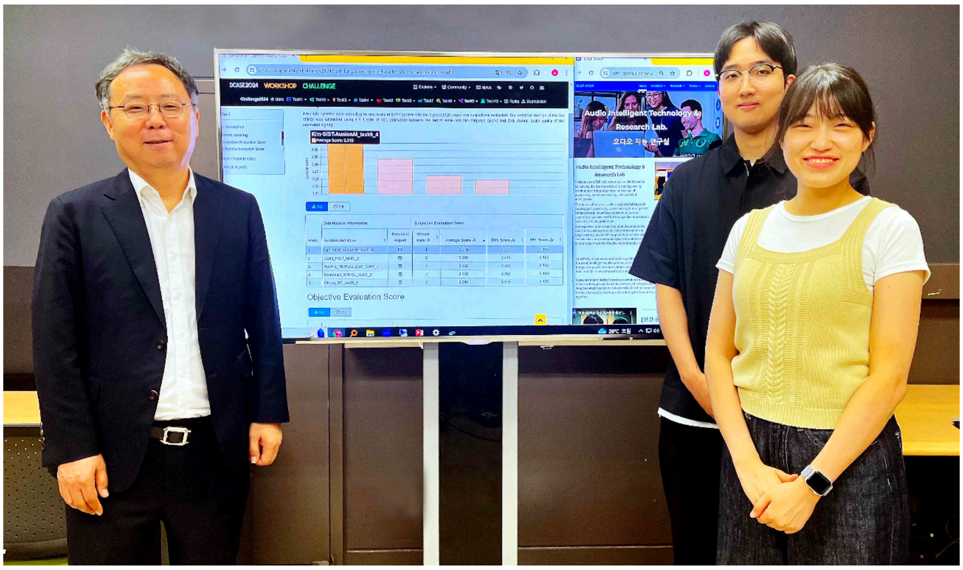
GIST proves its research capabilities in the field of artificial intelligence audio, taking first place in the international AI sound recognition competition “IEEE DCASE 2024 Challenge”

- Professor HongKook Kim's research team in the School of Electrical Engineering and Computer Science achieved first place in the language-queried audio source separation category
- Results of presenting a new technique for data augmentation based on Large Language Model (LLM)



▲ (From the left) School of Electrical Engineering and Computer Science Professor HongKook Kim, student Dohyeon Lee, and student Yuna Song

The Gwangju Institute of Science and Technology (GIST, President Kichul Lim) is attracting attention for its outstanding research achievements in the field of artificial intelligence (AI)-based voice signal processing.

Professor Kim Hong-guk's research team in the School of Electrical Engineering and Computer Science, who researches audio intelligence, won the 'Language-Queried Audio Source Separation (DCASE) IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events' competition. It ranked 1st in the ‘Audio Source Separation’ category and 3rd in the ‘Indoor Acoustic Event Detection’ category.

This is a global competition held since 2013 by the Signal Processing Society (AASP) under the Institute of Electrical and Electronics Engineers (IEEE). It competes on acoustic recognition technology that uses artificial intelligence (AI) to listen to sounds and judge the situation. This year, for about three months from April 1 to June 15, 108 leading institutions and universities participated and competed in 10 fields, including acoustic scene recognition and machine abnormality diagnosis confirmation.

The 'GIST-AunionAI' team, composed of Audio Intelligence Research Laboratory (AiTeR) students (integrated student Yuna Song, integrated student Dohyeon, and Professor HongKook Kim), supported by Professor Kim's startup company, AunionAI, won first place in the 'Language-Qualified Audio Source Separation, Task 9' category of the 'DCASE Challenge 2024', demonstrating the excellence of their research.

In addition, the GIST-HanwhaVision team formed together with Hanwha Vision researchers (GIST: master's student Sang-won Sang, integrated student Jong-yeon Park, Professor HongKook Kim; HanwhaVision: Executive Director Seung-in Noh, Senior Researcher Jeong-eun Lim, and Sulaiman Vesal) worked on the acoustic event detection task (Ranked 3rd in ‘DCASE Task 4)’.

‘Language query-based audio source separation (LASS) technology’ is a technology that separates audio signals according to the text entered by the user. By separating and generating audio sources through text queries, it provides the basis for developing a generative AI model that connects language and audio, and it can be used in various application fields such as automatic audio editing, multimedia content search, and augmented listening.

In this competition, the ‘GIST-AunionAI’ team developed a high-performance language query-based audio source separation technology by combining AI technologies that can express various audio intelligence. The AI model was improved through ▲ LLM (Large Language Model)-based prompt technology and data augmentation technology, ▲ pre-learning training model* and inference result fusion technology of existing models, and ▲ ensemble technology to improve AI capabilities.

* pre-learning training model: A large model learned on a large data set

‘Indoor and outdoor acoustic event detection technology’ is a technology that detects and distinguishes 27 different types of sounds that can occur in indoor and outdoor environments, including the sound of a vacuum cleaner, the sound of washing dishes, and the sound of a vehicle, through AI. It has the advantage of being able to detect acoustic events using only sound in situations where processing through cameras is limited, so it can be used in a variety of applications such as indoor and outdoor situation monitoring and vehicle monitoring.

The GIST-HanwhaVision team developed high-performance indoor and outdoor acoustic event detection technology by combining AI technologies that can express various audio intelligence. Excellent results were achieved by improving the AI model through ▲ auxiliary classifier-based model learning technology and ▲ various input characteristic extraction technology.

Professor HongKook Kim said, "As a result of collaboration between GIST Lab, AunionAI Co., Ltd., and Hanwha Vision, the AI model developed is very significant in the possibility that it can advance to commercialization rather than staying in the lab. In particular, we will continue to strive to improve the LLM-based audio generation and recognition AI model and apply it to various fields to contribute to the development of technology for a convenient and safe life."

The students of the 'GIST-AunionAI' team said, "Thanks to Professor HongKook Kim's guidance and generous support, we were able to achieve good results in the international competition. Not satisfied with this achievement, we will accelerate our research to continuously develop AI models for audio intelligence."

GIST's Audio Intelligence Research Laboratory (AiTeR), headed by Professor HongKook Kim, is researching various AI models related to speech and audio, and is collaborating with domestic industries, universities, research institutes, and overseas research institutes such as the Massachusetts Institute of Technology (MIT) to conduct research on acoustic event detection, speech synthesis, speech denoising, speech recognition, abnormal situation detection, multilingual recognition and translation, as well as speech source separation based on language quality.

'Audio source separation based on language quality' is an MIT international collaborative research project, supported by the GIST Science and Technology Innovation Company's 'Commercialization R&D Project' and the R&D Special Zone Promotion Foundation's 'Science and Technology Project to Open the Future of the Region', and 'Acoustic event detection research' is supported by HanwhaVision, the Ministry of Science, ICT and Future Planning and Evaluation Institute of Information and Communication, and the 'Media content voice language localization technology development project'.

GIST
Since 1993