as their mother tongue since birth, and the other had used it since becoming a teenager. All of the participants were proficient in reading and writing Japanese, and they were also able to understand questionnaires written in Japanese and answer by writing in Japanese. The two female participants have been enrolled in schools other than schools for the deaf, and they usually use multiple methods of communication such as communicating in writing in addition to JSL. The one male participant has not received inclusive education, and his ordinary method of communication is only JSL. They were not athletes or experts on ice hockey and curling, and all of them answered "Neither" on a question that checked whether they watched sports programs often or not. Furthermore, all participants were familiar with how to use a smartphone and/or tablet.

4.2 Procedure

First, the participants were asked to report their understandability of the JSL CG animation. Some studies have proposed methods for measuring the understandability of CG animation. We adopted a simple method in which the participants viewed short JSL video clips and then answered multiple-choice questions. There were three types of clips: sign language used by a human signer, manually produced JSL CG animation and automatically produced JSL CG animation. There were 30 questions, 5 about the clips of the human signer, 5 about the clips of a CG avatar that manually reproduced the motions of the human signer and 20 about the automatically generated CG animation clips. The manually produced animation clips were generated by arranging motion capture data in word orders corresponding to human signers and combining them by linear interpolation. Each clip was played 2 times for each question, and the participants were required to answer within 10 seconds.

Then, the participants were asked to evaluate the usability of the three display systems. They tried using each system and then evaluated the need for JSL videos, subtitles and other text or image information by using a 9-point rating scale in 0.5 increments (1: not effective - 5: very effective). The displayed game video was a five-minute summary of an ice hockey game. We displayed the JSL CG animation and the subtitles at the time that the game metadata was received for the actual game.

5. Results and Discussion

5.1 Understandability of JSL Animation

The results obtained by checking whether the participants had correctly answered the multiple-choice questions showed that the accuracy rate of the three participants was 100% for both the CG avatar and human signer. As for the time needed to answer the questions, all participants were able to answer all of the questions within 10 seconds. In the understandability test, we did not display Japanese subtitles at all. The accuracy rate of 100% was obtained with only JSL CG animation or only video of the human signer without subtitles. This demonstrates that the automatically generated JSL CG animation was as understandable as an actual human signer for fixed-pattern JSL sentences. We need to assess the acceptance of signing avatars as in prior evaluation studies, e.g., [8, 9].

5.2 Display Method

The participants gave their opinions on the systems by answering a questionnaire. Figure 6 shows the ratings for the display systems. The answers show that the tablet app (b) received the highest evaluation. The reason for this may be that, for this system, little eye
movement was needed to switch between the game video and the JSL CG animation. Furthermore, all content could be viewed with one easy-to-use tablet in hand. However, the system was not evaluated highly in terms of delay time and eye-movement fatigue. To alleviate fatigue, we will need to improve the user interface by finding a layout that minimizes eye movement. We will also consider implementing customizable functions and other notification functions that have been reported to be useful [10, 11].

5.3 Information Effectiveness

Figure 7 shows the results obtained for the evaluation experiment. "Game situation" and "Rule commentary" are game information provided with JSL CG animation. We found that the effectiveness of the information for each display system varied. Information such as the rule commentary in JSL, subtitles, tablet vibration and score information were highly evaluated for all systems. To evaluate the display systems in total including JSL CG animations and subtitles, we would only be able to verify the effectiveness when all information was displayed simultaneously. However, we evaluated the need for JSL CG animations independently from subtitles by using questions on a questionnaire sheet used for evaluation. As shown in Figure 7, although there were some differences depending on the display systems, the rating of the JSL CG animation for "Game situation" and "Rule commentary" was as high as that for subtitles. For a more precise evaluation, in the future, we would like to separately evaluate the case of displaying only JSL CG animation and that of only subtitles and conduct an evaluation to analyze which content is being looked at by using an eye tracking system.

As for the evaluation of each sport, the results showed that neither ice hockey nor curling needed typical information to be displayed, such as the current score. However, for both sports, the rule commentary with JSL CG animation and subtitles that the users could view at their own convenience was effective. Regarding the characteristics of each sport, the penalty information in ice hockey was very effective, but a lot of typical information in curling was considered to be insufficient when only the obvious information, such as the elapsed time and the player’s name, was provided. In the future, we will look into applying this system to other sports; for example, basketball and soccer share similar characteristics to ice hockey, and Boccia, played in the Paralympic Games, shares similar characteristics to curling. As it is considered that each sport has different effective information, we need to verify the system performance by applying the systems to other sports.

5.4 Delay Time

The delay time when displaying information could become a negative factor in viewing sports programs. Figure 8 shows the average delay time of the system processing. The average delay from the occurrence of an event to the display of JSL CG animation was 25 seconds for ice hockey and 31 seconds for curling. The average delay times are broken down in Figure 8 as follows. The values for "Data Analysis", which represents the time taken to create JSL sentences from game metadata and JSL templates, was 3.0 seconds for both ice hockey and curling. The values for "Animation Generation", which represents the time taken to generate the video clips of the JSL CG animation, was 11 seconds for ice hockey and 15 seconds for curling. The values for "Game Metadata Creation/Video Display", which represents the time taken to create the game metadata and display the JSL CG animation except for the two items mentioned above, was 11 seconds for ice hockey and 13 seconds for curling. The cause of the delay was primarily the time needed to generate the video clips of the animation and the time it takes to receive the event metadata. There was no specific numerical target for the allowable delay time that would be considered to be real-time, so we carried out an
evaluation by adding a delay time to the results of the experiment done using actual winter sports. The results obtained from the three participants regarding the allowable delay time showed that the average rate on a 9-point rating scale in 0.5 increments (1: unacceptable - 5: not mind at all) was 2.67 points. Further studies are needed in order to clarify a target value for the allowable delay time by comparing the evaluation of each type of content displayed with different delay times added. Except for the delay, it was confirmed that JSL CG animation can be automatically generated as a game progresses.

6. Conclusion

We developed a system for automatically generating JSL CG to assist deaf and hard of hearing people in viewing sports programs. A real-time experiment with actual sports games showed that information on games could be automatically generated as the game progresses. The results of an evaluation experiment indicated that the generated JSL CG animation was useful for understanding the information and that the preferable method for viewing displayed game video and information is to view them together on the same tablet device. We are continuously gathering opinions on the display systems from deaf and hard of hearing people and reflect those opinions by adding necessary functions and deleting unnecessary ones. In the future, we will continue to improve their accuracy and usability through evaluation experiments with more participants.

Acknowledgements

The authors would like to express their gratitude to the deaf and hard of hearing people and sign language interpreters who cooperated in our research.

References

1) S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt and S. Abbott: “The Development and Evaluation of a Speech to Sign Translation System to Assist Transactions”, International Journal of Human-Computer Interaction, pp.141-161(2003)
2) R. San-Segundo, J.M. Montero, R. Cordoba, V. Sama, F. Fernandez and L.F. D’Haro: “Design, Development and Field Evaluation of a Spanish into Sign Language Translation System”, Pattern Analysis and Applications, pp.203-224(2012)
3) M. Verlinden, C. Tijsseling and F. Prowein: “Sign Language on the WWW”, In Proceedings of the 18th International Symposium on Human Factors in Telecommunications (HFT2001), pp.197-204(2001)
4) M. Azuma, N. Hiruma, H. Sumiyosi, T. Uchida, T. Miyazaki, S. Umeda, N. Kato and Y. Yamanouchi: “Development and Evaluation of System for Automatically Generating Sign-Language CG Animation Using Meteorological Information”, In Proceedings of the 16th International Conference on Computers Helping People with Special Needs (ICCHP 2018), pp.233-238(2018)
5) A. Othman, O. El Ghoul and M. Jemni: “SportSign: A Service to Make Sports News Accessible to Deaf Persons in Sign Languages”, Lecture Notes in Computer Science 6180, pp.169-176(2010)
6) T. Uchida, T. Miyazaki, M. Azuma, S. Umeda, N. Kato, H. Sumiyosi, Y. Yamanouchi and N. Hiruma: “Sign Language Support System for Viewing Sports Programs”, In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS ’17), pp.339-340(2017)
7) M. Kipp, Q. Nguyen, A. Heloir and S. Matthes: “Assessing the Deaf User Perspective on Sign Language Avatars”, In Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS ’11), pp.107-114(2011)
8) H. Kacorri, M. Huenerfauth, S. Ebling, K. Patel and M. Willard: “Demographic and Experiential Factors Influencing Acceptance of Sign Language Animation by Deaf Users”, In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS ’15), pp.147-157(2015)
9) M. Kipp, A. Heloir and Q. Nguyen: “Sign Language Avatars: Animation and Comprehensibility”, In Proceedings of the International Conference on Intelligent Virtual Agents (IVA2011), pp.113-126(2011)
10) M. Seita: “Closed ASL Interpreting for Online Videos”, In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS ’16), pp.337-338(2016)
11) A.C. Cavender, J.P. Bigham and R.E. Ladner: “ClassInFocus: Enabling Improved Visual Attention Strategies for Deaf and Hard of Hearing Students”, In Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS ’09), pp.67-74(2009)

Tsubasa Uchida received his B.E. and M.E. degrees in Media and Telecommunications Engineering from Ibaraki University, Japan, in 2008 and 2011 respectively. He joined NHK in 2011 and since 2014 has been with NHK Science and Technology Research Laboratories. His current research interests include a sign language animation system for hearing-impaired persons.

Hideki Sumiyoshi graduated in 1980 from the Hiroshima Prefectural Technical High School and received a Ph.D. degree in Engineering from the University of Tokyo, Tokyo, in 2004. He joined NHK (Japan Broadcasting Corporation) in 1985. Since 2018, he has worked for NHK Engineering System. His research interests involve TV program production systems and Web systems.

Taro Miyazaki received his B. E. and M. E. degrees in Information Science from the Tokyo Institute of Technology in 2004 and 2006, respectively. After graduation, he joined NHK (Japan Broadcasting Corporation) and has been working at the Science and Technology Research Laboratories since 2011. From 2017 to 2018, he was a visiting scholar at the University of Melbourne. His research interests include machine learning, data mining, and information retrieval.

Makiko Azuma received her B.E. and M.E. degrees in Precision Engineering from the University of Tokyo, Japan, in 2008 and 2010 respectively. After graduation, she joined NHK, and worked at NHK Sendai from 2010 to 2013. Since 2013, she has been working at NHK Science & Technology Research Laboratories, and she is currently focused on automatic sign language animation generation and haptic interfaces.
Shuichi Umeda received a M.E. degree from the Nara Institute of Science and Technology in 1997. After having been in charge of CG production at a broadcasting station, he joined NHK Science & Technology Research Laboratories in 2009. There, he researched the characteristics of the human visual system, and he has been developing a practical CG generator for Japanese Sign Language.

Naoto Kato Ph.D., joined NHK in 1988. He has been engaged in research on natural language processing, especially on machine translation at NHK Science and Technology Research Laboratories (STRL). From 1994 to 1997 and 2004 to 2006, he was a visiting researcher at Advanced Telecommunications Research Institute International Laboratories and is now a Senior Research Engineer at NHK STRL.

Nobuyuki Hiruma is a Senior Research Engineer at NHK Engineering System, Inc. He received a B.E. degree from the University of Electro-Communications in 1982 and an M.E. degree in 1984 and Ph.D. degree in 2000 from the University of Tokyo. He joined NHK in 1984 and moved to NHK Science and Technology Research Laboratories (STRL) in 1988. His major research fields include the psycho-physics of human vision, human factors of information displays and automatic generation technology for sign language CG animation.

Hiroyuki Kaneko received his B.E. and M.E. degrees in Electric and Electronics Engineering from Sophia University, Tokyo, Japan, in 2000 and 2002, respectively. He joined NHK in 2002. He has been engaged in research on computer graphics based content production technology, especially on sign language animation in NHK Science and Technology Research Laboratories since 2006.

Yuko Yamanouchi received a B.E. degree in Electronic Engineering from Sophia University in 1986, and the M.E. degree and Ph.D. in Electronic Engineering from the Tokyo Institute of Technology in 1988 and 2005, respectively. She joined NHK in 1988. Since 1990 she has been with NHK Science and Technology Research Laboratories, and she is currently the head of the Smart Production Research Division. She has been engaged in research on image processing and computer graphics techniques for TV program production.