

소리 데이터 활용을 위한 주요 기술 동향에 관한 연구

이인우, 최지우

한국과학기술정보연구원

inulee@kisti.re.kr, liewhite@kisti.re.kr

A Study on the Key Technological Trends for Sound Data Utilization

Lee In Woo, Choi Ji Woo*

Korea Institute of Science and Technology Information.

요약

본 논문은 소리데이터 AI 활용을 위한 주요 기술에 대한 내용을 조사하고, 향후 집중이 필요한 연구 분야에 대한 조망을 하고자 하였다. 최근 AI 활용에 있어서 소리데이터의 중요성이 늘어나고 있으며, 이를 위한 연구개발 활성화가 필요한 시점인 만큼, 향후 체계적인 연구개발을 진행하고, 다양한 연구주체의 연구개발 활성화를 위해 최근 소리데이터 AI 활용에 있어서 주요한 기술 개발 주제를 확인하고, 이를 통해 향후 집중 연구 분야에 대한 제안을 하고자 하였다.

I. 서론

컴퓨팅 성능의 향상과 AI 기술의 진보로 다양한 AI 활용 솔루션이 선보여지고 있고, 이에 관한 관심이 그 어느 때보다 높은 상황이다. 기존에 익숙한 이미지, 텍스트를 활용한 경우는 활용이 가능한 수준의 솔루션으로 일상생활에서 쉽게 접할 수가 있으며, 음성과 관련된 AI 기술 역시 다양한 방식으로 쉬이 접할 수 있게 되었다.

이렇듯 다양한 데이터를 활용한 AI 솔루션이 주목받는 상황에서, 사람의 음성뿐만이 아닌 다양한 소리를 활용한 AI 기술에 대한 관심 역시 증대되고 있다. 특정한 소리를 통해 문제를 발견하거나 진단하는 기존의 인력 중심의 기술과 노하우에 의존하던 방식을 AI를 통해 풀어내고자 하는 노력이 이루어지고 있으며, 이러한 노력이 어느 정도 성과를 거두어 일부 활용되고 있다. 해외에서는 AI 활용을 위한 다양한 음향 데이터셋의 구축과 이를 활용하기 위한 기술 개발, 그리고 실제 적용 사례가 다수 존재하고 있으며, 국내에서도 몇몇 주요 연구기관 및 민간기업을 중심으로 관련된 연구개발이 진행되고 있다.

소리는 사람이 정보를 습득하는 오감 중 청각을 통해 받아들이는 데이터로, 시각 정보를 포함한 다양한 정보와의 결합을 통해 종합적인 상황을 파악할 수 있게 해준다. AI 활용에서도 마찬가지로, 기존의 이미지, 영상만을 통해 확인하기 어려운 부분을 포함하여, 다양한 정보를 추가적으로 제공할 수 있어 그 활용도가 높은 분야로 주목받고 있다. 하지만 아직까지 국내에서는 연구가 각자의 분야에서 다양하게 이루어지고는 있으나, 폭넓은 연구가 가능한 데이터셋의 구축이나, 관련된 체계적인 연구개발 방향이 잡혀있지 않은 상황이다.

본 논문에서는 소리와 AI 활용을 위해 필요한 다양한 요소기술 중 주요 기술의 동향에 대한 조사를 진행하여, 현재 관심이 집중된 기술 분야를 조망하고, 앞으로의 기술 개발 방향성에 대한 제언을 제공하고자 한다.

II. 본론

소리데이터 AI 활용을 위해 필요한 기술은 매우 다양하다. 소리데이터 AI 기술이라는 큰 카테고리의 기술을 구성하는 구성요소로서의 기술들, 즉 요소기술을 조사하기 위해서는 우선 그 기준이 필요하다. 데이터 과학에서 이해가 쉬운 기준으로는, 데이터 생애주기가 있다. 일반적으로 데이터의 생성부터 파기까지의 과정을 의미하여, 데이터 특성이나 활용 목적, 세부 프로세스 등에 따라 몇 가지 단계로 나눌 수 있다. 소리의 활용이라는 측면에서는 소리데이터의 수집, 처리, 저장, 분석의 단계 정도로 간략하게 정리할 수 있으며, 이를 기준으로 관련 기술을 파악하였다.

소리데이터의 수집, 처리, 저장의 단계에서는 매우 일반적이며, 널리 알려진 기술이 위주로 활용되고 있다. 따라서 소리데이터 AI 활용에 직접적으로 관련이 있으면서, 최신 기술 동향 반영이 가능한 분석 단계의 기술을 위주로 서술한다.

실제 AI를 활용한 소리 분석과 관련하여 주요하게 다루어지는 기술 중 음향 처리와 관련된 부분에는 음향 이벤트 감지, 음원 분리, 음원 위치 추정 등이 있다.

음향 이벤트 감지는 특정한 오디오 이벤트를 식별하는 기술로, 실제 환경에서는 다양한 소음 속에서 특정한 소리를 감지하거나 분류하기 위해 사용되는 기술이다. 이를 위해 Time-Dependent한 신호 데이터 처리에 효과적인 RNN과 스펙트럼 패턴 인식에 유용한 CNN을 주로 활용하고 있다. 기존에는 인위적으로 생성된 데이터셋을 활용하였으나, 최근에는 실제 복잡한 환경에서 수집된 라벨링 된 데이터를 활용하는 노력이 진행되고 있다.

음원 분리는 다양한 음향이 혼재된 데이터에서 특정한 음원을 분리해 내는 기술로, 복잡한 음향 환경에서 원하는 소리를 추출하여 활용할 수 있도록 돕는다. 이 경우 주로 시간-주파수 분석을 기반으로 한 신경망을 사용하며, 음향 신호를 푸리에 변환 등을 통해 변환 후, 이를 비지도 학습 등을 통해 다양한 각 음원의 특징을 자동으로 학습하게 된다.

음원 위치 추정은 특정한 소리가 발생하는 위치를 찾아내는 기술로, 기존

의 음향 이벤트 감지에 더해 정확한 위치 파악까지 가능하게 한다. 소리 신호가 가지는 방향성과 거리 정보를 분석하며, 단일 음원의 위치 추정부터, 다수 음원 위치 추정까지 이루어진다. 이를 위해서는 다채널 입력 데이터가 필요하며, 이를 통해 방향, 높이, 거리를 추정할 수 있다. 마이크 어레이를 활용한 빔포밍 기술이나 시간차를 이용한 TDOA 기술이 있으며, 멀티 채널 데이터를 활용한 딥러닝 학습을 통해 방향과 거리 추정을 하기도 한다. 이러한 기술들은 사건, 보안 침입 감지, 위험한 상황이나 환경 모니터링, 의료 분야 등에서 적용이 될 수 있다.

소리데이터 AI 활용에 있어서 가장 큰 장애는 데이터 확보와 관련된 부분인데, 이와 관련하여 실제 데이터 확보에는 한계가 있는 만큼 데이터 증강과 관련된 부분이 중요하다. 기존의 음향 데이터를 자르거나, 늘리는 등의 기본적인 방식 이외에, 복잡한 환경적인 요소를 더하기 위한 공간음향 기술이 있다. 공간음향 기술은 청취자가 현실과 유사한 음향을 경험할 수 있도록 청각적인 정보를 현실과 최대한 유사하게 재현하기 위한 기술이다 [1]. 공간에서의 음향 특성은 음파가 접하는 경계면의 소재 등에 따라 흡수(Absorption)와 반사(Reflection), 산란(Scattering)과 확산(Diffusion)의 특성이 달라질 수 있다. 이러한 요인으로 인해 공간에서는 복잡한 음향 환경이 만들어지게 되는데, 이러한 요인으로 인해 음향 탐지, 음원 위치 추정 모델의 정확도에 부정적인 영향을 미칠 수 있다. 성능 향상을 위해서는 이러한 부분에 대한 데이터 보강이 필요하며, 다양한 환경에서의 공간 음향 시뮬레이션 기술을 적용한 데이터 활용이 가능하다. 공간음향 시뮬레이션에는 음파 기반 음향 모델과 기하 음향 모델이 있는데, 음파 기반 음향 모델의 경우 저주파와 복잡한 지오메트리에 대해 더 정확하지만, 계산 집약적이고, 기하 음향 모델의 경우 고주파와 단순한 공간 형태에 대해서는 더 빠르지만 복잡한 음향 환경에서는 다소 부정확한 모습을 보인다. 음파 기반 모델에는 FEM(finite Element Method), FDTD(Finite Difference Time Domain)이 포함되며, 공간을 작은 섹션을 나누어 음파가 각 섹션에서 시간에 따라 어떻게 이동하는지 수식을 통해 계산한다. 기하 모델에는 레이트레이싱과 ISM(Image Source Method)이 있는데, 레이트레이싱은 개별 Sound Ray 경로를 추적하여 반사, 회절, 흡수 기반으로 전파 방향과 강도를 예측하는 방법이며, ISM은 레이트레이싱의 간소화된 버전이라고 할 수 있다[2][3].

III. 결론

본 논문에서는 소리데이터 AI 기술 중 최근 주로 다루어지는 음향 이벤트 감지 기술, 음원 분리 기술, 음원 위치 추정 기술, 공간음향 기술을 확인하였다.

음향 이벤트 감지, 음원 위치 추정 기술과 관련하여서는 앞으로 실제, 실시간 데이터베이스 구축 및 처리, 그리고 적용된 딥러닝 모델의 범용성 및 강건성 확보, 실용성을 위한 모델 최적화 등의 과제가 있다[4]. 이러한 과제를 포함한 AI를 활용한 음향 처리 기술은 DCASE Challenge와 같은 주요 연구 커뮤니티에서 이루어지는 챌린지에서도 다루어지고 있으며, 국내 연구자들도 이들 챌린지에서 좋은 성과를 거두고 있다. 소리데이터 활용에 관해 AI와 데이터 연구자들의 관심이 모이고, 국내 연구 커뮤니티가 활성화된다면, 국내에서도 관련 연구 생태계가 활성화될 수 있을 것으로 보인다.

공간음향 기술은 역사는 오래되었으나, 음향 콘텐츠 제작을 위한 부분에서 주로 다루어져 왔다. 최근에는 음향 콘텐츠 제작, 그리고 재생 기술에 활용되는 상황까지 왔는데, 이를 활용하여 소리데이터 AI 활용에 도움이 될 수 있을 것으로 보인다. 음향처리 관련 기술 연구에 있어서 활용 가능

한 데이터 양의 문제가 있는 만큼, 다량의 다양한 데이터 확보를 위해서는 시뮬레이션을 활용한 데이터 증강 기술이 더욱더 중요해질 것으로 보이며, 이에 관한 관심이 더욱더 필요할 것으로 보인다.

ACKNOWLEDGMENT

이 논문은 2023년도 한국과학기술정보연구원(KISTI)의 자체사업으로 수행된 연구입니다.(과제번호: J-23-NB-C01)

참고 문헌

- [1] 이용주 등, 체험형 미디어 서비스를 위한 공간음향 기술 동향, 전자통신동향분석, 34(30), 2019, PP.13-22.
- [2] Finnur Pind, Wave based virtual acoustics, PhD Thesis, Technical University of Denmark, 2019.
- [3] RATHNAYAKE, R. A.; WANNIARACHCHI, W. K. I. L. "Image source method based acoustic simulation for 3-D room environment," International Journal of Scientific & Technology Research 8(11), pp. 222-228, 2019.
- [4] Zaheer, Ruba, et al. "A Survey on Artificial Intelligence-based Acoustic Source Identification." IEEE Access 11, 2023.