

英 옥스퍼드 대학 주최, 「VoxSRC 챌린지 2022」

지스트 오디오지능연구실 3위 수상

- 전기전자컴퓨터공학부 김홍국 교수 연구팀 학생 및 연구원 화자 분할 부문서 3위 수상
- 대화하는 여러 화자의 음성에서 사람별 음성 구간 분류하는 인공지능 모델 설계



▲ 전기전자컴퓨터공학부 김홍국 교수의 오디오지능연구실(AiTeR) 소속 학생 및 연구원으로 구성된 GIST-AiTeR 팀이 영국 옥스퍼드 대학 VGG 그룹 주최로 진행된 「VoxSRC 챌린지 2022」 트랙4 '화자 분할 부문'에서 3위를 수상했다. (오른쪽부터 석사과정 박경완 학생, 석박통합과정 박동건 학생, 유예찬 연구원, 석사과정 김지원 학생)

오디오지능 분야를 연구하는 지스트(광주과학기술원, 총장 김기선) 학생들이 **국제 화자 인식(Speaker Recognition) 경진대회**에서 입상했다.

전기전자컴퓨터공학부 김홍국 교수의 오디오지능연구실(AiTeR) 소속 학생 및 연구원으로 구성된 **GIST-AiTeR 팀**(석박통합과정 박동건 학생, 석사과정 박경완, 김지원 학생, 유예찬 연구원)은 영국 옥스퍼드 대학 VGG 그룹 주최로 7월 19일부터 8월 14일까지 온라인으로 진행된 「VoxSRC 챌린지 2022」* 트랙4 '화자 분할 부문'에서 3위를 수상했다.

* **VoxSRC(VoxCeleb Speaker Recognition Challenge)**: 영상처리 AI 모델을 연구 개발하는 영국 옥스퍼드 대학 VGG(Visual Geometry Group) 그룹이 다양한 환경에서 실제로 녹음된 빅데이터를 활용한 화자 인식 분야의 기술 개발을 장려하기 위해 2019년부터 매년 개최하는 국제 화자 인식 경진대회

'화자 분할 기술'은 2인 이상 복수인 화자들의 음성이 담긴 오디오를 개별 화자별로 분할해 '누가 언제 말했는지'를 효과적으로 구분하기 위한 기술이다. 이 기술은 비즈니스 회의나 인터뷰 등 복수의 화자가 대화식으로 진행하는 오디오 데이터를 인식하기 위한 음성 관련 어플리케이션에 필수적이다.

이번 경진대회에서 GIST-AiTeR 팀은 복수 화자들의 대화 음성으로부터 화자별로 음성 구간을 분류하는 기술을 개발했다. 이를 위해 ▲음성 신호에서 주변 잡음을 제거하는 기술 ▲인공지능 모델 중 트랜스포머 기법을 활용한 화자를 인식하는 기술 ▲화자별 발성 구간 판별 및 화자가 동시에 발성하는 구간을 분할하는 기술 ▲군집화 기법을 활용한 분리된 발성 구간에 대해 오류를 보정하는 기술 ▲인공지능의 능력을 향상시키기 위한 다양한 기법들을 병합하는 앙상블 기술을 포함하는 인공지능 모델로 문제를 해결했다.



▲ 「VoxSRC 챌린지 2022」에 참여한 GIST-AiTeR 팀의 박동건 석박사통합과정생이 발표를 하고 있다.

GIST-AiTeR 팀의 참여 학생들은 “지도교수님과 학교 측의 아낌없는 지원 덕분에 국제대회에서 좋은 성과를 얻을 수 있었다”며 “이번 대회를 준비하며 얻은 경험과 개발된 기술을 인공지능 기반 다화자 대화체 음성인식, 화자 변화에 적응적인 음성합성 등에 접목하여 실생활에서 활용 가능하도록 성능 개선을 통해 우수한 논문 발표와 상용화에 노력하겠다” 고 말했다.

이번 대회는 화자 인식 데이터를 활용한 인공지능 학습 모델 개발을 위해 비즈니스 회의 또는 인터뷰 데이터를 이용한 화자 분류에 대한 목적으로 열렸으며, 시상식은 9월 22일(목) 인천 송도컨벤시아에서 열린 음성 및 AI 관련 국제학술대회인 <인터스피치(Interspeech) 2022 학회>에서 진행됐다.

한편 GIST-AiTeR 팀이 소속된 지스트 오디오지능연구실(AiTeR, 지도교수 김홍국)에서는 화자뿐만 아니라 음성 잡음 제거, 음성 인식, 음향 사건 감지, 이상 상황 감지, 성대 질환 감지 등 다양한 연구를 진행하고 있다.

이번 연구는 2022년도 문화체육관광부 및 한국콘텐츠진흥원의 인공지능 기반 어린이 독서활동 지원 로봇 및 서비스 콘텐츠 개발 사업, 과학기술정보통신부 및 정보통신기획평가원의 미디어 콘텐츠 음성 언어 현지화 기술개발 사업의 지원으로 수행됐다.