



청각 장애인을 위한 딥 러닝 기반 소리 방향 및 종류 식별 시스템

Deep Learning based Sound Direction and Classification System for Deaf People

저자 (Authors)	김세영, 김현웅, 박찬호, 정목동 Seyeong Kim, Hyeonung Kim, Chanho Park, Mokdong Chung
출처 (Source)	한국정보과학회 학술발표논문집 , 2017.06, 1896-1898 (3 pages)
발행처 (Publisher)	한국정보과학회 KOREA INFORMATION SCIENCE SOCIETY
URL	http://www.dbpia.co.kr/Article/NODE07207793
APA Style	김세영, 김현웅, 박찬호, 정목동 (2017). 청각 장애인을 위한 딥 러닝 기반 소리 방향 및 종류 식별 시스템. 한국정보과학회 학술발표논문집, 1896-1898.
이용정보 (Accessed)	경희대학교 국제캠퍼스 163.180.98.*** 2019/04/03 13:38 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독 계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

청각 장애인을 위한 딥 러닝 기반

소리 방향 및 종류 식별 시스템

김세영¹ 김현웅¹ 박찬호¹ 정목동¹¹부경대학교 컴퓨터공학과

sseyoung513@naver.com, kimhu92@gmail.com, superus@naver.com, mdchung@pknu.ac.kr

Deep Learning based Sound Direction and Classification System
for Deaf PeopleSeyeong Kim¹ Hyeonung Kim¹ Chanhoo Park¹ Mokdong Chung¹¹Pukyong National University Department of Computer Engineering

요 약

청각 장애인은 소리로 알 수 있는 수많은 정보를 놓치고 있다. 경보를 듣지 못해 위험에 빠지거나, 초인종 소리를 듣지 못해 택배 배달원을 돌려보낸다. 이를 해결하기 위한 보조기기가 많지 않고 시중에 나와 있는 제품 또한 소리의 방향만 알려주지만 할 뿐이다. 이에 본 논문에서는 소리 방향을 알려주는 청각 장애인용 보조기기에 딥 러닝을 이용한 소리 종류 식별 기능을 추가한 시스템을 제안한다. 제안한 시스템의 활용은 소리를 진동과 이미지를 통하여 알려 주므로 청각 장애인들이 더 안전하고 편리한 삶을 사는 데 도움을 줄 것이다.

1. 서 론

청각 장애인을 위한 보조기기는 소리를 통한 일상생활 정보를 알려주거나 위험 경보를 알려주기보다는 보청기 및 인공와우와 같은 의사소통을 돕는 것에 집중되어있으며 비싼 가격으로 인해 사용자가 부담을 느낀다 [1]. 또한, 사용 시 불편하거나 수년의 적응 기간이 필요하기도 하다.

따라서 청각 장애인은 소리를 통한 정보 취득에 애로 사항을 가지게 되어 다른 사람과의 대화나 일상생활에서의 소리 정보뿐만 아니라 화재정보나 자동차 경적과 같은 경고 소리도 듣기 어려워 위험한 상황에 노출된다.

국내에 유통되고 있는 장애 보조기기 1,000여 가지 중 이러한 문제점을 해결하기 위한 소리의 종류나 방향을 식별하는 보조기기는 다섯 가지 미만이었다. 더구나 소리의 방향과 종류를 함께 식별하면서 휴대할 수 있는 시스템은 찾아볼 수 없었다 [2]. 소리의 방향 식별이 가능한 휴대용 기기는 2016년 아메리카 최대의 발명 무역 박람회인 INPEX International Award에서 금메달을 수상한 유퍼스트(주)의 ‘누구나 넥밴드’ 만을 찾을 수 있었으나, 해당 제품은 소리의 종류를 알 수 없으므로 사용자가 소리의 종류를 직접 확인하는 과정에서 제품의 의도와는 달리 위험한 상황에 노출될 수가 있다는 문제가 있다 [3].

따라서 본 논문에서는 딥 러닝에 기반을 두는 소리 종류 식별 모듈을 구현하고, 소리의 방향과 종류를 식별하여 청각 장애인에게 알려 주는 통합적인 시스템을 제안한다. 이를 통해 청각 장애인이 촉각으로 소리의 방향을, 시각으로 소리의 종류를 알게 되어 더욱 안전하고 편리한 삶을 사는 데 도움을 주고자 한다.

본 논문의 구성은 다음과 같다. 2절 관련 연구에서는 청각 장애인을 위한 보조기기 및 소리 종류를 식별하기 위한 기술을 살펴보고, 3절에서는 소리 종류 식별 기능을 더한 청각 장애인을 위한 개선된 보조기기 시스템의 구성을 살펴본다. 4절에서는 제안하는 시스템 구성을 위한 소리 종류 식별 모듈의 구현 및 테스트 결과를 보이고 마지막으로 5절에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구

2.1 딥 러닝

딥 러닝은 다 계층 인공신경망 기반의 기계 학습 기술로 자율적으로 이해, 학습, 예측 적응 및 잠재적으로 독자적인 작동하는 시스템을 포괄한다[4].

딥 러닝은 물체 인식과 자동차를 위한 장애물 센서 연구를 중심으로 적용되고 있으며, 번역 및 음성인식, 이미지 식별, 인공지능 비서, 인공지능 바둑 프로그램(알파고) 등 실생활의 다양한 분야에 사용되고 있다.

본 논문에서 제안하는 시스템의 소리 종류 식별 모듈 구현에도 딥 러닝 기술을 사용한다.

2.2 TensorFlow

TensorFlow는 구글 브레인 팀에 의해서 개발된 데이터 흐름 그래프를 사용한 수치 계산을 위한 오픈소스 소프트웨어 라이브러리이다.

원래 기계 학습 및 심층 신경 네트워크 연구를 수행하기 위해 개발되었지만, 다양한 다른 영역에 적용할 수 있다. 또 오픈소스 소프트웨어 라이브러리로 만큼 수많은 기업과 개발자들이 다양한 분야에서 사용하고 있다 [5].

본 논문에서는 이 라이브러리를 활용하여 딥 러닝을 구현해 소리 종류를 식별하고자 한다.

2.3 소리 종류 분류

소리 종류 분류 연구 중에 지도학습 방법으로 UrbanSound8K [6]을 입력 데이터 집합으로 하여 딥 러닝을 통해 소리 종류 분류 모듈을 만들어 테스트한 경우가 많았다.

이 중 하나는 UrbanSound8K와 BBC SoundFX를 사용하여 입력 데이터 집합으로 삼고 2개의 은닉 계층을 두고 각각의 계층에 3000개의 뉴런을 두어 정확도는 최대 72.2%가 나왔다 [7]. 또 다른 경우는 UrbanSound8K만 입력 데이터 집합으로 삼고 각각 2개, 3

개의 은닉 계층과 각각의 계층에 수백 개의 뉴런을 두어 정확도는 약 80%대 중반이 나왔다 [8][9]. 이 연구들을 본 논문에서 구현한 소리 종류 식별 모듈과 비교하고자 한다.

3. 제안 시스템 구성

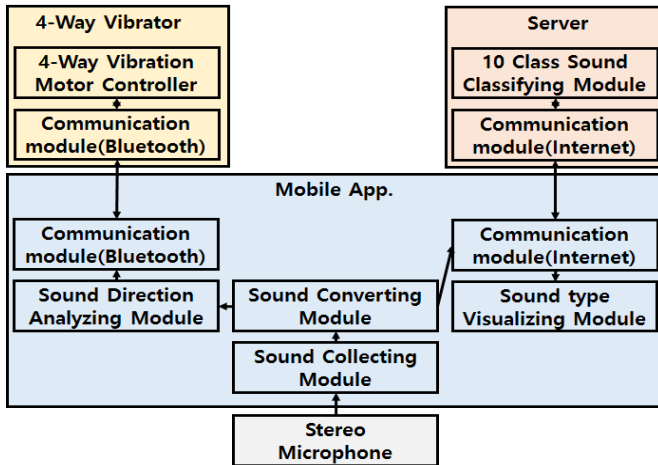


그림 1. 시스템 상세 아키텍처

본 논문의 시스템 상세 아키텍처는 그림 1과 같이 소리를 수집하는 Stereo Microphone, 소리 데이터 변환 및 방향 분석을 하는 Mobile App, 소리 방향을 알리는 4-Way Vibrator, 소리 종류를 식별하는 Server. 총 4가지 모듈 집합으로 구성되어있다.

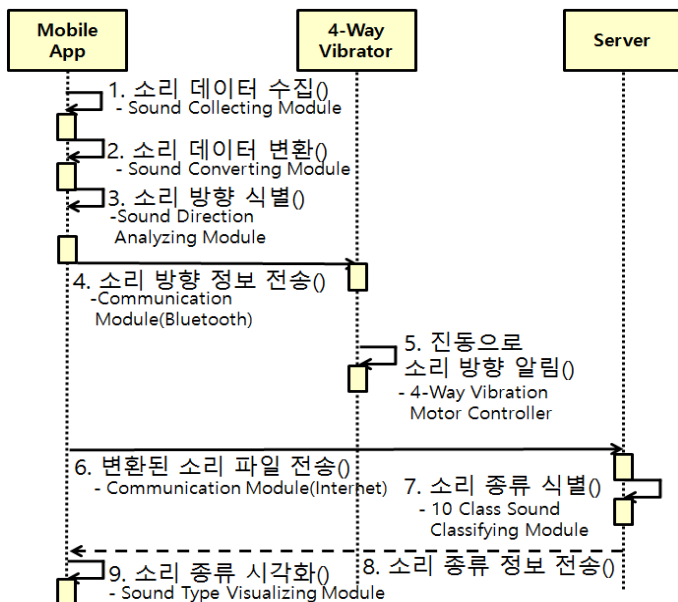


그림 2. 시스템 순차 다이어그램

그림 2는 시스템의 순차 다이어그램으로 각 모듈의 기능과 데이터의 흐름을 나타낸 것이다.

4. 소리 종류 식별 모듈 구현 및 평가

4.1 구현

4.1.1 모듈 훈련 및 테스트를 위한 데이터 집합

본 시스템은 UrbanSound8K라는 8,732개의 짧은 길이(4초 이하)인 10종류의 소리를 모아놓은 wav 데이터 집합을 이용하여 소리 중

류 식별 모듈을 구현하였다. 모든 소리 수의 20%(1,747개)를 테스트 데이터로 사용하였다. 나머지(6,985개) 중 20%(1,397개)를 유효성 확인 데이터로 하였고 모든 소리 데이터에서 이들 데이터를 빼 나머지(5,588개)를 소리 종류 식별 모듈의 훈련 데이터로 삼았다.

4.1.2 소리 종류 식별 모듈 구현 및 테스트를 위한 환경

표 1. 구현 환경

Memory	7.8 GB
Processor	Intel® Core™ i5-2400 CPU @ 3.10GHz × 4
Graphic Card	Gallium 0.4 on NVC1
OS	Ubuntu 16.04.2 LTS, 64-bit
Programming Language	python 2.7
Modelling	TensorFlow

4.1.3 소리 특징 추출 및 입력 데이터 선정

먼저 UrbanSound8K에 Python 라이브러리인 LibROSA [10]를 이용하여 특성을 추출하였다. 사용한 특성은 표 2와 같고 합하여 모두 행으로 늘어 놓아 총 193개의 특성 데이터를 만들었고 입력 데이터로 사용하였다.

표 2. 사용한 소리 특성

사용 특성 이름
Mel-frequency cepstral coefficients
chromagram from a waveform or power spectrogram
Mel-scaled power spectrogram
spectral_contrast(spectral contrast)
tonnetz(tonal centroid features)

4.1.4 딥 러닝 신경망 구성

관련 연구(2.4)에서 참고한 기존 모듈처럼 신경망은 완전 연결 신경망으로 3개의 은닉 계층에 각각 뉴런을 300개, 200개, 100개 설정하였다. 은닉 계층 및 출력 계층의 가중치 및 편향의 초기화는 평균이 0이고 표준 편차가 입력 데이터의 크기(193)의 제곱근의 역수인 정규분포를 따르는 임의의 값으로 설정한다.

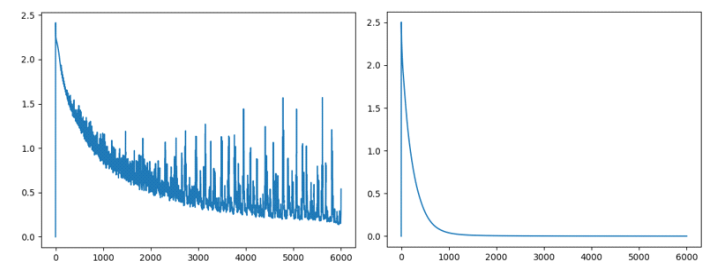


그림 3. 비용 함수 그래프, (좌) 기존 방식, (우) 제안 방식

활성화 함수와 optimizer, learning rate을 변경시켜가며 가장 정확도가 높은 모듈 구성을 해보았다. 그림 3은 활성화 함수와 optimizer, learning rate을 다르게 사용한 결과를 보여준다. 좌측은 활성화 함수로 Sigmoid, optimizer로 Gradient Descent optimizer, learning rate은 0.1을 사용한 기존 방식이고, 우측은 활성화 함수로 ReLU 함수 [11], optimizer로 Adam optimizer [12], learning rate은 0.0001을 사용한 제안 방식이다. 따라서 비용의 흔들림이 적고 빠르게 0으로 수렴해 나가는 제안 방식을 선택하였고, 훨씬 빠르고 정확도가 높았다.(상세 내용은 4.2.5 평가에서 볼 수 있다.)

제안한 방식은 첫 번째, 세 번째 은닉 계층의 활성화 함수는 ReLU 함수를, 두 번째 은닉 계층의 활성화 함수는 Hyperbolic-tan 함수를 사용하였다. 다중 종류 식별 문제를 다루기에, 상호배타적인 각 종류에 대해 확률분포가 담긴 출력 벡터를 얻기 위해 출력 계층은 softmax 활성화 함수를 사용하였다. 비용 함수는 크로스 엔트로피를 사용하고 비용 함수의 최소화를 위해 Adam optimizer

를 사용하였다. 이때 learning rate은 1/10,000이고 총 6,000회 훈련 시켰다.

4.2 평가

4.2.1 소리 종류 식별 모듈 평가

정확도는 기존 방식과 본 논문에서 구현한 방식의 소리 종류 식별 모듈들을 각각 따로 구성한 후 UrbanSound8K 데이터 집합에서 테스트 데이터로 별도로 남겨둔 1,747개의 소리로 모듈에서 추정된 소리의 추정 종류와 원래 종류의 일치 여부를 확인하여 측정하였다. 표 3은 기존 방식과 제안한 방식을 비교한 것이다. 표 3을 보면 제안한 모듈의 정확도가 약 92%로 기존 방식의 정확도보다 약 7.6% 개선이 되었다. 또한 학습률은 1/1,000배이지만 오히려 훈련 속도는 기존 방식보다 약 42% 감소하였다. 정확도 측정 시간은 기존 방식보다 약 26% 감소하였다.

표 3. 기존 방식과 제안 방식의 비교

비교 항목	기존 방식	제안 방식	개선도
정확도	84.1442 %	91.8145 %	약 7% 개선
학습률	0.1	0.0001	-
훈련 소요시간	1286.802 초	746.214 초	약 42% 개선
정확도 측정 소요시간	0.0168 초	0.0123 초	약 26% 개선

그림 4는 기존 방식과 제안 방식 각각으로 구성한 소리 종류 식별 모듈에서 정확도 측정 시 각 소리 종류별로 틀린 확률을 나타낸다. child playing과 drilling만 약 1% 정도 제안 방식의 정확도가 낮으나, 그 외의 경우 제안 방식이 정확도가 약 1%~21% 개선되었다.

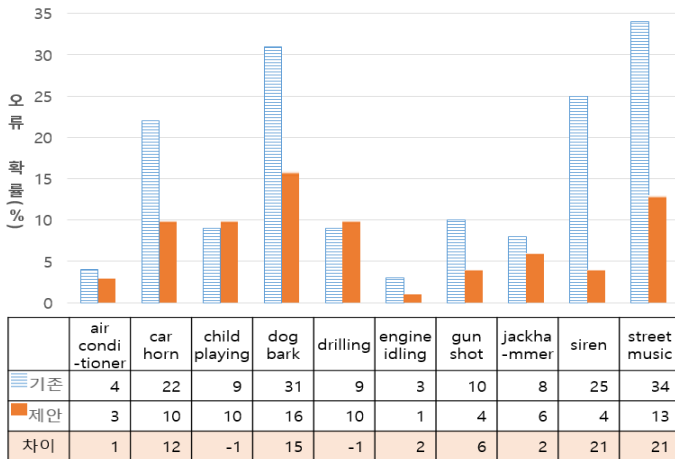


그림 4. 각 소리 종류별 식별 오류 확률

4.2.2 제안 시스템 평가

표 4. 기존 청각 보조기기와 제안 시스템의 비교

비교 목록	보청기	인공 와우	설치형 경보기	누구나 넥밴드	제안 시스템
소리 방향	○	○	×	○	○
소리 종류	○	○	×	×	○
의사소통	○	○	×	×	×
휴대성	○	○	×	○	○
편리성	×	×	○	○	○
적용 기간	×	수 년	×	×	×
낮은 가격	×	×	○	○	○
중증 장애인 사용 가능	×	○	○	○	○

표 4는 기존 시스템과 제안 시스템의 차이점을 보여준다. 제안 시스템은 소리 방향 및 종류를 식별할 수 있는 휴대용 청각 보조 기구로서 가격이 비싸서 혹은 중증 장애인이라 보청기를 못 쓰는 청각 장애인들에게 위험한 소리들의 방향과 종류를 가르쳐주어 더욱 안전한 그리고 편리한 일상생활을 도울 수 있다고 기대된다.

5. 결론 및 향후 연구

본 논문에서는 딥 러닝을 이용한 소리 종류 식별 모듈을 구성하였고 1,747개의 소리로 그 모듈의 정확도 측정 결과 기존 방식보다 약 7.6% 정확도가 향상되어 약 92%의 확률로 10가지의 소리 종류를 올바르게 예측하였다. 이를 통해 딥 러닝에 기반을 두는 소리 종류 식별 모듈을 구현하고, 소리의 방향과 종류를 식별하여 청각 장애인에게 알려 주는 통합적인 시스템을 제안하였다. 제안 시스템은 높은 정확도로 청각 장애인이나 난청인이 소리 방향뿐만 아니라 종류를 알 수 있게 된다. 즉 청각 장애인 및 난청인이 더욱 안전한 그리고 편리한 일상생활을 할 수 있을 것으로 기대된다. 특히 보청기를 사용할 수 없거나 의사소통 소리를 제외하고 잡음을 처리하는 인공와우를 사용하는 청각 장애인에게는 커다란 도움이 될 것으로 보인다.

향후에는 소리 종류 식별 모듈에 향후 Convolution neural network 등의 더욱 진보된 딥 러닝 기술을 사용하고, 더 많은 소리 데이터와 종류가 있는 Google AudioSet을 이용하여 더욱더 정확도를 높여 보고 이를 기반으로 하여 더 정확도가 높은 소리 방향 및 종류 식별 청각 보조기기를 만들고자 한다.

참 고 문 헌

- [1] 황화성, 장애인 보조기기 지원 및 활성화 방안 연구, (재)한국 장애인개발원, pp 34-40, 2016.
- [2] 경기도재활공학서비스연구지원센터, “국내유통 중인 보조기기 검색” ; <http://atrac.or.kr/support>, 2017.
- [3] 유퍼스트(주), “누구나 넥밴드” ; <http://theufirst.com>, 2017.
- [4] D. Cearley, M. Walker, B. Burke, “Trend No. 1: Artificial Intelligence and Advanced Machine Learning”, Top 10 Strategic Technology Trends for 2017, p.4, 2017.
- [5] TensorFlow, “About TensorFlow” ; <https://www.tensorflow.org/>, 2015.
- [6] J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research”, In Proceedings of the ACM International Conference on Multimedia. ACM, pp. 1041-1044, 2014.
- [7] MK Lim, DH Lee, HS Park, US Park, JH Kim, JS Park, GJ Jang, “Audio Event Classification Using Deep Neural Networks”, Journal of the Korean society of speech sciences, Vol. 7, Issue. 4, No. 4, pp. 27-33. 2015.
- [8] Aaqib Saeed, “Urban Sound Classification, Part 1” ; <https://aqibsaeed.github.io/2016-09-03-urban-sound-classification-part-1/>, 2016.
- [9] 박해선, “Urban Sound Classification” ; <https://tensorflow.blog/2016/11/06/urban-sound-classification/>, 2016
- [10] librosa development team, “LibROSA” ; <https://librosa.github.io/librosa/>, 2013.
- [11] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines”, in Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807-814, 2010.
- [12] P. Kingma and J. Ba. Adam, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.