

주의집중 기반의 합성곱 양방향 게이트 순환 유닛을 이용한 코골이 소리 검출 방식

Snoring sound detection method using attention-based convolutional bidirectional gated recurrent unit

김민수,¹ 이기용,¹ 김형국[†]

(Min-Soo Kim,¹ Gi Yong Lee,¹ and Hyung-Gook Kim^{1†})

¹광운대학교 전자융합공학과

(Received January 5, 2021; revised February 17, 2021; accepted February 24, 2021)

초 록: 본 논문은 수면 무호흡 환자의 중요한 증상 중의 하나인 코골이 사운드 자동 검출 방식을 제안한다. 제안된 방식에서는 수면 중 발생하는 소리 신호를 입력받아 소리 발생 구간을 검출하고, 검출된 소리 구간으로부터 변환된 스펙트로그램을 주의집중 기반의 합성곱 양방향 게이트 순환 유닛 기반의 분류기에 적용하였다. 적용된 주의집중 메커니즘은 합성곱 양방향 게이트 순환 유닛 모델을 확장하여 코골이 소리에 대한 차별적 특징 표현을 학습함으로써 코골이 검출 성능을 향상시켰다. 실험 결과는 제안하는 코골이 검출 방식이 기존 방식보다 약 3.1% ~ 5.5%의 정확도 향상을 보여준다.

핵심용어: 코골이 검출, 소리 발생 구간 검출, 주의집중 메커니즘, 합성곱 양방향 게이트 순환 유닛

ABSTRACT: This paper proposes an automatic method for detecting snore sound, one of the important symptoms of sleep apnea patients. In the proposed method, sound signals generated during sleep are input to detect a sound generation section, and a spectrogram transformed from the detected sound section is applied to a classifier based on a Convolutional Bidirectional Gated Recurrent Unit (CBGRU) with attention mechanism. The applied attention mechanism improved the snoring sound detection performance by extending the CBGRU model to learn discriminative feature representation for the snoring detection. The experimental results show that the proposed snoring detection method improves the accuracy by approximately 3.1% ~ 5.5% than existing method.

Keywords: Snoring detection, Sound activity detection, Attention mechanism, Convolutional bidirectional gated recurrent unit

PACS numbers: 43.60.Bf, 43.80.Qf

1. 서 론

수면 무호흡증의 증상 중 하나인 코골이는 수면 중 상기도의 협착으로 인해 공기의 흐름에 떨림이 생겨 소음이 발생하는 현상으로 수면 방해로 인한 만성 피로, 집중력 장애, 인지기능 저하와 같은 증상들을 발생시킨다. 이와 더불어 코골이가 오래 지속

될 경우에는 고혈압, 심장질환, 뇌졸중 등 다양한 생명 기관에 합병증을 유발할 수 있는 원인이 되기 때문에 코골이는 정확한 검사와 치료를 필요로 하는 위험한 질병이다. 코골이를 검사하기 위한 방법으로는 수면 장애 진단을 위한 표준검사 방법인 수면 다원 검사가 있으나 이는 적지 않은 시간 및 비용을 발생시키며 집 이외의 곳에서 수면을 취해야 한다는

†Corresponding author: Hyung-Gook Kim (hkim@kw.ac.kr)

Department of Electronics Convergence Engineering, Kwangwoon University, 20 Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea

(Tel: 82-2-940-5574, Fax: 82-2-913-5006)



Copyright©2021 The Acoustical Society of Korea. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

불편함을 유발시킨다. 이러한 이유로 음향 분석을 통해 수면 장애 호흡 사운드를 감지할 수 있는 스마트폰 기반 솔루션을 필요로 하고 있다. 기존의 수면 장애 호흡 사운드를 분석하는 방법으로는 수면 무호흡 증후군 환자의 저주파 진동 특성^[1]을 이용하기 위해 단순 코골이 환자와의 스펙트럼 포락선 차이를 분석하여 포먼트 주파수의 표준 편차를 적용하는 방법^[2]이 사용되었으며 스펙트럼으로부터 잡음 성분을 제거하기 위한 다양한 필터링 방법^[3]들이 적용되었다. 최근에는 이러한 수면 장애 호흡 사운드의 주파수 특징을 딥러닝 방식으로부터 추출하는 연구들이 활발히 진행되고 있다. 그중 수면 장애 호흡 사운드의 주파수 및 시계열 특징을 모두 효과적으로 분석할 수 있는 합성곱 순환 신경망(Convolutional Neural Network, CNN)과 순환신경망(Recurrent Neural Network, RNN)^[4,5]이 결합된 CRNN을 통해 코골이 검출 성능을 향상시키는 연구 성과가 제시되고 있다.^[6] 또한 시퀀스 데이터에 대한 관련 특성 표현을 학습하는 데 탁월한 성능을 보이고 있는 주의집중 메커니즘을 순환 신경망에 결합하는 방식^[7]이 적용되어 환경 사운드 분류의 정확도가 탁월하게 향상되는 유망한 결과를 얻고 있다.

이에 본 논문에서는 수면 중 녹음된 오디오 신호로부터 주파수 및 시간적 상관관계를 모두 학습하고 오디오 신호의 중요 맥락 특징을 세밀하게 학습하기 위해 합성곱 양방향 게이트 순환 유닛(Convolutional Bidirectional Gated Recurrent Unit, CBGRU)과 주의집중 메커니즘을 적용한 코골이 검출 방식을 제안한다.

본 논문의 구성은 다음과 같다. II장에서는 제안하는 코골이 검출 방식의 구조에 대해 설명한다. III장에서는 실험을 통한 주의집중 기반의 합성곱 양방향 게이트 순환 유닛을 이용한 코골이 검출 결과를 제시하고 IV장에서는 결론을 맺는다.

II. 코골이 검출 방식의 구조

Fig. 1은 본 논문에서 제안하는 코골이 검출 단계를 나타내는 블록도이다.

먼저, 학습 단계에서는 데이터베이스의 오디오 신호들이 소리 발생 구간 검출을 위해 전처리 과정을

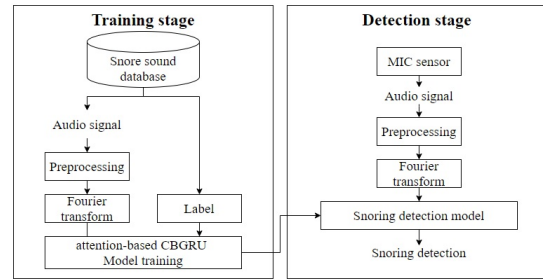


Fig. 1. Block diagram of snoring sound detection.

로 입력된다. 전처리 과정을 통해 검출된 소리 발생 구간의 오디오 신호는 스펙트로그램으로 변환된 후에 주의집중 기반의 CBGRU 모델에 적용되어 학습을 진행한다. 학습이 완료된 후 검출 과정에서는 마이크를 통해 실시간으로 입력된 오디오 신호로부터 전처리 과정을 통해 소리의 발생 구간이 검출되고, 검출된 구간으로부터 변환된 스펙트로그램이 학습 단계에서 생성된 모델에 입력되어 코골이 발생 여부가 결정된다.

2.1 전처리 과정

전처리 과정에서는 녹음된 오디오의 잡음 및 무음 구간과 소리 발생 구간을 구분 짓기 위해 오디오 신호를 2 ms(16 samples)의 프레임 단위로 분할한다. 그리고 전체 프레임의 에너지를 통해 오디오 신호의 에너지와 자기 상관 분산도(Autocorrelation Vector Variance, AVV)를 계산하여 소리 활성 여부를 판단한다. s 번째 오디오 신호의 에너지와 AVV의 계산과정은 다음과 같다.

$$E_j(s) = \sum_{i=16(j-1)+1}^{16j} a^2(i), \quad (1)$$

$$E(s) = \sum_{j=1}^J E_j(s), \quad (2)$$

$$v(s) = \frac{\sum_{j=1}^J \{E_j(s) - (E(s)/J)\}^2}{J}, \quad (3)$$

여기서 i, j, J 는 각각 분할된 프레임의 시간 변수, 번호, 개수이다. 그리고 $a(i), E_j(s), E(s)$ 는 이산 오

디오 신호, j 번째 분할된 프레임과 s 번째 오디오 신호의 에너지를 나타낸다. 또한, Eq. (2)의 16은 프레임의 길이이고 Eq. (3)의 $v(s)$ 는 AVV이며 s 번째 오디오 신호의 에너지 분산 값을 나타낸다.

계산된 입력 오디오 신호의 에너지와 AVV는 임계값과의 비교를 통해 두 값이 모두 임계값보다 작은 경우에는 잡음 구간으로, 그리고 반대의 경우에는 소리 발생 구간으로 판별된다. 이때 초기 임계값은 데이터베이스에서 무작위로 선택된 오디오 신호의 3s 길이의 평균 에너지와 평균 AVV로 설정하였다. 또한 소리 발생 구간 검출의 정확도 향상을 위해 고정된 임계값이 아닌 다양한 잡음 환경 및 잡음 레벨의 상황에 맞춰 변화하는 적응적 임계값 갱신 방법^[8]을 사용하였다.

2.2 스펙트로그램을 통한 소리 특성 비교

오디오 신호로부터 시간 영역에 대한 주파수별 특성을 표현하기 위해 단시간 푸리에 변환을 사용한다. 단시간 푸리에 변환은 1s 길이의 오디오 신호를 16ms(128 samples)의 해밍 창을 통해 분할하고 50%의 중첩 과정을 거친 후, 128-bin의 스펙트로그램으로 변환시킨다.

Fig. 2는 본 논문에서 코골이 검출을 위해 사용된 대표적인 수면 상태에서 발생하는 4가지 소리의 파형과 변환된 스펙트로그램으로 각 소리의 에너지 분포 특징을 명확하게 확인할 수 있다. Fig. 2(a)는 코골이 증상이 없는 사람의 호흡음에 대한 파형과 스펙트로그램을 나타내며, 정상 상태 호흡음의 파형은 낮은 진폭과 일정한 주기를 가진 신호로 스펙트로그램에서는 150Hz~300Hz의 주파수 대역에서 하모닉 성분으로 표현된다. Fig. 2(b)는 코골이 소리의 파형과 스펙트로그램으로 시간 축에서 큰 진폭과 주기성을 갖는 코골이의 스펙트로그램이 50Hz~1200Hz의 주파수 대역에서 높은 주파수에서 에너지가 점점 약해지는 퍼커시브 성분과 약한 하모닉 성분으로 표현된다. Fig. 2(c)는 기관지 천식 환자가 수면 상태일 때 발생하는 천명음의 파형과 스펙트로그램이다. 천명음의 파형은 주기성을 가지며 호흡음보다는 높은 주파수 대역의 신호로 스펙트로그램에서는 200Hz~500Hz의 주파수 대역을 가지고 있다. 또한 700Hz 대

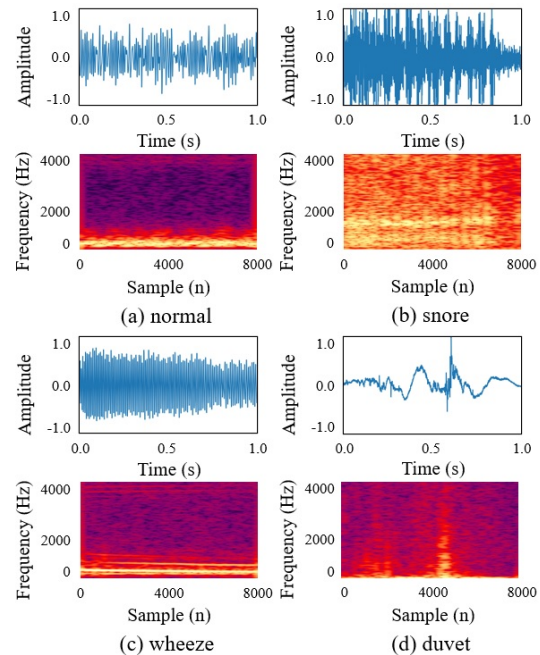


Fig. 2. (Color available online) Waveforms and spectrograms.

역에서 큰 에너지를 보여주며 호흡음보다 하모닉 성분이 더욱 강하게 표현된다. Fig. 2(d)는 수면 상태에서 사람이 뒤척이면서 나는 이불 소리의 파형과 스펙트로그램을 나타낸다. 이불 소리의 파형은 주기성이 없이 불규칙하게 발생하는 소리로서 스펙트로그램에서는 일정한 주파수 대역을 가지지 않는 소수의 타악기 성분으로 표현된다.

2.3 코골이 검출 부

본 논문에서 제안하는 CBGRU와 주의집중 메커니즘을 적용한 코골이 검출 방식의 구조는 Fig. 3(a)에 나타나있다. 오디오 신호로부터 변환된 스펙트로그램은 합성곱 블록에 입력되어 압축된 윤곽 특징으로 추출되며, 추출된 윤곽 특징은 BGRU로 입력되어 시간에 따른 특징 정보의 상관관계를 계산하게 된다. 이때 순환 신경망의 장기 의존성 문제를 해결하기 위해 시간에 따른 특징 정보의 중요도를 다르게 설정하는 주의집중 층이 적용된다. 이후 주의집중 메커니즘이 적용된 CBGRU로부터 출력된 최종 특징값은 소프트맥스 층을 통해 호흡음, 코골이, 천명음, 이불 소리 중의 하나로 예측된다.

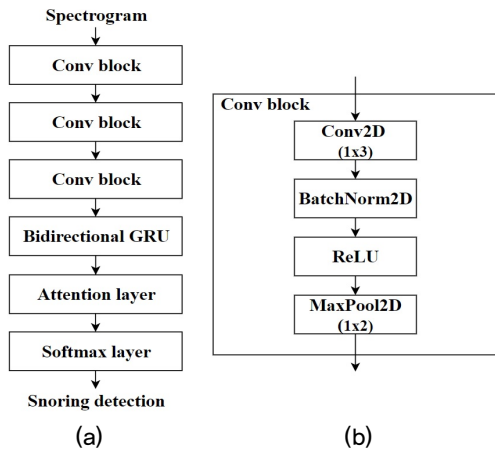


Fig. 3. (a) Structure of proposed attention-based CBGRU, (b) convolutional block.

2.3.1 합성곱 양방향 게이트 순환 유닛

오디오 신호로부터 추출된 스펙트로그램은 시간-주파수 영역에서의 진폭 변화를 표현한 이미지 데이터로 이미지 분류 분야에서 효과적으로 적용되고 있는 CNN을 통해 스펙트로그램의 윤곽특징을 효과적으로 추출할 수 있다. 하지만 CNN은 긴 시간 동안에 발생하는 진폭 변화의 상관관계를 추출할 수는 없기 때문에 스펙트로그램의 시간 상관관계를 효과적으로 분석할 수 있는 방법이 필요하다. 따라서 본 논문에서는 시계열 데이터의 시간적 상관 특성을 반영하기 위해 순환신경망 중 가장 단순한 구조와 빠른 속도의 장점이 있는 GRU를 CNN과 결합한다. 이를 통해 시간-주파수 영역의 윤곽특징과 시간적 상관관계 정보를 모두 효과적으로 반영할 수 있다.

제안하는 구조의 CNN은 Fig. 3과 같이 3개의 합성곱 블록으로 구성되며 각 합성곱 블록은 합성곱 필터(Convolutional filter, Conv), 배치 정규화(Batch Normalization, BN), 선형 유닛 활성화 함수(Rectified Linear Units, ReLU), 맥스 풀링으로 구성된다. 또한, 각 합성곱 블록은 1x3크기의 합성곱 필터가 16, 32, 64개로 구성되어 시간 축방향의 정보를 유지하며 주파수 영역의 다양한 특징을 추출한다. CNN으로부터 추출된 특징은 Fig. 4와 같이 순방향과 역방향으로 나누어진 양방향 GRU로 입력된다. 이때 각 방향의 GRU 셀 개수는 64개이며 이를 통해 시계열 시퀀스에서 과거 및 미래 정보를 최대한 활용하여 강력한 시간적 상관관계 표현이 반영된다.

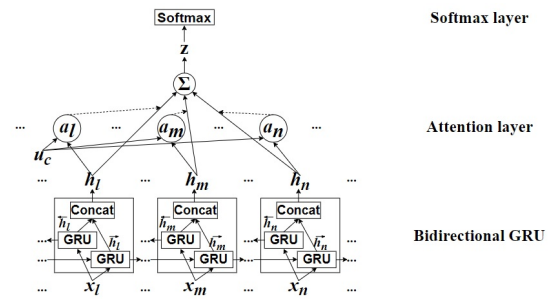


Fig. 4. Structure of attention mechanism.

2.3.2 주의집중 메커니즘

CBGRU로부터 출력된 각 시간 상태의 상관적 특징 벡터는 코골이 검출 분류에 모두 동일하게 기여하지 않는다. 즉, 일부 시간 상태의 특징 벡터는 다른 시간 상태의 특징 벡터보다 더 많은 가중치가 반영된 정보를 제공할 수 있다. 이러한 이유로 코골이 검출에 더 많은 정보의 가중치를 포함하는 시간 상태의 특징 벡터를 위주로 분류기에 적용하기 위해 본 논문에서는 주의집중 메커니즘을 결합한다.

Fig. 4는 본 논문에서 사용한 주의집중 메커니즘의 구조도이다. 주의집중 메커니즘의 연산 과정은 다음과 같은 식으로 표현된다.

$$h_m = \text{concat}[GRU_f(x_m), GRU_b(x_m)], \quad (4)$$

$$u_m = \tanh(W_c h_m + b_c), \quad (5)$$

$$\alpha_m = \frac{\exp(u_m^T u_c)}{\sum_m \exp(u_m^T u_c)}, \quad (6)$$

$$z = \sum_m \alpha_m h_m, \quad (7)$$

여기서 h_m 은 GRU의 입력 값 x_m 을 기반으로 계산된 순방향 GRU 층의 셀 상태와 역방향 GRU 층의 셀 상태를 결합한 m 번째 시퀀스의 주석 벡터이다. 각각의 주석 벡터는 코골이 검출을 위한 시퀀스 간의 맥락 정보를 추출하기 위해 단일 은닉 계층을 갖는 다중퍼셉트론에 입력되고 은닉 표현 u_m 은 Eq. (5)에 의해 계산된다. 다음으로 Eq. (6)과 같이 u_m 과 시퀀스 레벨 문맥 벡터 u_c 의 유사도에 소프트맥스 함수를 적용하여 주의집중 가중치 α_m 이 측정된다. 이후 z 는

Eq. (7)의 식을 통해 시퀀스 주석 벡터의 가중치 합으로 계산되고 문맥 정보를 요약하여 시계열 데이터의 강력한 특징을 반영한다. 문맥 벡터 u_c , 다중퍼셉트론의 매개 변수 W_c 및 b_c 는 무작위로 초기화되고 학습 과정에서 최적 값을 찾아간다. 마지막 단계에서 확률 분포의 예측은 다음과 같은 소프트맥스 층을 사용하여 계산된다.

$$probability = \text{softmax}(W_r z + b_r), \quad (8)$$

여기서 W_r 및 b_r 는 소프트맥스 층의 매개 변수이며 아담(adam) 최적화 방식^[6]을 통해 예측된 레이블과 실제 레이블 간의 교차 엔트로피 오류를 최소화하도록 학습이 진행되어 획득된다.

III. 실험

3.1 실험 데이터

본 논문에서 코골이 검출 실험을 위해 16 m² 크기의 방에서 피험자의 동의하에 수면 중에 발생하는 소리를 수집하였다. 각 소리는 피험자의 머리에서 약 20 cm 떨어진 지점에 위치한 상용 안드로이드 스마트폰을 이용하여 16 bit의 해상도와 8 kHz의 샘플링주파수로 일주일 동안 녹음되었으며 피험자 10명의 성별과 상태에 따른 특성은 반영하지 않았다. 데이터베이스는 1 s 길이의 호흡음 1250개, 코골이 1130개, 천명음 1120개, 이불 소리 1000개로 구성되며 전체 4500개의 소리는 공정한 성능 평가를 위해 각 클래스 별로 동일한 비율로 4-fold 교차 검증 방법에 적용된다.

3.2 실험 결과

주의집중 기반의 CBGRU를 검증하기 위해 다양한 방식의 실험을 진행하여 성능을 비교하였다. 우선 CNN은 합성곱 필터만을 사용하여 특징 값을 추출하여 소리를 분류하는 방식으로 본 실험에서는 3개의 합성곱 블록만을 사용하여 실험을 진행하였다. BGRU는 리셋 게이트와 업데이트 게이트로 구성된 GRU를 사용함으로써 매개 변수가 적고 학습 속도가

빠르다는 장점을 가지고 있는 순환 신경망 방식이며 본 실험에서는 64개의 셀 개수를 적용하여 실험을 진행하였다. 다음으로 CBGRU는 CRNN의 한 형태로 본 실험에서는 앞선 실험을 진행한 CNN과 BGRU의 실험 조건을 그대로 결합하여 실험을 진행하였다. 또한, residual CNN은 CNN에 특정 층의 입력과 출력을 연결하는 잔류 학습 방식을 적용한 신경망으로 4개의 잔류 블록을 사용하여 실험을 진행하였다. 마지막으로, 주의집중 메커니즘의 효과를 검증하기 위해 동일한 BGRU와 CBGRU에 주의집중 메커니즘을 결합하여 실험을 진행하였다. 모든 신경망에 입력된 데이터는 같은 파라미터를 통해 변환된 스펙트로그램을 사용하였고 learning rate는 0.001, batch size는 300으로 설정하여 100 Epoch까지 학습을 진행했다.

Table 1은 본 논문에서 진행한 실험의 결과이다. 가장 기본적인 심층 신경망 구조인 CNN은 92.1%의 정확도를 보이며 순환 신경망인 BGRU는 CNN보다 조금 높은 93.3% 정확도를 보인다. 이는 시계열 데이터에서 시간이 흐름에 따라 변화하는 특징 값을 고려하지 못하는 CNN보다 모든 시간 정보를 고려하여 학습을 진행하는 순환 신경망이 더 우수하다는 것을 나타낸다. 그러나 순환 신경망에 너무 많은 특징 값이 입력되면 기울기 손실문제와 장기의존성 문제를 동반한다. 따라서 CRNN의 한 종류인 CBGRU는 93.8%의 정확도를 보이며 CNN과 BGRU의 단점을 줄이고 장점을 접목시킨 신경망이라는 것을 알 수 있다. 또한 residual CNN은 94.5%의 정확도를 보이며 CNN에 잔류 블록을 결합한 형태로 CNN의 층을 깊게 쌓으면서 연산 속도는 줄이고 성능은 더욱 우수한 것을 확인할 수 있다. 주의집중 기반의 BGRU는 BGRU보다 2.9% 상승한 것에 비해 본 실험에서 제안하는 주의집중 기반의 CBGRU는 CBGRU보다 3.8%

Table 1. Result of snoring detection.

Method	Accuracy
CNN	92.1 %
BGRU	93.3 %
CBGRU	93.8 %
residual CNN	94.5 %
attention-based BGRU	96.2 %
attention-based CBGRU	97.6 %

상승하였다. 이를 통해, 오디오 신호로부터 추출한 스펙트로그램을 주의집중 기반의 CBGRU에 적용한 제안된 방식이 2차원 이미지 데이터 학습에 적합하면서 시간 영역의 특징 값도 고려할 수 있는 가장 우수한 신경망임을 확인할 수 있다.

IV. 결 론

본 논문에서는 주의집중 기반의 합성곱 양방향 게이트 순환 유닛을 이용하여 수면 호흡장애 증상인 코골이 검출 방식을 제안하였다. 실험을 통해 기존의 심층 신경망 방식과 비교하였을 때, 본 논문에서 제안하는 방식이 가장 우수한 성능을 보여주었다. 향후연구로서는 데이터베이스의 규모를 확장하여 더 많은 종류의 수면 호흡 장애를 검출하는 방식에 관해 연구할 계획이다.

감사의 글

본 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원과 2020년도 정부(교육부)의 4단계 BK21 사업의 지원을 받아 수행된 기초연구사업임 (NRF-2018R1D1A1B07041783), (F20YY8101058).

References

1. J. A. Fiz, J. Abad, R. Jane, M. Riera, M. A. Mananas, P. Caminal, D. Rodenstein, and J. Morera, "Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea," *Eur. Respir. J.* **9**, 2365-2370 (1996).
2. J. Sola-Soler, R. Jane, J. A. Fiz, and J. Morera, "Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea," *Proc. IEEE-EMBS*, **3**, 2527-2530 (2003).
3. J. Kim, T. Kim, D. Lee, J. W. Kim, and K. Lee, "Exploiting temporal and nonstationary features in breathing sound analysis for multiple obstructive sleep apnea severity classification," *Biomed. Eng. Online*, **16**, 1-18 (2017).
4. B. Arsenali, J. V. Dijk, O. Ouweltjes, B. D. Brinker, D. Pervernagie, R. Krijn, M. V. Gilst, and S. Overeem, "Recurrent neural network for classification of snoring and non-snoring sound events," *Proc. 40th*

Ann. Int. Conf. IEEE EMBC. 328-331 (2018).

5. S. J. Lim, S. J. Jang, J. Y. Lim, and J. H. Ko, "Classification of snoring sound based on a recurrent neural network," *Expert Syst. Appl.* **123**, 237-245 (2019).
6. F. Vesperini, L. Romeo, E. Principi, A. Monteriù, and S. Squartini, "Convolutional recurrent neural networks and acoustic data augmentation for snore detection," *Neural Approaches to Dynamics of Signal Exchanges*, **151**, 35-46 (2020).
7. Z. Zhang, S. Xu, S. Zhang, T. Qiao, and S. Cao, "Attention based convolutional recurrent neural network for environmental sound classification," *Proc. PRCV*. 261-271 (2019).
8. S. H. Ryu and H. G. Kim, "Audio mixer algorithm for enhancing speech quality of multi-party audio telephony" (in Korea), *J. Acoust. Soc. Kr.* **32**, 541-547 (2013).
9. D. P. Kingma and J. Ba. "Adam: a method for stochastic optimization," *arXiv preprint arXiv: 1412.6980* (2014).

저자 약력

▶ 김 민 수 (Min-Soo Kim)



2021년 2월 : 광운대학교 전자융합공학과 학사
2021년 3월 ~ 현재 : 광운대학교 전자융합공학과 석사과정

▶ 이 기 용 (Gi Yong Lee)



2020년 2월 : 광운대학교 전자융합공학과 학사
2020년 3월 ~ 현재 : 광운대학교 전자융합공학과 석사과정

▶ 김 형 국 (Hyoung-Gook Kim)



1999년~2002년 : 독일 SIEMENS/Cortologic AG 책임연구원
2002년 ~ 2005년 : 독일 베를린 공과대학교 Assistant Professor
2005년 ~ 2007년 : 삼성종합기술원 수석연구원
2007년 3월 ~ 현재 : 광운대학교 전자융합공학과 교수