Doc T-25/10

REPORT OF TOPIC CHAIRMAN
PRODUCTION TOPIC AREA

Topic Chairman	**: Kazim PEKTAS – TRT-Türkiye**
Period of Report	**: April 2025 to September 2025**
Date of Report	**: 11 September, 2025**

Reports from Project Managers in Topic on Production

A total of four projects are under the Production Topic Area, as below:

1. **Ultra High Definition TV and Advanced Immersive Sensory Media Systems P/UHDAISM**
2. **Systems for people with special needs P/SN**
3. **IT -based Production System and file-based workflows P/ITPS**
4. **Big Data, Personalisation and Artificial Intelligence P/BDAI**

For this reporting period I have received five contributions from NHK Japan:

Project topic: Ultra High Definition TV and Advanced Immersive Sensory Media Systems (P/UHDAISM)
Project Manager: ***Mr. Kimihiro Tomiyama***(NHK, Japan)

- **Harmonized Operation between Audio and Video Renderers for Presentation of Immersive Content**
- **Content-Driven AR Communication in the Coming Future**
- **3D Display Using Light Source Array Toward realizing 3D/2D switchable display**

Project topic: Systems for people with special needs(P/SN)
Project Manager: ***Ms. Reiko Takou*** (NHK, Japan)

- **Audio-Description Production Technology for Live TV Sports Programs**

Project topic: Big Data, Personalisation and Artificial Intelligence (P/BDAI)
Co Project Manager: ***Mr. Taro Miyazaki*** (NHK, Japan)

- **News Metadata Generation System**

Harmonized Operation between Audio and Video Renderers for Presentation of Immersive Content

1. Toward live-action immersive content where users can move freely looking around and listening

NHK has developed the technology for the harmonized operation between audio and video renderers, which enables users to watch videos and listen to audio from their preferred locations and orientations within a virtual space on a variety of devices. Figure 1 depicts the overview of the developed system. In the system, video data of immersive content is composed of scene description and video objects. Audio data of immersive content is composed of audio metadata and audio signals. A video renderer and an audio renderer process video data and audio data according to user's location and orientation, respectively. Then, the video and audio corresponding to user's location and orientation are produced. Additionally, by rendering video according to the display capabilities of each device, such as a head-mounted display (HMD) or tablet, users can enjoy immersive content on their preferred devices.

The utilized technologies in the system comply with the latest MPEG standards, including the scene description (ISO/IEC 23090-14, MPEG-I Part 14) to arrange video objects in a time-series space and the audio rendering method for Virtual Reality (VR) content (ISO/IEC 23090-4, MPEG-I Part 4). Furthermore, some features and functionalities have been independently expanded.

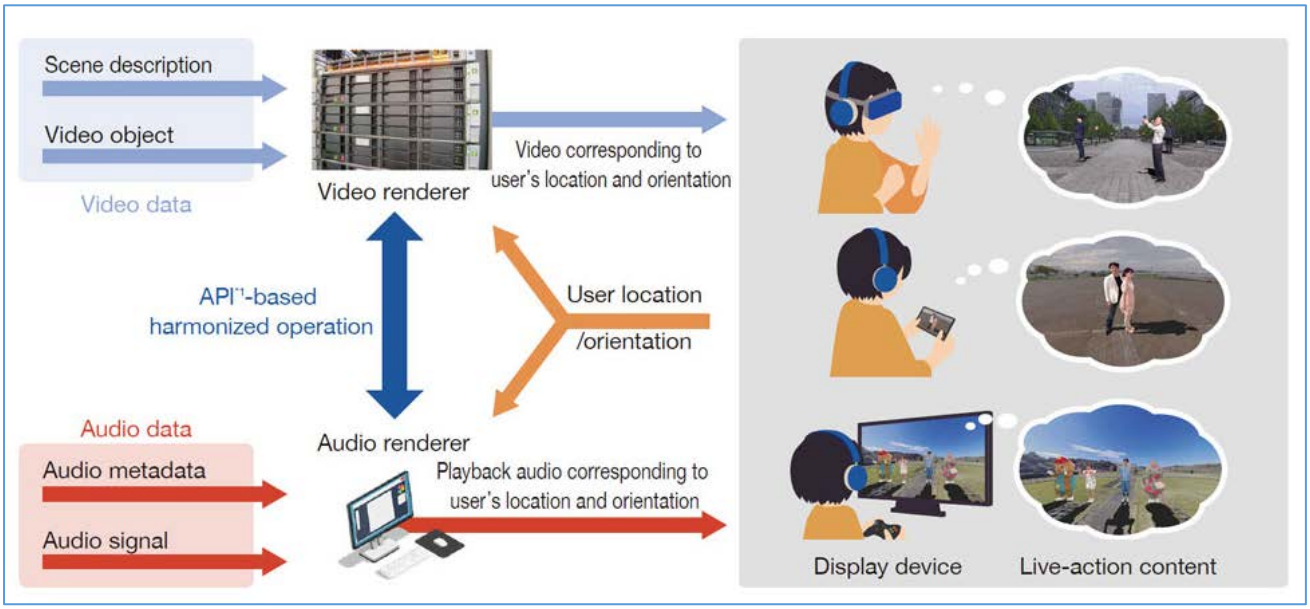


Figure 1: Overview of developed system

2. Developed technology for the harmonized operation between audio and video renderers

They have developed a mechanism to harmonize a video renderer and an audio renderer by having them share the user's position information to enable the presentation of the video and audio in a synchronized manner. Figure 2 shows the protocol among the audio renderer, video renderer, and user's device.

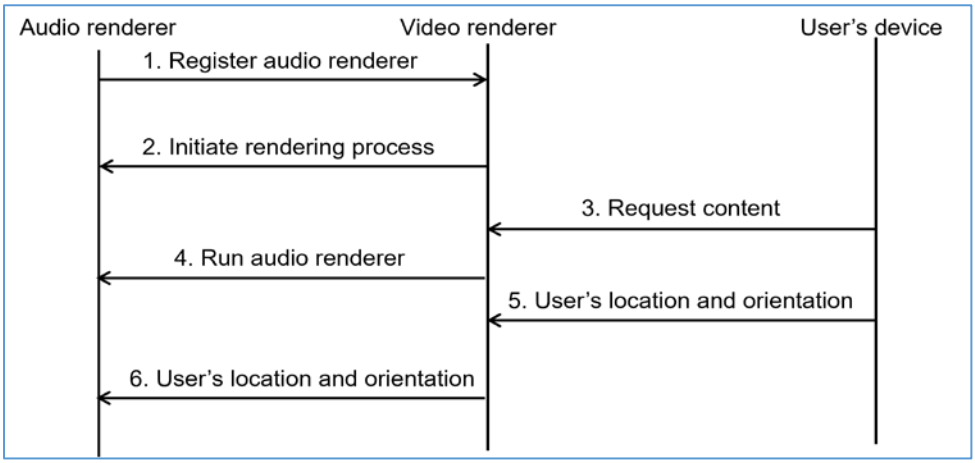


Figure 2: Protocol among audio renderer, video renderer, and user's device

1. The audio renderer registers itself to the video renderer. This allows the video renderer to identify the audio renderer with which it should be paired.

2. In response to the registration, the video renderer notifies the audio renderer that the rendering process to be initiated.
3. The user's device requests the content user want to enjoy to the video renderer. On the basis of the request, the video renderer initiates rendering the video data.
4. The video renderer runs the audio renderer while it starts playing.
5. The user's device continuously informs the video renderer of the user's viewing location and orientation.
6. The video renderer shares the user's location and orientation with the audio renderer.

Owing to this mechanism, the video and audio corresponding to the user's location and orientation are output from video renderer and audio renderer and they are presented at the user's device in synchronization with each other. Our implemented user's devices are shown in Figure 3.

(a) HMD with headphone

(b) Controller with headphone

(c) Tablet device with headphone

Figure 3: Implemented user's devices

Content-Driven AR Communication in the Coming Future

As a concept for future content, NHK STRL is introducing the image of content that expands spatially through AR glasses. You can experience how to operate the content, react to each scene, and communicate while watching the content.

Figure 4: AR communication in the near future to enjoy content together

- Sharing of the viewing experience beyond space
  You will be able to communicate beyond space while viewing the same content. Sharing each other's view will create conversations and discoveries, triggered by the content.
- Enhancement in the operability and convenience of video content
  New functions, such as gesture recognition, will make basic operations, such as playing back the scene you want to watch again and inputting your reactions, simpler while you enjoy content.

- Display of scene-linked information
  You will be able to see other viewers' reactions to the scene you are watching by displaying them around the content.

3D Display Using Light Source Array Toward realizing 3D/2D switchable display

NHK STRL is researching technology to display natural three-dimensional (3D) images according to the viewer's eye position. By using an electrically switchable light source array*1, a display that allows users to select and view 3D or two-dimensional (2D) images according to user preference has been developed.

3D images can be displayed through the reconstruction of light rays in various directions by displaying light source array images on an LCD panel for light-ray control. When a white image is displayed on the LCD panel and functions as a plane light source, 2D images can be displayed, enabling 3D/2D switching or 3D/2D simultaneous display according to user preference.

By estimating the viewer's eye position using a camera and generating elemental image array*2 in real time according to the eye position, a 3D image display with a wide viewing zone*3 has been developed.

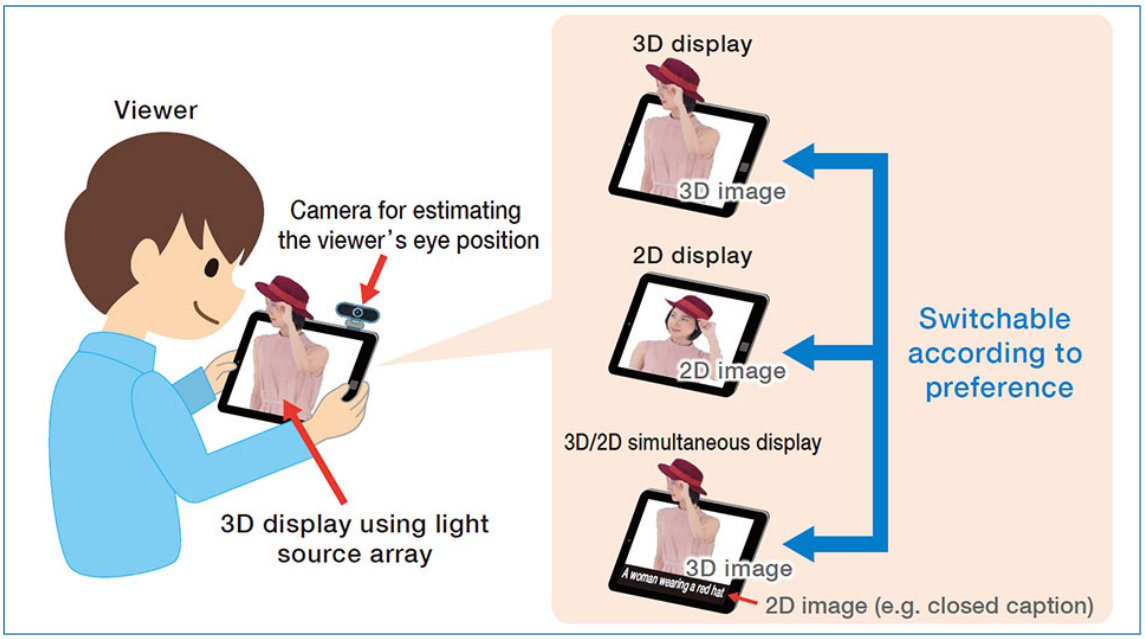


Figure 5: 3D/2D switchable eye-tracking three-dimensional display

*1: Light source array: A number of light sources arranged at equal intervals for reconstruction of 3D images.
*2: Elemental image array: An image generated to reproduce light rays emitted from an object in various directions
*3: Viewing zone: The range in which 3D images can be viewed.

Audio-Description Production Technology for Live TV Sports Programs

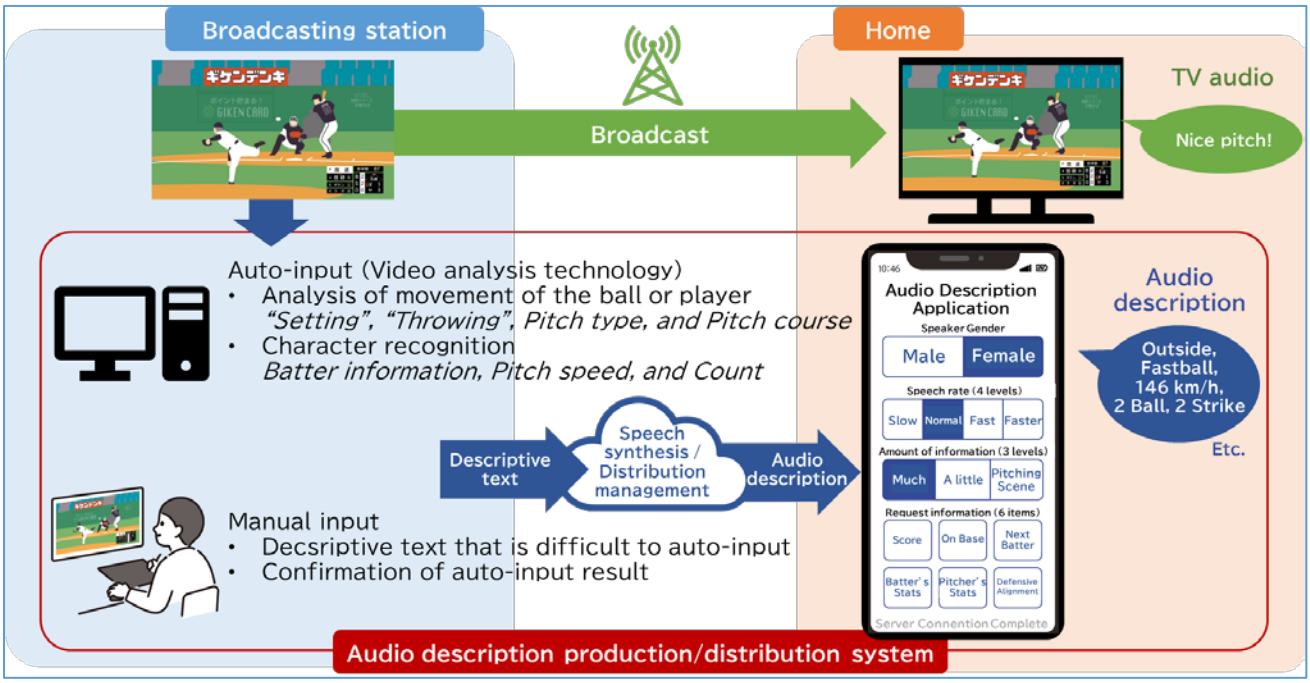
NHK is developing a system to produce and deliver audio descriptions for live sports broadcasting

programs to help more people, including those with visual impairments, fully understand such programs.

Features

- This system generates descriptive texts about plays and game conditions, converts them into audio descriptions, and delivers them to the audio description app before the broadcast video is displayed on the home's television. The audio descriptions are then played in time to match the TV program at home.
- The descriptive texts are either generated by the video/image analysis technology they developed or manually entered by the operator using their input tool.
- Their current focus is live baseball broadcasts. The video analysis technology identifies player's movements ("setting" and "throwing"), pitch types ("breaking ball," "fastball"), and pitch course (e.g., "high inside," "low outside"). The image analysis technology recognizes batter information (such as player name, uniform number, right/left-handed hitter), pitch speed, and count from visual information overlaid on the video.
- The audio description app allows users to adjust the speech rate or amount of information. Users can also listen to information such as score, on-base data, and next batter, which are extracted from external data by the system, at any time they like.

In future work, they will focus on the practical application of an audio description production and distribution system for baseball, while also expanding its scope to other sports and advancing R&D on technologies for efficient audio description production.



Overview of audio description production and distribution system

News Metadata Generation System

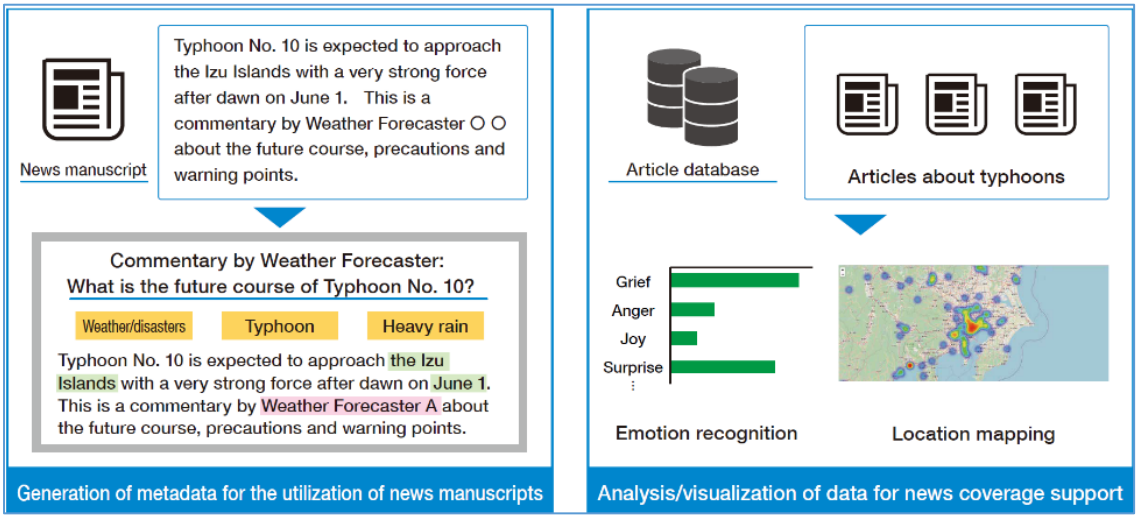
NHK Science and Technology Research Laboratories (NHK-STRL) is conducting research on natural language processing technologies to enhance the efficiency of news content production. This article introduces a system designed to generate expressive content with minimal manual effort by automatically producing metadata for news manuscripts and visualizing the results.

Metadata Generation Using AI for News Manuscript Utilization

When publishing news manuscripts online, NHK assigns multiple labels that succinctly represent the content of each article. Given the existence of thousands of label types, manually selecting the appropriate ones is a

highly time-consuming task. To address this, the system employs AI technology to automatically assign suitable labels to news articles.
NHK-STRL has also developed a system that automates various processes essential to news production, including title generation, keyword extraction, and article anonymization. By combining human oversight with system-generated outputs, the overall efficiency of content production can be significantly improved.

Data Analysis and Visualization to Support News Coverage
In the event of a large-scale disaster or similar situation, the system visualizes the regions begin covered, allowing users to quickly understand the coverage status and plan strategic reporting accordingly. Furthermore, the system visualizes the emotional polarity (positive or negative) of news articles related to each region using emotion recognition technology, thereby contributing to the identification of local issues.



Future Outlook
NHK-STRL aims to deploy these systems in production environments by 2026. By around 2028, the goal is to establish technologies capable of simultaneously processing not only natural language but also other types of information, such as image data.