

잡음 환경에서 Generative Adversarial Network를 이용한 소리 음질 향상

최용주*, 이종욱**, Huasang Wang*, 박대희**, 정용화**

*고려대학교 컴퓨터정보학과, **고려대학교 컴퓨터융합소프트웨어학과

e-mail: aaa928@korea.ac.kr

Sound Enhancement with Generative Adversarial Network under Noise Conditions

Yongju Choi*, Jonguk Lee*, Huasang Wang*, Daihee Park*, Yongwha Chung*

*Department of Computer and Information Science, Korea University

*Department of Computer and Convergence Software, Korea University

요 약

4차 산업혁명이 도래하면서 정보 통신 기술 및 융합 기술의 발전에 힘입어 소리 데이터를 이용한 연구가 활발하게 진행되고 있다. 소리 데이터를 이용한 학술적 프로토타입 연구들을 실제 환경에서 운용하기 위해서는 소리 취득 시 발생하는 다양한 잡음 환경에서도 원시 데이터(raw data)에 근접한 정보를 취득할 수 있는 시스템의 강인함이 보장되어야 한다. 본 논문에서는 SEGAN(Speech Enhancement Generative Adversarial Network) 모델을 활용하여, 전처리 및 후처리 과정이 필요 없이 원시 데이터를 대상으로 하는 end-to-end 방식의 소리 음질 향상 시스템을 제안한다. 제안하는 시스템은, 축산업 분야의 돼지 호흡기 질병 소리 데이터를 이용하여 실험하였으며, 여러 가지 잡음 상황(인위적인 잡음, 실제 환경 잡음)에서 소리 음질이 개선됨을 실험적으로 검증하였다.

1. 서론

전통적으로 음성인식 분야는 IT 관련 연구자들의 중요한 연구 분야 중 하나이며, 최근에는 해당 연구 결과를 실제 시스템에 적용하는 실용화 연구들이 증가하고 있다[1]. 특히 4차 산업혁명 시대의 정보 통신 기술(information communication technology) 및 융합 기술(convergence technology)의 발전에 힘입어 음성인식 분야만이 아닌 다양한 분야에서 소리 데이터를 활용한 연구들이 보고되고 있다[2-5].

축산업 분야에서 소리 데이터를 이용한 IT 관련 연구를 살펴보면, Chung 등[2]은 돼지 호흡기 질병으로 인한 기침 여부를 MFCC 특징 정보와 단일 클래스 탐지기인 SVDD(Support Vector Data Description)를 이용하여 호흡기 질병을 탐지하고 SRC(Sparse Representation Classifier)를 이용하여 질병의 종류를 분류하는 이중 구조를 제안하였다. Lee 등[3]은 돼지의 호흡기 질병을 효과적으로 탐지하기 위하여, 시간 영역과 주파수 영역의 다양한 소리 특징 중 호흡기 질병 탐지에 유효한 특징들만을 선택 및 조합하는 방법에 관한 연구 결과를 발표하였다.

위에서 언급한 동물의 발성음을 기반으로 호흡기 질병을 탐지하는 최근의 연구 동향을 살펴보면, 소리 신호해석에 의한 돼지 호흡기 질병의 탐지가 현실적으로 가능함을 확인하는 학술적 노력들이 일정 부분 성공적인 결과를 도출하였으나, 실제 돈사 환경에서 발생하는 잡음의 영향력

을 충분히 고려했다고는 볼 수 없다. 학술적 프로토타입 시스템을 실제 돈사 환경에서 성공적으로 운용하기 위해서는 소리 취득 시 발생하는 다양한 잡음에서도 강인함이 보장되어야만 한다.

최근 이와 같은 잡음에 대한 영향력을 제어하고, 잡음 환경에서 시스템의 인식 성능을 높이기 위하여 Choi 등[4]은 이미지를 질감 정보로 변환하여 잡음 문제를 해결하는 DNS(Dominant Neighborhood Structure) 알고리즘을 소리 데이터에 최초로 적용함으로써 잡음 환경에서도 시스템의 우수한 성능을 보장하는 연구 결과를 발표하였고, Lee 등[5]은 소리 데이터를 스펙트로그램으로 변환한 후, CNN(Convolutional Neural Network)의 커널 기법을 이용하여 잡음 환경에서도 이상 상황을 효과적으로 탐지하는 연구 결과를 발표하였다.

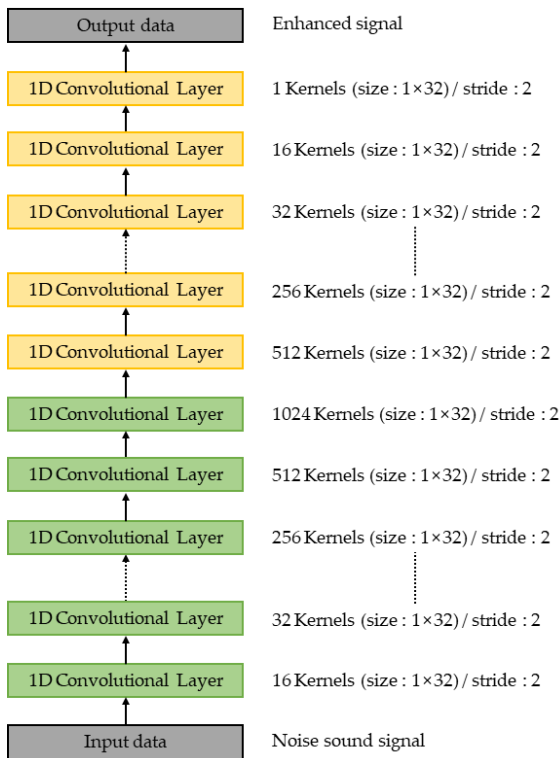
본 연구에서는 소리 데이터에 대한 변형 없이 원시 데이터 그 자체인 소리 신호에서 잡음을 제어하는 새로운 방법을 제안한다. 해당 방법론은 Pascual 등[6]이 제안한 SEGAN(Speech Enhancement Generative Adversarial Network) 모델을 활용하여, 전처리 및 후처리 과정이 필요 없이 원시 데이터를 대상으로 하는 end-to-end 방식의 소리 음질 향상 시스템(그림 1 참조)이다. 실제 돈사 환경에서 취득한 소리 정보를 이용하여 제안한 방법론의 타당성을 실험적으로 검증한다.



(그림 1) End-to-end 방식의 소리 잡음 제어

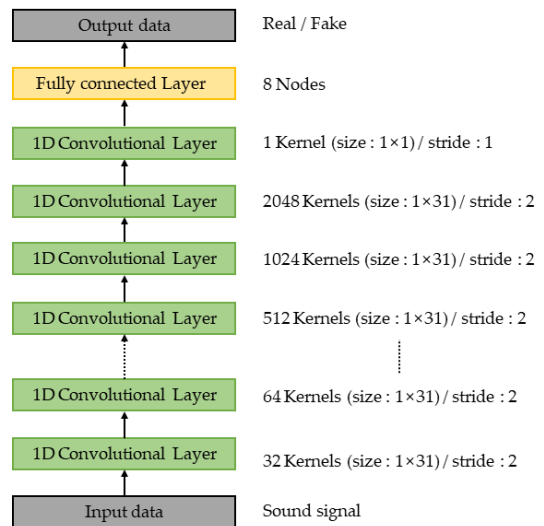
2. 제안하는 시스템

본 논문에서 제안하는 시스템의 모델 구조는 그림 2와 3과 같다. 그림 2는 소리 데이터에서 잡음을 제어하는 기능을 수행함과 동시에 새로운 소리 신호를 만들어내는 Generator 부분이며 11개의 1D convolutional layer와 11개의 1D de-convolutional layer로 구성되어있다. 모든 layer에서 커널의 크기는 1×32로 고정하였고 커널의 개수는 convolutional layer에선 {16, 32, 32, 6, 64, 128, 128, 256, 256, 512, 1024} de-convolutional layer에선 {512, 256, 256, 128, 128, 64, 64, 32, 32, 16, 1}로 설정하였다. convolutional layer와 de-convolutional layer는 서로 대칭을 이루며 skip connection 구조를 이룬다.



(그림 2) Generator 구조도

그림 3은 Generator에서 생성된 소리 신호가 진짜 잡음이 제어된 신호인지 아닌지를 판별하는 Discriminator이며 12개의 convolutional layer와 1개의 fully connected layer로 구성된다. 모든 layer에서 커널의 크기는 1×31로 고정하였고 커널의 개수는 {32, 64, 64, 128, 128, 256, 256, 512, 512, 1024, 2048, 1}로 설정하였다.



(그림 3) Discriminator 구조도

3. 실험 및 결과 분석

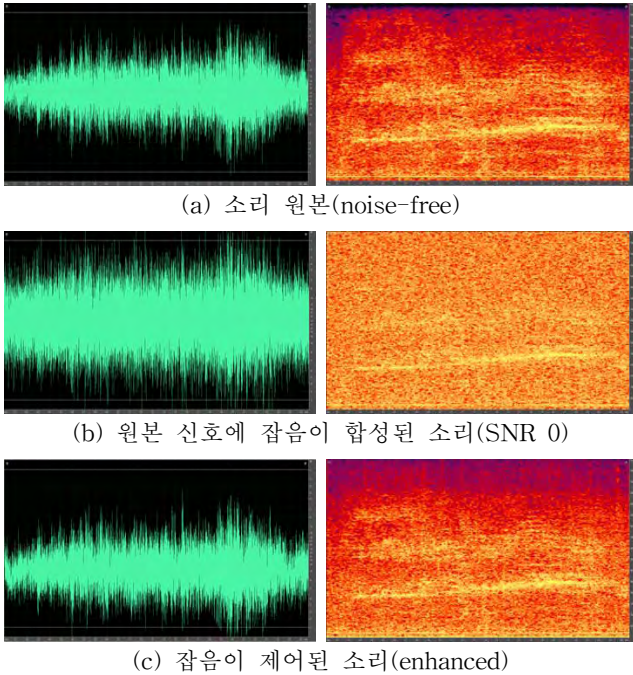
3.1 실험 데이터 획득

충청남도과 경상남도에 위치한 돈사에서 25~30kg의 총 36마리의 돼지(Yorkshire, Landrace, Duroc)을 대상으로 실험에 사용할 데이터를 수집하였다. 호흡기 질병으로 의심되는 돼지의 혈액을 채취한 후, 바이러스 검사와 혈청 분석으로 PMWS, PRRS, MH에 감염된 돼지와 질병에 걸리지 않은 돼지를 확인하여 분류하였다[2]. 음성 수집은 각 개체로부터 1m의 거리에서 디지털 캠코더(JVC GR-DVL520A, Japan)로 취득하였으며, 수집된 소리는 0.13~2.66초의 길이를 가지며 44,100Hz로 샘플링하였다. 또한, 실제 환경 잡음 데이터 생성을 위해 돈사에서 발생하는 실제 잡음인 돼지 발걸음 소리, 라디오 소리, 문 여는 소리와 인위적인 백색 잡음(SNR 비율: 18, 15, 12, 9, 6, 3, 0dB)으로 합성한 데이터를 추가하였다. 환경 잡음 중 돼지의 발걸음 소리는 1~2마리가 움직일 때의 소리(weak footsteps)와 여러 마리가 움직일 때의 소리(strong footsteps)로 나누어서 수집하였고, 돼지의 안정을 위해 틀어주는 라디오 소리(radio operation), 그리고 관리인이 사료를 주거나 배설물 청소를 할 때 문을 여닫는 소리(door opening)를 수집하였다. 수집한 데이터는 총 710개이며 정상 소리 350개와 호흡기 질병 소리 360개(PMWS: 150, PRRS: 140, MH: 70)이다.

3.2 실험 내용 및 결과 분석

본 논문에서 제안하는 시스템을 검증하기 위해, 돈사에서 수집한 소리 데이터를 이용하여 SEGAN 모델을 학습하였으며(그림 2, 3 참조), 실험은 Ubuntu 16.04, TensorFlow 0.12.1 환경에서 진행하였다. 모델 학습은 백색 잡음(SNR 18, 12, 6, 0)과 환경 잡음(weak footsteps, radio operation, strong footsteps, door opening)을 이용한, 총 8가지 잡음 상황을 추가한 데이터 5,680개를 이용

하여(710개 × 8가지 잡음 상황) 진행하였으며, ground truth는 잡음이 추가되지 않은(noise-free) 깨끗한 데이터를 이용하였다. SEGAN을 학습할 때 설정한 하이퍼 파라미터로는 Generator의 학습률은 0.001, Discriminator의 학습률은 0.0001, batch size는 100, Leaky ReLU(alpha = 0.2), dropout 비율은 50%이며 Xavier 함수를 이용하여 모든 노드를 초기화하였고 전체 학습 횟수는 50,000회로 설정하였다.



(그림 4) 소리 신호 비교 그림(파형, 스펙트로그램)

그림 4는 잡음이 없는 소리 원본(noise-free), 원본 신호에 잡음이 합성된 소리(SNR 0), 잡음이 제어된 소리(enhanced)에 대한 파형(왼쪽)과 스펙트로그램(오른쪽)에 대한 예시이다. 그림 4에서 보듯이 SEGAN을 통해 생성된 잡음이 제어된 소리 데이터가 잡음이 합성된 데이터와 비교하여 확연하게 잡음이 제어된 것을 시각적으로 확인할 수 있었다. 그리고 잡음이 제어된 정도를 정량적인 수치로 확인하기 위하여 실험 성능 지표로 PESQ(Perceptual Evaluation of Speech Quality)와 SSNR(Segmental SNR) 그리고 코사인 유사도를 사용하였다. PESQ는 ITU-T P.862.2에서 권장하는 광대역 버전

을 사용하여 소리 품질을 평가하는 지표로 -0.5~4.5의 값을 가지며, SSNR는 프레임 단위 신호 대 잡음비로 0~∞의 값을 갖는다. 마지막으로 코사인 유사도는 두 벡터 사이의 코사인 거리를 계산하여 역수를 취한 값으로 0~1 사이의 값을 갖는다. 세 지표 모두 값이 클수록 좋은 성능을 의미한다. 본 논문에서 제안하는 시스템의 성능 실험 결과는 표 1과 같다. 세 지표의 수치 값은 각 잡음 상황별 평균값으로 계산하였으며, PESQ, SSNR, 코사인 유사도 3개의 지표 모두 백색 잡음과 실제 환경 잡음에서 잡음이 제어되어 음질이 향상된 것을 정량적으로 확인할 수 있었다.

4. 결론 및 향후 연구

본 논문에서는 소리 취득 시 발생하는 다양한 잡음을 제어하기 위하여 SEGAN 방법론을 도입하였으며, 소리 데이터의 변형 없이 원시 데이터 그 자체를 활용한 end-to-end 방식의 소리 음질 향상 시스템을 제안하였다. 제안된 시스템은 실제 국내 돈사에서 취득한 돼지 소리 데이터들을 이용하여 제안한 모델을 학습하였으며, PESQ, SSNR, 코사인 유사도의 3개의 지표를 통해 음질이 향상된 것을 실험적으로 확인하였다.

향후 연구 과제로는 SEGAN 모델의 일부를 변형하여 데이터 증폭(data augmentation) 과제도 고려해볼 예정이며, 본 연구에서 사용한 돼지 호흡기 질병 데이터 이외에 다양한 산업군의 데이터와 open databases를 이용하여 검증 실험을 수행할 예정이다.

감사의 글

본 연구는 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2018R1D1A3B07044938).

참고문헌

[1] I. Ozer, Z. Ozer, and O. Findik, "Noise Robust Sound Event Classification with Convolutional Neural Network," Neurocomputing, Vol. 272, pp. 505-512, 2018.
 [2] Y. Chung, S. Oh, J. Lee, D. Park, H. Chang, and S. Kim, "Automatic Detection and Recognition of Pig

<표 1> 성능 비교표

Noise conditions	PESQ		SSNR		코사인 유사도	
	Noisy	Enhanced	Noisy	Enhanced	Noisy	Enhanced
백색 잡음	2.2920	2.4137	6.5216	9.2815	0.8838	0.9155
환경 잡음	1.5939	2.1482	6.4849	9.5276	0.8705	0.9159

- Wasting Diseases Using Sound Data in Audio Surveillance,” *Sensors*, Vol. 13, No. 10, pp. 12929–12942, 2013.
- [3] J. Lee, L. Jin, D. Park, Y. Chung, and H. H. Chang, “Acoustic Features for Pig Wasting Disease Detection,” *International Journal of Information Processing and Management*, Vol. 6, No. 1, pp. 37–46, 2015.
- [4] 최용주, 이종욱, 박대희, 정용화, “질감 분석과 CNN을 이용한 잡음에 강인한 돼지 호흡기 질병 식별”, *정보처리학회논문지: 소프트웨어 및 데이터 공학*, Vol. 7, No. 3, pp. 91–98, 2018.
- [5] 이종욱, 최용주, 박대희, 정용화, “CNN 기반의 소리 잡음에 강인한 돼지 호흡기 질병 탐지 및 식별 시스템”, *한국정보기술학회논문지*, Vol. 16, No. 5, pp. 1–13, 2018.
- [6] S. Pascual, A. Bonafonte, and J. Serra, “SEGAN: Speech Enhancement Generative Adversarial Network,” *arXiv preprint arXiv:1703.09452*, 2017.