실시간 복합음향 잡음 저감을 위한 WaveNet-VNNs 기반 ANC 시스템 설계 및 구현

이상민*, 최승렬**, 김태리***, 이예지****, 신애리*****, 김우현*****, 유길상*****

영남대학교*, 인천대학교**, 경북대학교***, 상명대학교****, 고려대학교*****
cross01wmin@naver.com, csy1736@naver.com, telle2409@knu.ac.kr, thisprevision@gmail.com,
shinairi0129@gmail.com, kimwoo320@naver.com, ksyoo@korea.ac.kr

Design and Implementation of a WaveNet-VNNs-Based ANC System for Real-Time Composite Acoustic Noise Reduction

Sangmin Lee*, Seungryeol Choi**, Taeri Kim***,Yaeji Lee****, Aeri Shin*****, Woohyun Kim*****, Gilsang Yoo*****

Yeungnam University*, Incheon National University**, Kyungpook National University***,
Sangmyung University****, Korea University*****

요 약

소음은 스트레스 유발 등 정신적 피해뿐 아니라 장기간 노출 시 심혈관 질환, 청력 손상과 같은 신체적 문제까지 초래할 수 있다. 이러한 소음을 저감하기 위한 능동형 소음 제어(ANC)는 소음의 반대 위상 신호를 생성하여 소음을 상쇄하는 방식으로 주목받고 있다. 그러나 기존 ANC 연구는 주로 시간 지연 및 저감 성능에 초점을 두고 있어 복합 환경에서의 적용에는 한계가 있다. 본 연구에서는 복합음향 중 원하는 소리만을 선택적으로 저감할 수 있도록 WaveNet-VNNs기반 인과적 확장합성곱을 적용하여 신호의 장기 시간 의존성을 학습하고, Volterra Neural Networks를 통해 고차 비선형 특성을 모델링함으로써 역위상 신호를 생성하는 새로운 ANC 시스템을 제안하였다. 실험결과, 음성 분리 및 분류, 소음 저감 과정을 경량화해 계산량을 최소화하면서도 실시간 처리 성능을 확보하였으며, dBA -37.61과 NMSE -35.28의 우수한 저감 성능을 확인하였다. 본 시스템은 실시간 맞춤형 소음 저감 솔루션의 구현을 위한 기술적 기반을 마련하고, ANC 기반 소음 제어 연구의 적용성·범용성·실용성을 높이는 데 기여할 것으로 기대된다.

Ⅰ. 서 론

소음은 현대 사회에서 삶의 질을 저하시키는 주요 환경 요인 중 하나이다. 환경부 보고에 따르면 소음은 스트레스 증가, 수면 방해, 심혈관계 질환 등 인체 건강에 부정적인 영향을 미치는 것으로 알려져 있으며, 이러한 문제를 해결하기 위해 다양한 소음 저감 연구가 진행되어 왔다. 기존 연구에서는 물리적 차폐를 기반으로 하는 수동적 소음 제어 방식이 주류를 이루었으나 인공 지능 및 공학 기술의 발전은 능동형 소음 제어(ANC, Active Noise Control)의 적용 가능성을 확대시키며 물리적 차폐를 보완하거나 대체할 수 있는 새로운 가능성을 열어 주고 있다.

기존 ANC 관련 연구들은 주로 비선형 특성의 반영, 시간 지연(time latency) 문제의 개선, 그리고 전체 소음 저감 성능 향상에 초점을 맞추어 발전해 왔다. 하지만 이러한 시스템들은 모든 입력 음향을 일괄적으로 저감한다는 한계가 있어, 실환경 적용 시 필수 정보가 포함된 신호(예: 음성 안내, 경고음 등)까지 함께 감소시키는 문제가 존재한다. 만약 특정 소리만을 선별적으로 저감하는 방식이 구현된다면, 일상생활뿐 아니라 산업 현장과 같은 복합 환경에서도 필요한 소리는 유지하면서 유해한 소음만을 효과적으로 줄일 수 있어 소음으로 인한 신체적 피해를 획기적으로 감소시킬 수 있다. 따라서 본 연구에서는 복합 음향 환경에서 특정 소리만을 선택적으로 저감할 수 있는 실시간 능동형 소음 제어(ANC) 시스템을 설계하고 구현하였다. 이를 위해 WaveNet-Volterra Neural Networks(WaveNet-VNNs)를 기반으로 한 음성 분리 및 분류 모듈을 설계하고, 해당 모듈을 ANC 프레임워크에 통합함으로써 필수 음향은 유지하면서 원치 않는 소음만을 저감할 수 있는 선택적 소음 제어 기법을 제안하였다. 본 연구는 핵심 기술은 다음과 같다. 첫째, 기존 ANC 기술의 한계로 지적되던 전체 음향 일괄 저감 문제를 해결할 수 있는 선택적 소음 제어 구조를 제시하였다. 둘째, 복합 환경에서도 실시간성이 유지될 수 있도록 모델 경량화와 연산 최적화를 수행하였다. 셋째, 제안된 시스템은 실제 환경에서의 테스트를 통해 실효성을 검증함으로써 차세대 맞춤형 소음 제어 솔루션의 기술적 기반을 제공한다.

Ⅱ. 본론

제안한 선택적 소음 저감 시스템의 전체 구조는 그림 1과 같다. 먼저 입력된 음향 신호는 음성 분리 모듈을 통해 여러 구성 요소로 분리되며, 이후 분류 단계에서 저감 대상 음향(A)과 비저감 대상 음향(B)으로 구분된다. 분류 과정에서 A로 판별된 신호에 대해서만 능동형 소음 제어를 적용함으로써 해당 신호에 대한 반대 위상 신호를 생성하고 제거한다. 이는 단일 채널 입력 환경에서도 소음을 독립적으로 분리하여 실시간으로 역위상 신호를 생성함으로써 효율적인 선택적 소음 저감을 가능하게 한다.

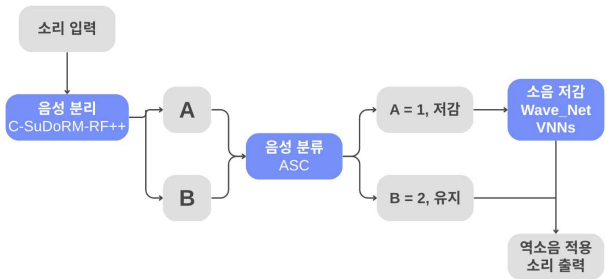


그림 1. 제안한 복합음향 선택 저감용 ANC 모델의 전체 구성도

(1) 데이터 수집 및 전처리

소음 저감 모델 학습은 기존 연구와의 정량적 비교를 위해 타 논문과 동일한 데이터 및 전처리 조건을 적용하였다. 소음 저감 학습에는 DEMAND 및 MN-SNSD 데이터를 벤치마킹 데이터로 활용하였으며, 분리 및 분류 모델 학습에는 서울교통공사 1-8호선 안내방송 음원과 도시 환경 소리(철도

및 항공기 소음)를 수집하였다. 모든 음원은 .wav, 16 kHz, mono 형식으로 통일하였으며, 30초 단위로 랜덤 분할하여 방송(spk1) 및 소음(spk2) 클래스로 저장한 후, 데이터 불균형 문제를 보정하기 위해 인덱스 기반 랜덤 복제를 적용하였다. 두 음원의 혼합은 더 짧은 신호 길이에 맞추어 진행하였으며, 목표 SDR 5 dB가 되도록 두 신호의 에너지 비율을 계산하여 조정하였다. 최종 데이터셋은 train:val:test = 8:1:1의 비율로 분할하였고, 분할 전 원천 데이터의 규모는 철도 49시간 30분, 항공기 24시간 55분이다.

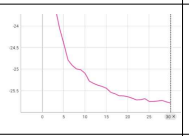
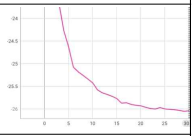
(2) 소음 분리 및 분류

소음 분리 과정에서는 C-SuDoRM-RF++[1]의 인과적 변형 구조를 기반으로 하여 단일 채널 혼합 음향으로부터 안내 방송과 소음을 실시간으로 분리된다. 제안한 분리 모델은 Encoder-U-ConvBlock-Mask Estimator-Decoder 구조로 구성되며, 모든 합성곱 연산을 인과적 1D convolution으로 대체함으로써 스트리밍 기반 실시간 처리가 가능하도록 설계하였다. 분리된 신호는 경량 1D CNN 기반 Audio Segment Classifier(ASC)를 통해 방송과 소음으로 분류되며, 본 분류기는 전역 평균 풀링과 sigmoid 기반 확률 산출 방식에 따라 각 프레임의 음향을 판별한다. ASC는 소음으로 판별된 구간만을 이후의 소음 저감 단계로 전달하는 게이트 역할을 수행함으로써 전체 시스템의 연산 효율이 향상되도록 하였다.

(3) 소음 저감

소음 저감 단계에서는 WaveNet-VNNs[2] 구조를 기반으로 인과적 확장 합성곱을 적용하여 신호의 장기 시간 의존성을 학습하고, Volterra Neural Networks를 통해 고차 비선형 특성을 모델링함으로써 역위상 신호를 생성하였다. 본 모델의 성능과 실시간성 간의 균형을 확인하기 위해 기존 모델의 설정과 비교하며 블록 수(num_stacks과 dilations) 구성에 따른 연산량과 저감 성능을 비교한 결과는 표 1과 같다.

표 1. 블록 수 변화에 따른 dBA/NMSE 성능 및 추론 속도 비교

블록 수	기존 모델	num_stacks = 3 dilations = 8	num_stacks = 2 dilations = 9
train_loss	-		
dBA(dB)	-35.30	-37.61	-37.81
NMSE(dB)	-43.97	-35.28	-41.31
추론속도(ms/sec)	10.60	10.30	9.50

num_stacks=3, dilations=8 설정은 dBA(A-weighted decibel) 기준에서 가장 높은 소음 저감 성능을 나타냈으나, 연산량 증가로 인해 추론 속도가 가장 느리게 측정되었다. 반면, num_stacks=2, dilations=9 설정은 NMSE(Normalized Mean Squared Error)가 개선되었으나 dBA 성능은 다소 저하되었으며, 연산 효율상 속도 개선 역시 제한적으로 나타났다. 그림 2의 결과는 원음 d(n), 역소음 u(n), 그리고 감쇠된 신호 e(n)를 나타내고 있으며, 이는 입력 신호로부터 소음이 효과적으로 감쇠되고 있는 것으로 나타났다.

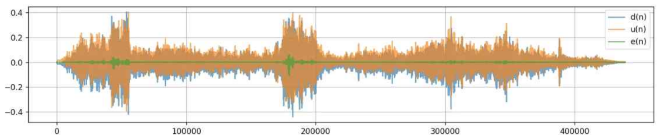


그림 2.제안한 저감 모델의 실행 결과

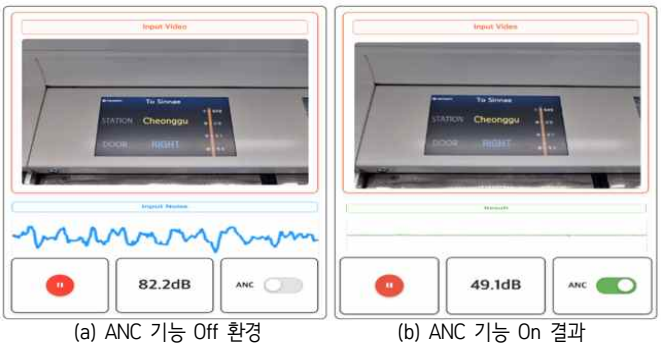
(4) 공동 학습

공동 학습((Joint Training)에서는 분리, 분류, 저감 단계로 구성된 다중 작업(Multi-task) 구조를 하나의 통합 프레임워크로 학습시키기 위해 각 단계별 손실 항목을 최적의 가중치로 조합한 Total Loss 함수를 설계하였으며, 공동 학습에서의 Total Loss 구성 요소 및 가중치 값은 표 2와 같다. 공동 학습을 통해 분리 모듈과 ANC 모듈 간의 과도한 경쟁을 억제하면서 전체 파이프라인의 품질을 균형 있게 향상시키고, 통합 손실 최적화를 기반으로 실제 추론 단계에서 18.9 ms/sec의 지연 시간을 달성하여 실시간 처리 성능[3]을 충족한다.

표 2. 공동 학습에서의 Total Loss 구성 요소 및 가중치

구성 요소	목적	가중치
Final_quality	전체 파이프라인 사용자 체감 품질 향상	0.12
Separation	음성 분리 단계 성능 반영	0.45
ANC_total	소음 저감 단계 성능 반영	0.28
Classifier	분류 인식 단계 성능 반영	0.15

(5) 모델 구현 및 실제 환경 적용



(a) ANC 기능 Off 환경 (b) ANC 기능 On 결과

그림 3. 모델 구현 및 실제 환경 적용 결과

모델의 성능은 서울 지하철 6호선 환경에서 수집한 실제 데이터를 대상으로 검증하였다. 안내 방송음과 배경 소음을 동시에 포함하는 입력 신호를 사용하여 분리 및 저감 성능을 실험한 결과는 그림 3과 같다. 현장 실험 결과, 평균 30 dB 내외의 소음 저감 효과가 확인되었다. 특히, 철도 주행 소음이 효과적으로 억제되면서 안내방송이 또렷하게 전달됨을 확인하였다.

Ⅲ. 결론

본 연구에서는 분리-분류-저감 파이프라인을 기반으로 실시간 환경에서 특정 소음을 선택적으로 처리하는 새로운 ANC 시스템을 제안하였다. 모듈 간 오버헤드를 최소화하고 스트리밍 처리를 전제로 설계함으로써 연산 효율성과 지연 관리 측면에서 유의한 이점을 확보하였다. 실험 결과, 비선형 및 복합 소음 환경에서도 dBA와 NMSE 지표의 개선을 안정적으로 달성하였으며, 교통, 모빌리티, 웨어러블 등 실시간 처리가 필수적인 응용 분야에 즉시 적용 가능한 수준의 일관된 성능을 확인하였다. 향후 연구에서는 임베디드 환경 최적화, 데이터 다양화, 자기 적응 학습(self-adaptive learning) 기법을 통해 시스템의 적용 범위를 확장하고 성능향상을 진행할 계획이다.

참 고 문 헌

[1] Tzinis, E., Wang, Z.-Q., & Smaragdis, P. (2020). SuDoRM-RF: Efficient and compact convolutional neural networks for audio source separation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 1683-1696.

[2] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. arXiv preprint arXiv:1609.03499.

[3] 3GPP. (2024). Universal Mobile Telecommunications System (UMTS); LTE; Enhanced Voice Services (EVS) codec; ANSI C code (fixed-point) (ETSI TS 126 442 V18.0.0, May 2024). European Telecommunications Standards Institute (ETSI).