Universal Service Utilizing Real-time Sports Data Feeds

KUMANO Tadashi UCHIDA Tsubasa KANEKO Hiroyuki

Abstract

We developed a technology to automatically generate play-by-play text using competition data broadcast from OBS and provide them with synthesized speech and subtitles, as well as sign language CG for the Olympic and Paralympic Games Tokyo 2020. The technology was implemented as a universal service to add to live streaming video. The live sporting updates were provided on a special Internet site in a format that was easy to understand and tailored to users' requirements, such as for people with visual or hearing impairments, people who cannot view a screen while driving a car, or people on a train who cannot hear the sound. This article describes the automated generation technology used for voice play-by-play with subtitles and sign language CG play-by-play.

Keywords : Visual impairment, Hearing impairment, Information accessibility, Sign language, Automated play-by-play, Speech generation, Speech synthesis, Sign language CG

1. Introduction

A widely used practice in program production is for announcers to add audio descriptions such as play-by-play and narration to video content. This is useful for conveying information that cannot be understood without viewing the video and enables the same information to be conveyed to all viewers irrespective of any impairments they might have (this is called "information accessibility"). The information is also helpful for many viewers in that it helps them better understand the content that is difficult to understand even by viewing the video. In the past, non-broadcasting services such as live-streaming services sometimes abandoned providing speech and subtitles due to human resource limitations, which was a problem from the perspective of allowing more people to enjoy the content.

In addition, even programs that include play-by-play and narration are mainly intended for viewers without visual impairment and are often inadequate for visually-impaired viewers. Broadcasting with live commentary is an audio-added service to compensate for this. However, in live broadcast programs, in particular, the burden on the person providing the commentary to judge the information to be commented on and speak at the right time while listening to the broadcast audio is high, and this service has not yet become a continuous service.

Although subtitles are available as a means of providing information to people with hearing impairments, for those who are deaf from birth or have lost their hearing at an early age, sign language is their native language, and understanding information provided in Japanese and viewing it for long periods can be a significant burden.

There is a need to develop technologies for solving such problems that can automatically generate speech,

KUMANO Tadashi
Science & Technology Research Laboratories, Japan Broadcasting Corporation
E-mail kumano.t-eq@nhk.or.jp
UCHIDA Tsubasa (IEICE Member)
Science & Technology Research Laboratories, Japan Broadcasting Corporation
E-mail uchida.t-fi@nhk.or.jp
KANEKO Hiroyuki (IEICE Member)
Science & Technology Research Laboratories, Japan Broadcasting Corporation
E-mail kaneko.h-dk@nhk.or.jp
The Journal of The Institute of Electronics, Information and Communication Engineers, Vol.105, No.8, Supplement, pp.82-88, August 2022

subtitles, and sign language CG (Computer Graphics) that provide appropriate explanations of the situation instead of people, especially for live broadcast content. In response to this, NHK has been conducting research and development on the automatic generation of play-by-play commentary. Sports content generally contains a large amount of video information, and play-by-play commentary is also indispensable for those who can view the video to understand the game's content. Therefore, we developed a technology to automatically generate play-by-play scripts from available real-time competition data synchronized with the game's progress and convey them vocally using a speech synthesizer or by sign language using CG animation technology.

This article describes the "robot play-by-play service with subtitles" and the "sign language CG play-by-play service" implemented by NHK as universal services at the Olympic and Paralympic Games Tokyo 2020 (hereinafter referred to as "Tokyo 2020 Games").

2. Real-time Competition Data

Data such as team or athlete names, scores, penalties, and each athlete's movements have been generated and distributed in real-time, especially in large-scale sporting events. In the case of the Olympic and Paralympic Games, the Olympic Broadcasting Services (OBS) distributes real-time data called ODF (Olympic Data Feed) [1] for all events. Based on this data, broadcasters produce various contents such as superimposed text and graphs on the broadcast screen, Web content, etc.

In ODF, "messages" are delivered each time there is a change in message content, which is organized into a single XML (eXtensible Markup Language) document for each type of information, such as "participating athletes" and "game status", including current score and stats, and a "list of events" that occurred during the game. Although the types of messages delivered and the specific information in the messages differ depending on the competition, specifications are defined so that they can be handled in a unified format.

3. Mechanism for Generating Play-by-play Scripts

Figure 1 shows the overview of our system that automatically generates voice play-by-play from competition data broadcast in real-time. Real-time data is in a predefined format, updated at the venue as the game progresses, and distributed with every update. The play-by-play generation system receives the data sequentially and always maintains the latest status of the

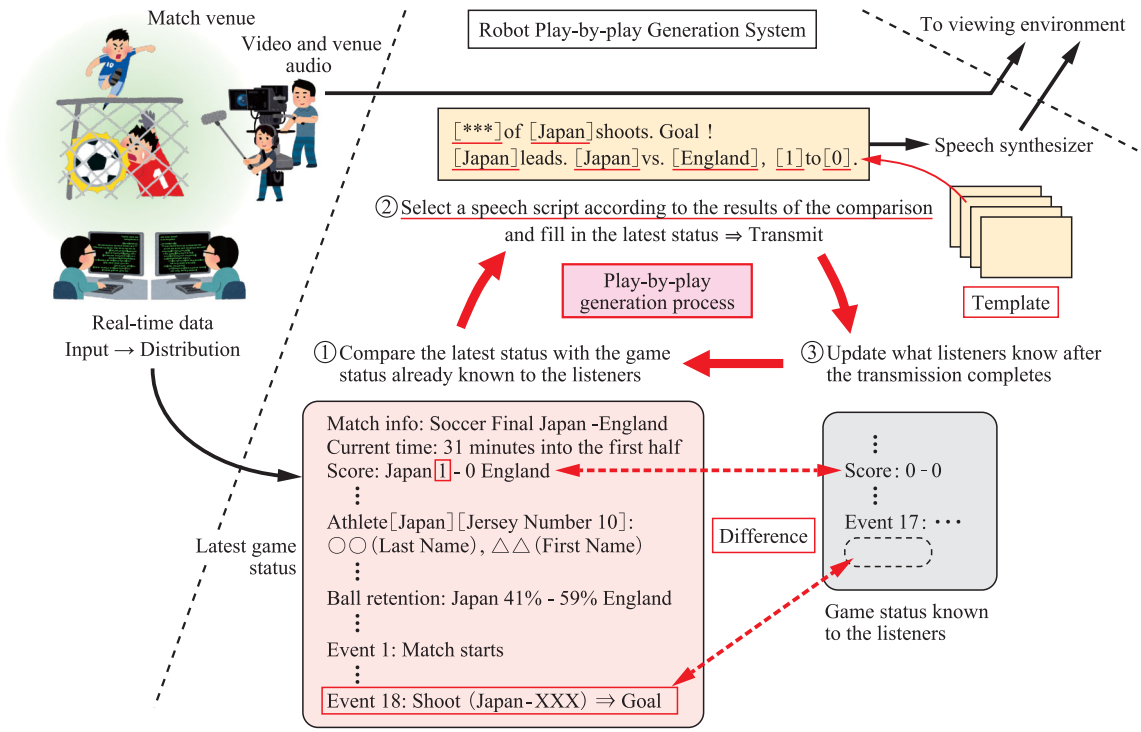

Figure 1 Mechanism of the Robot Play-by-play System

game. In addition, the system maintains the game's status, which the listeners who have heard the previous play-by-play comments should know to find the differences between the previous and current game status. Specifically, the following steps are repeated from the game's start to end[(2)].

① Compare the latest status with the game status already known to the listeners and check whether there is anything new that needs to be conveyed.
② If there is anything new to be conveyed as a result of step ①, select one from the collection of prepared incomplete play-by-play scripts according to the result of the comparison, and complete it by filling in the latest game status. If nothing is to be conveyed, return to step ①.
③ Convert the completed script into voices using a speech synthesizer and send it out together with the video and audio from the venue.
④ Reflect the conveyed information in the game status known to the listeners.
⑤ Return to step ①.

We call a collection of incomplete (some information needs to be filled) play-by-play scripts required in step ② a template. The templates must be prepared for each sort of target event, for each voice and sign language CG play-by-play. Each script in a template has some conditions to be selected according to the type of difference detected in step ①. They can also have a special condition, such as "when a certain time has elapsed since a certain information has been conveyed without any change", which enables the system to convey the same information periodically, such as a competitor's name or current score that has not changed for a long time. This is useful for those who cannot continuously check the information on the screen or who have started watching a game from the middle.

4. Production and Distribution for Each Service

This section describes the method of producing and distributing service contents for each robot play-by-play and sign language CG play-by-play from generated scripts by the method described in Section 3. Figure 2 shows an overview of the entire universal service system.

4.1 Robot Play-by-play with Subtitles

The generated play-by-play scripts were synthesized into voice speech for our audio play-by-play service named "robot play-by-play". The synthesized voice was superimposed on the video and audio content from the venue and provided via the Internet. In addition, the textual scripts to be read by the speech synthesizer are also delivered to enable viewing with subtitles on a Web browser, thereby improving service for those who cannot hear. Figure 3 shows an image of the delivered content of the robot play-by-play with subtitles.

We developed a dedicated speech synthesizer suited

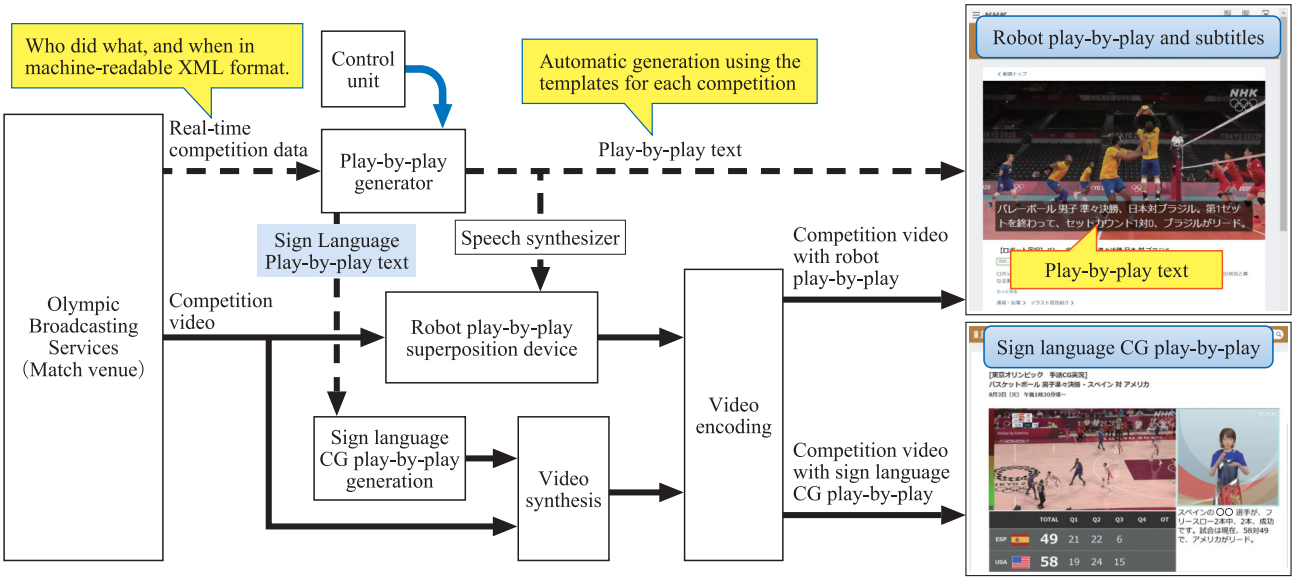

Figure 2 Overview of Universal Service System (©2021-International Olympic Committee-All Rights Reserved)

for sports play-by-play. This is achieved by preparing many voice samples that include expressions used in play-by-play by recording the human voice reading and having the DNN (Deep Neural Network) learn them. In addition, we achieved a more realistic and understandable speech play-by-play by implementing a mechanism for the speech synthesizer, which can recognize special symbols embedded in speech scripts indicating where to place emphasis and pause.

4.2 Sign Language CG Play-by-play

As in the case of robot play-by-play, the sign language CG play-by-play production system first generates textual sign language scripts by filling the latest status of a game into an incomplete script selected from a template, then converts them into sign language CG animations. The system configuration is shown in Figure 4[3].

Figure 3 Screenshot of the Robot Play-by-play Service with Subtitles (©2021-International Olympic Committee-All Rights Reserved)

Each sign language script is described as a sequence of sign language unit IDs. Every ID represents a different sign language expression that could be used in play-by-play, including words, common phrases, numerical expressions, competition terms such as "defensive fouls" and "traveling", athlete names and other proper nouns included in the competition data.

The sign language CG generator converts the ID sequences mentioned above into CG, representing a series of actions. It consists of a sign language motion database that stores the motion data of CG characters corresponding to each ID, CG model data such as CG characters and studio sets, and a rendering engine that generates sign language CG animations using these data. Optical motion capture was used to record the motion data for the sign language motion database. Markers were attached to the head and joints of limbs, etc., and their movements were captured by more than 40 infrared cameras, recording and storing the data. Figure 5 shows how the recording was done in the motion capture studio and hand with markers attached.

The generated sign language CG animation was laid out with the score and subtitle information generated together with the speech text from the competition data. It was output with the competition video in the upper left corner of the image blank.

The video synthesizer combined the competition video and the animation output from the sign language CG generator through a delay device to produce a single video stream. The final composite screen consisted of the competition video in the center, with the generated sign language CG animation on its right, the score display using the distribution data on the bottom, and subtitles generated from the distribution data on the lower right. On the left side of the competition video, a

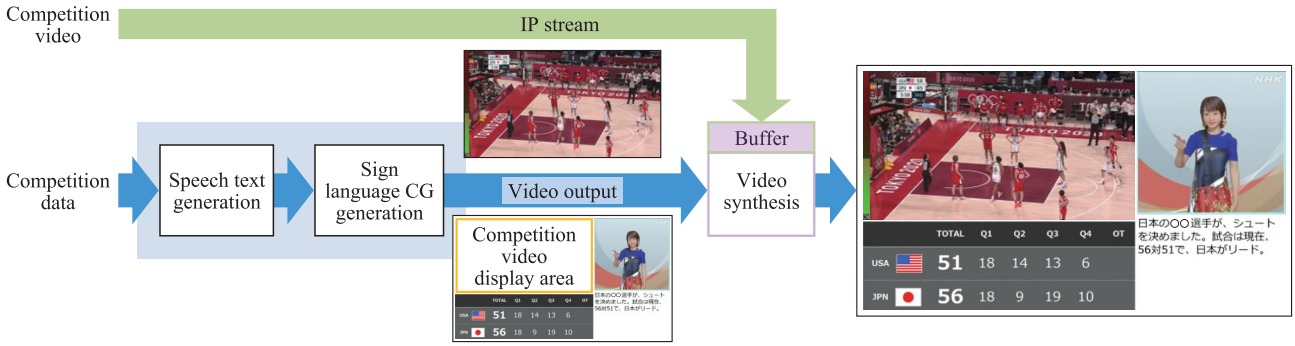

Figure 4 Structure of the Sign Language CG Play-by-play System (©2021-International Olympic Committee-All Rights Reserved)

graph visualizing the strength of the audio included in the competition video is displayed, enabling viewers to obtain comprehensive information, such as visually capturing the excitement of the spectators at the venue.

The composite video generated by this system was provided as though streaming a video on a website. An example of the website screen is shown in Figure 6. The fouls during the game are immediately displayed in the text at the bottom of the composite video. When the user clicks on the text, another male character appears on the right side and explains the foul in sign language. This feature was developed in response to many requests from users with hearing-impaired in a preliminary questionnaire survey, such as "I cannot understand what happens when a foul occurs just by looking at the video", and "I understand that an athlete stops playing and the

(a) Studio recording

(b) Attaching markers to the hand

Figure 5 Motion Capture Recording

(a) Animated character

(b) Natural (photorealistic) character

Figure 7 Comparison of Animated CG and Photorealistic CG Representing 3 to 0.

Figure 6 Example of the Sign Language CG Play-by-play Service Screen (©2021-International Olympic Committee-All Rights Reserved)

game is suspended, but I want to know the details of why they were suspended".

A new CG character was developed for the sports play-by-play, which is more natural (photorealistic), like a live-action character (Figure 7), instead of the previously developed weather sign language CG[4] developed by NHK for weather information, which was an animated character with large, caricature eyes. In sign language, not only are the shape and movement of the hands, but also the lifting and lowering of the eyebrows, the widening of the eyes, the shape of the mouth, and the expression and posture of the entire face are used as grammar to convey information. For the photorealistic character of this project, various initiatives were implemented to express natural facial expressions in response to local changes in the motion data.

5. Service Achievements

For the Tokyo 2020 Games, we provided the "Robot Play-by-play Service with Subtitles" and "Sign Language CG Play-by-play Service" via the Internet on NHK's special website.

The information in the real-time data differs significantly from competition to competition, making it challenging to provide meaningful play-by-play for competitions with little helpful information, and input delays significantly impact the quality of the actual situation as most of the data is input manually. The selection of competitions was made based on a comprehensive consideration of these factors.

At the Olympic Games, the robot play-by-play service with subtitles was provided in 10 competitions (weightlifting, canoeing, swimming, table tennis, tennis, beach volleyball, basketball, badminton, volleyball, and rowing), and 175 videos of 518 hours were distributed. At the Paralympics, 110 videos of 506 hours were distributed in seven competitions : swimming, wheelchair tennis, wheelchair basketball, wheelchair rugby, sitting volleyball, table tennis, and badminton. The service was especially popular for rowing and canoeing, and we believe it helped people watch live streaming of rarely broadcasted sports. The site was well-received on social media, with users commenting that it was "simple and easy to understand" and "follows the flow of the sport and explains the situation with fairly quick responses", and the system was used not just by people with disabilities but also by a wide range of people.

Sign language CG play-by-play was used for basketball at the Olympics, and 14 videos of 28 hours were distributed. It was implemented in two sports for the Paralympics, wheelchair basketball and wheelchair rugby, and 13 videos of 26 hours were distributed. This was NHK's first attempt to add sign language CG to live sports broadcasts, and the service received positive comments from a wide range of viewers, not just those with hearing impairments. The naturalness of the sign language expression using high-quality CG characters was highly evaluated on social networking services, with comments such as "The sign language is quite easy to understand", and "The fine facial expressions that are important for sign language have been studied, and the soft hand movements are amazing!". In general, sign language interpreters usually take turns every 15 minutes, but in extended programs such as sports, there is a significant burden on the interpreters. In addition, interpreters with specialized knowledge must be found for sign language in sports, which is infrequently expressed in daily life, making it challenging to provide sports play-by-play using sign language. However, this technology has made it possible to provide fully-automated sign language CG play-by-play.

6. Conclusion

This article describes the robot play-by-play with subtitles and sign language CG play-by-play for the Tokyo 2020 Games as an approach to universal service utilizing real-time data feeds. We developed a technology to automatically generate play-by-play scripts from sports data transmitted in real-time during competitions and convert them into synthesized speech and sign language CG animations. We also reported implementing a universal service that adds them to live streaming video.

This technology will automate the production of audio play-by-play, subtitles, and sign language for live content, which has been challenging to provide due to human resource issues and will enable the increase of universal content that more people can enjoy. Public broadcasting aims to provide all people with the necessary information quickly and accurately. We will continue to focus on research and development of universal services in cooperation with the parties concerned.

References

(1) International Olympic Committee, "Olympic data feed." http://odf.olympictech.org/

(2) T. Kumano, M. Ichiki, K. Kurihara, H. Kaneko, T. Komori, T. Shimizu, N. Seiyama, A. Imai, H. Sumiyoshi, and T. Takagi, "Generation of automated sports commentary from live sports data," Proc. IEEE Int. Symp. Broadband Multimedia Systems and Broadcasting, June 2019.

(3) T. Uchida, H. Sumiyoshi, H. Kaneko, and M. Sano, "Delay reduction and usability improvement of automatic sign language CG generation system for sports program," ITE Winter Annual Convention, 15B-2, Dec. 2019.

(4) M. Azuma, N. Hiruma, H. Sumiyoshi, T. Uchida, T. Miyazaki, S. Umeda, N. Kato, and Y. Yamanouchi, "Development and evaluation of a system for automatically generating sign-language CG animation using the meteorological information," Proc. 16th Int. Conf. Computers Helping People with Special Needs, July 2018.

(Received February 28, 2022 ; Revised March 18, 2022)

KUMANO Tadashi

He graduated from the Tokyo Institute of Technology in 1993 with a degree in Computer Science. He completed his Master's in Computer Science from the Graduate School of Science and Technology, Tokyo Institute of Technology, in 1995. He joined NHK in the same year. He has been engaged in research and development of natural language processing, machine translation, sports play-by-play generation, and speech synthesis at the Science and Technology Research Laboratories since the same year. From 2000 to 2004, he was seconded to ATR Spoken Language Communication Research Laboratories. Currently, he is a senior researcher in the Smart Production Research Department.

UCHIDA Tsubasa (IEICE Member)

He graduated from Ibaraki University in 2008 with a Bachelor's degree in Media and Telecommunication Engineering. He completed his Master's course in Media and Telecommunication Engineering at the Graduate School of Science and Technology, Ibaraki University, in 2011. He joined NHK in the same year. After working at Sendai Broadcasting Station, he has been engaged in the research and development of sign language CG at the Science and Technology Research Laboratories since 2014. Currently belongs to the Smart Production Research Department.

KANEKO Hiroyuki (IEICE Member)

He graduated from Sophia University in 2000 with a B.E. in Electrical and Electronic Engineering. He completed his M.E. in Electrical and Electronic Engineering at the Graduate School of Science and Technology, Sophia University, in 2002. He joined NHK in the same year. After working at Sendai Broadcasting Station, he has been engaged in research on content description, production, and sign language CG generation technology at the Science and Technology Research Laboratories since 2006.