5 Smart Production - Universal service

5.1 Automatic closed-captioning technology

Closed-captioning service not only serves the needs of people with hearing disabilities and the elderly by conveying speech in TV programs in text but also offers a useful function to general viewers watching programs in a noisy environment or an environment where audio cannot be played such as public places. Also, there is a demand from viewers for more programs with closed captions, including live broadcast programs and programs produced by regional broadcasting stations. Speech recognition technology is used for real-time closed captioning, but many regional broadcasting stations are faced with issues such as a shortage of staff to correct speech recognition errors and a considerable amount of time required to prepare necessary equipment and system. To address these issues and expand closed-captioning service, we began a trial service that distributes speech recognition results on the internet without correcting recognition errors with the aim of evaluating the extent to which uncorrected speech recognition results can help viewers understand programs.

Speech recognition technology for internet delivery

A service to deliver uncorrected recognition results on the internet requires a higher recognition accuracy. We therefore employed a speech recognition technology that we developed for the transcription of video footage, which requires a higher level of recognition than for program speech. This technology can generate highly precise recognition results sequentially for speech inputs streamed over network without waiting for the end of a sentence. We verified the recognition accuracy of this speech recognition technology when used for news and information programs produced by regional broadcasting stations and demonstrated that a recognition accuracy of 90% to 97% can be achieved for news items and live coverage and information items.

Meanwhile, the recognition accuracy declined significantly for interviews with local residents, which contain inarticulate speech, an accent peculiar to the region or different phrases from those used in standard Japanese. The recognition results of these parts make it difficult to understand the program due to incorrectly recognized words. Also, many interviews in news and information programs are open captioned and do not need additional closed captions. We therefore elaborated a way of presenting closed captions by stopping speech recognition and displaying “...” in closed captions for parts where the recognition accuracy is expected to decline.

The Chinese characters of personal names used in programs produced by regional broadcasting stations cannot be correctly identified in some cases because they are not always the same as those of names used in programs produced in Tokyo, which are used as training data for speech recognition. Therefore, we decided to display personal names in katakana, which is the Japanese phonetic characters.

NHK broadcasts regional programs of each regional broadcasting station in a broadcast time frame starting 6:10 PM on weekdays. Providing this trial service across the nation would require as many sets of speech recognition equipment as there are regional broadcasting stations, incurring a large-scale capital investment and operation costs for the installation and maintenance. We therefore built a system efficiently by aggregating speech recognition and delivery equipment on the cloud (Figure 5-1). We began this trial service in three NHK broadcast stations of Shizuoka, Kumamoto and Fukushima in February 2019.

[References]

5.2 Audio description technology

We are researching “audio description” technologies, which produce voice explanations for live broadcast programs so that people with visual impairment can enjoy live sports programs better. We studied “automated audio description,” which supplements human commentaries with auxiliary voice explanations for visually impaired people, “robot commentary,” which provides commentaries for internet services in place of human announcer, and a speech synthesis technology, which is the base of these technologies.

Automated audio description

As with manually produced audio descriptions, an automated audio description should not overlap with program speech to the extent possible. We studied a technique to estimate a desirable timing for inserting a commentary from the variation trend of acoustic features of speech. We developed a prototype system that operates in real time and exhibited it at the NHK STRL Open House 2018. The results of comparison between
the timings estimated by this technique and the timings which are visually confirmed to be ready for an insertion demonstrated a certain level of effectiveness\(^{(1)}\). We also conducted evaluation experiments participated by visually impaired people on favorable hearing conditions under a situation where speech overlaps cannot be avoided. The results clarified the conditions which give little influence even if an overlap occurs and the conditions which must be prevented, demonstrating guidelines for system design.

Additionally, we investigated the acoustic features of easy-to-hear commentary speech and studied the feasibility of expanding service to non-sports programs in cooperation with visually impaired people and program producers.

### Robot commentary

Robot commentary is a technology to automatically generate commentary manuscripts describing the situation in games from real-time competition data provided in international sports competitions and to read out the manuscripts with synthesized speech\(^{(2)}\). We provided a robot commentary service for some sports via internet delivery at a sports event held in 2018. With the aim of enriching the content of speech and providing commentaries for more sports in an event to be held in 2020, we improved the speech algorithm and studied a way of collecting necessary data.

---

**Figure 5-2. Mechanism of automated audio description and robot commentary**

---

### Speech synthesis technology

As a practical application of the speech synthesis technology using deep neural networks (DNNs) that we developed in FY 2017, we developed the voice of CG reporter “Yomiko” for a program “News Check 11.” We prepared necessary data and modified the operation program to enable the handling of diverse speech expressions according to production effect needs and improved utterance skills such as intonations, poses and conversational tones through actual use in programs. This technology is also utilized for news reading service on the internet. Concurrently, we began developing a new speech synthesis technology that could bring further quality improvement of synthesized speech and higher efficiency of the preparation of training data for a speech model, which currently takes significant costs and time\(^{(3)}\).

We developed a DNN speech synthesis technology to realize speech with an announcer-equivalent quality for a trial service in which the radio weather information of regional broadcasting stations will be partly provided by speech synthesis. We achieved high-quality speech by limiting the content of speech to weather forecast programs. We began test broadcasts in prefectoral-area radio broadcasting by the Kofu station in March 2019.

Additionally, we helped promote the use of Yomiko for internet content by using a speech rate conversion technology that we previously developed. We added a production effect to read tongue twisters fluently and a function to shorten the speech of news commentary video to fit in a specified time by varying the speech speed. We also continued to support the operation of Chinese learning applications, “Seiko Kakuninkun (Tone Checker)” and “Sorijita Kakuninkun (Retroflexion Checker),” in an Educational TV program “Learn Chinese on TV” by applying a speech processing technology that we previously developed.

---

### References


---

### 5.3 Machine translation technology

To provide information to foreigners promptly and efficiently, we are conducting research on machine translation for texts of news reports, broadcast programs and newspaper articles and on news scripts with reading assistance information to help understand Japanese news.

### Machine translation of news and program content

Broadcasters translate Japanese into foreign language to provide information for non-native speakers. For the speedy and efficient production of foreign language content, we are researching machine translation\(^{(4)}\). The mainstream of recent machine translation technologies is a method that collects a huge amount of parallel translation data and trains a translation model using neural networks. In FY 2018, we created 500,000 pairs of Japanese and English news sentences by translating Japanese news manuscripts manually. Using this data, we prototyped a Japanese-English machine translation system. In addition, we developed an interface that allows easy manual correction of the machine translation results by mapping Japanese and English words and assigning the degree of translation reliability for English words. This research was conducted in cooperation with the National Institute of Information and Communications Technology (NICT).

Furthermore, we are researching English-Spanish machine translation with the aim of supporting the production of Spanish closed captions for video-on-demand (VOD) service programs on NHK WORLD. In FY 2018, we newly prepared
Machine translation of newspaper articles

With the aim of facilitating communication between non-Japanese and Japanese in business scenes, we are researching machine translation technologies for conversations and small talk in meetings and social occasions and newspaper articles in cooperation with five external institutions. NHK is in charge of a newspaper article translation technology. In FY 2018, we conducted Japanese-English machine translation experiments using parallel data generated by automatically mapping Japanese sentences and English sentences of Jiji Press. The experiments achieved better results both in objective evaluation and subjective evaluation than existing translation systems. We also identified research subjects that we should work on in the future, such as support for translation of new words, under-translation, mistranslation, over-translation, and translation with context. This research was supported by NICT as part of a project titled “R&D of Deep Learning Technology for Advanced Multilingual Speech Translation.”

News scripts with reading assistance information

In FY 2018, we produced 64 articles of news scripts with reading assistance information for experimental purpose and investigated the effect of such news scripts for non-native speakers in Japan. The results demonstrated that area of assistance information helps improve their understanding of news. This research was conducted in cooperation with Utsunomiya University.

5.4 Information presentation technology

To convey information to all viewers including those with vision or hearing impairments in an easy-to-understand manner, we made progress in our research on a technology to generate sign language computer graphics (CG) from sports information and a technology to convey the status of sports events and motion information to tactile sensation. We also began research on information presentation using the sense of smell.

Sign language CG for presenting sports information

To enrich broadcasting services for viewers who mainly use sign language, we are researching a technology for automatically generating sign language animations using CG for information about sports events. We developed a system that automatically generates sign language CG and Japanese closed captions by combining competition data delivered during a game and templates prepared in advance and displays them on the web browser in synchronization with the game status. We conducted experiments on the real-time automatic generation and presentation for ice hockey and curling matches and exhibited the results at the NHK STRL Open House 2018. Evaluation experiments using the generated content demonstrated guidelines for the information that should be presented and a screen layout in actual service (Figure 5-3)(9).

In our research on machine translation from Japanese sentences to sign language CG for sports news, we prototyped an automatic translation system using a syntax transfer method, which changes the order of words by converting Japanese syntactic structures to sign language ones, to support sentences with complicated syntactic structures. The conversion of syntactic structures uses data applied with machine learning from the results of Japanese syntax analysis and the results of sign language syntax analysis that we developed in FY 2017. We also developed a sign language CG production assistance system that allows the user to modify translation errors manually by changing the order of sign language words or replacing an incorrect word with a correct one. The results of evaluation experiments using sign language CG animations generated by this system demonstrated the need for a function to appropriately present the words or phrases to be replaced for efficient modification(2). We also began developing a function to add appropriate facial expressions in accordance with the impression of the context and words and studying evaluation experiments on the production assistance system assuming actual operation.

With the aim of expanding the parallel corpus necessary for improving the accuracy of machine translation, we began R&D on a technology to convert sign language video into text using image recognition. We verified previous studies on sign language recognition using deep learning and conducted evaluation experiments using training data for Japanese Sign Language (JSL). We confirmed the effectiveness of deep learning and obtained knowledge about training data necessary for improving the recognition rate of JSL.

We have been gathering user feedback about the understandability of our weather report sign language CG for the Kanto Region via an evaluation website that we released in NHK Online in February 2017. To expand the coverage area of scheduled forecasts from the prefectural capitals of the seven prefectures in the Kanto Region to those of the nation’s 47 prefectures, we prototyped a weather report sign language CG generation system and verified its operation. Part of this study was conducted in cooperation with Kogakuin University.
Haptic presentation technology for touchable TV

We are researching a technology for conveying the information of movements in video to people’s skin. Our main purpose is to convey information such as the moving directions of a ball and players and the timings when the ball hits the floor, wall or racket in fast-moving sports content that is difficult to convey with speech information. For the tactile presentation of motions and timings, we decided to use three types of stimuli, vibration, sliding and acceleration. In FY 2017, we identified fundamental conditions for the perception and discrimination of these stimuli and demonstrated the feasibility of conveying motions and timings. In FY 2018, we conducted experiments by artificially adding tactile stimuli to the video and speech of virtual sports content to verify whether it is possible to understand the flow of the game and which team has scored a point. The experiments, participated by visually impaired people, demonstrated that it is possible to understand the game status even without visual information. An experiment in which the ball moving back and forth in tennis was presented by sliding, the linear movement of a pressure stimulus to the skin, identified perception and recognition characteristics such as the difficulty in conveying the changes of motion speed while demonstrating the feasibility of conveying the direction of ball motion.

We developed a wristband haptic device that gives stimuli to the wrist using four vibrators through such improvement of the cube vibration device developed in FY 2017 as unnecessary to hold it by hand (Figure 5-4). Using this device, we expressed player actions in volleyball (serve, receive, set, attack) and the ball hitting in or outside of the line by vibration. The results of evaluation experiments participated by people with visual impairment and people with hearing impairment demonstrated that a certain level of information can be conveyed. For an experiment on simulating and conveying impact by using vibration with large acceleration, we developed a ball-type haptic device that presents vibrations on the palm (Figure 5-5) and confirmed the possibility of understanding the game status without visual information by varying and combining parameters such as the size of amplitude to be presented and the volume and type of sound information. We plan to continue our research to achieve more enriched content representation using AR/VR technology and real-time presentation from sports event information. Part of this research was conducted in cooperation with the University of Tokyo and Niigata University.

In our research on a technology for effectively presenting 2D information that is difficult to describe in words, such as diagrams and graphs, to people with visual impairment, we continued with our development of a finger-leading presentation system with a tactile display. This system combines a tactile display that conveys information using the unevenness and vibrations of pin arrays that move up and down and a method for conveying important points by leading fingers with a kinetic robot arm. In FY 2018, we conducted development to enable the system to be used for many purposes, including helping visually impaired people learn characters and assisting deaf-blind people with their communication. A common way to convey characters to visually impaired people who have not mastered braille or deaf-blind people is for a caregiver to trace characters with his/her finger on the deaf-blind person’s palm. We developed a method for guiding the user’s finger according to the stroke order of characters and demonstrated through evaluation experiments that this method achieves a higher character recognition rate than that of character-tracing on the palm. This showed that this system is effective not only as a learning tool but for helping deaf-blind people with their communication. Part of this research was conducted in cooperation with Tsukuba University of Technology.

Olfactory information presentation method

To provide richer viewing experiences, we began research on an olfactory information presentation method. We investigated previous studies and the latest trend of olfactory information presentation technologies and studied broadcast content for which the addition of olfactory information is effective. Also, we began studying a method for effective olfactory information presentation with simultaneous presentation of video.

[References]