https://doi.org/10.7735/ksmte.2024.33.1.27

J. Korean Soc. Manuf. Technol. Eng.
ISSN 2508-5107(Online)

골전도 헤드폰 형태로 추출된 골전도 음성 신호의 딥러닝 활용

송희주[a], 유선아[a], 손세강[a], 장웅기[a], 황향희[a], 김현욱[a], 김병희[a,*], 이형석[a,*]

Application of Deep Learning Models for Bone-Conducted Speech Signals Extracted in the Form of Bone Conduction Headphones

Heeju Song[a], Seona Yu[a], Shikang Sun[a], Woong Ki Jang[a], Hyang-Hee Hwang[a], Hyun-Ouk Kim[a], Byeong-Hee Kim[a,*], Hyungseok Lee[a,*]

[a] *Department of Smart Health Science and Technology, Kangwon National University*

ARTICLE INFO

Article history:
Received 12 January 2024
Revised 25 January 2024
Accepted 26 January 2024

Keywords:
Bone conduction
Bone-conducted speech signals
Automatic speech recognition
Deep learning

ABSTRACT

In this study, we used deep learning to align bone-conducted speech signals with air-conducted speech signals, aiming to replace traditional air conduction microphones in voice-based services capturing surrounding sounds. We fabricated headphones, placing bone conduction microphones on the rami (the branches of a bone in the jaw area), in line with traditional bone conduction headphone configurations. Using LSTM, CNN, and CRNN models, we created databases that aligned bone-conducted speech signals with their air-conducted counterparts and tested them with bone-conducted speech signals captured via our custom-made headphones. The CNN model demonstrated superior performance in accurately distinguishing three English words (“apple,” “hello,” and “pass”), including their voiceless pronunciations. In conclusion, our study shows that deep learning models can effectively use bone-conducted speech signals extracted from the rami for automatic speech recognition (ASR), paving the way for future ASR technology that precisely recognizes only the speaker’s voice.

1. 서 론

소리를 듣고 인식하는 방법은 공기전도와 골전도 방식으로 분류할 수 있다[1]. 공기전도는 일반적으로 사람이 소리를 듣는 방식으로, 공기를 타고 외이에 도달한 진동이 중이에 위치한 고막을 진동시키면서 내이의 달팽이관에서 전기에너지로 변환된 후 뇌로 전달되는 과정을 거치게 된다[2]. 골전도의 경우, 진동이 중이의 고막을 거치지 않고 뼈와 근육을 통해서 바로 내이에 전달된다는 점에서 차이를 보인다[3]. 공기전도 마이크는 인식률이 높아 다양한 음성 기반 서비스에 활용되고 있지만[4], 다양한 외부 방해 요소를 함께 인식한다는 한계점이 있다. 예로, 주변의 대화 내용에 의하여 혼선을 야기하거나 큰 소음에 묻혀 사용자의 발언을 정확히 인식하지 못하는 점을 고려해볼 수 있다. 골전도 마이크의 경우 신체에 접촉된 면으로부터 소리 정보를 수집하기 때문에 착용 상태에 따

* These authors equally contributed to this work.
* Corresponding author. Tel.: +82-33-250-6374
E-mail address: kbh@kangwon.ac.kr (Byeong-Hee Kim).
* Corresponding author. Tel.: +82-33-250-6309
E-mail address: ahl@kangwon.ac.kr (Hyungseok Lee).

라 명료도가 달라질 수 있지만[5], 사용자의 목소리만을 녹음할 수 있어 외부의 방해 요소로부터 자유롭다는 이점을 보인다.

현재 시중 판매되고 있는 이어폰은 공기전도 마이크를 탑재하고 있으며, 골전도 헤드폰 역시 스피커만 골전도 기반일 뿐 마이크는 공기전도 방식을 사용하고 있다. 앞서 언급한 공기전도의 마이크가 지닌 한계점을 극복하기 위하여 골전도 음성 신호를 활용한 사례가 존재하기는 하지만, 골전도 음성 신호를 바탕으로 공기전도 음성 신호를 개선하는 방향으로 알고리즘이 개발되고 있어[6] 골전도 음성 신호의 단독 활용은 아직 연구가 부족하다. 또한, 해당 알고리즘은 노이즈를 효과적으로 제거할 수는 있지만, 공기전도 신호와 골전도 신호의 특징 불일치 때문에 잔여 잡음 신호를 발생시켜 오히려 명료도를 저하시킬 수 있다는 문제점이 존재한다[7].

딥러닝 기술의 발전에 따라 골전도 음성 신호의 단독 활용을 시도한 연구가 등장하게 되었는데, Venkata Subbaiah Putta 등에 따르면 조용한 환경에서 인도의 타밀어 단어로 녹음된 공기전도 음성 신호와 골전도 음성 신호를 3가지 딥러닝 모델로 학습시켰더니 다섯 가지 단어를 잘 구별할 수 있었다고 한다[8]. 그러나 해당 연구에서 사용된 MEMS 음향 센서는 외부의 박수 소리에 반응할 정도로 민감하기 때문에 완벽하게 사용자의 목소리만 인식할 수 있는 골전도 마이크라고 보기에는 어려움이 있으며, 큰 기판에 탑재된 센서의 구조는 제품화하기에는 적합하지 않은 형태를 보이고 있다. 따라서 본 연구에서는 시끄러운 환경에서 사용자의 목소리만을 인식하는 골전도 마이크를 선별하고, 골전도 음성 신호를 공기전도 음성 신호에 매칭시켜 구축한 데이터베이스에 실제 골전도 이어폰과 유사한 형태 조건에서 추출된 골전도 음성 신호를 입력할 경우 단어를 잘 구별할 수 있는지 확인하고자 하였다.

자동 음성 인식(ASR, automatic speech recognition)은 다양한 딥러닝 기술을 사용하여 음성 신호를 글자로 변환하는 기술이다. LSTM (long short-term memory) 모델은 순환 신경망(RNN, recurrent neural network)의 한 유형으로, 순차 데이터에서 장거리 의존성을 파악하는 데에 특화되어 있어 음성의 시간적 측면을 모델링하는 데에 적합하다[9]. LSTM은 시간이 지남에 따라서 정보를 기억하고 잊어버리는 과정을 반복하기 때문에 음성 인식 과정 중 문맥을 파악하는 데에 도움이 되기도 한다. 이미지 처리 기술로 널리 알려져 있는 CNN (convolutional neural networks) 모델의 경우 오디오에서의 파형(waveform)과 스펙트로그램(spectrogram)로부터 관련 패턴과 계층 구조 특징을 추출할 수 있다. 스펙트럼 특징과 시간적 특징 모두 추출할 수 있어 자동 음성 인식 기술에서 종종 활용되고 있다[10]. 마지막으로 CRNN (convolutional recurrent neural networks) 모델은 RNN과 CNN의 장점을 결합한 하이브리드 모델이다[11]. 합성곱 층(convolutional layer)을 통합하여 오디오 특징으로부터 고차원적 표현형을 추출할 수 있게 할 뿐만 아니라, 시간 의존성 특징을 캡처하기 위한 순환 층(recurrent layer)을 추출할 수 있도록 하기 때문에 다양한 언어가 공존하는 음성 데이터를 구분하고 텍스트화 하는 데에 특화되어 있다고 할 수 있다. 이처럼 여러 특성을 지닌 세 가지 딥러닝 모델을 본 연구에 활용함으로써 골전도 헤드폰 형태로 추출된 골전도 음성 신호를 딥러닝을 통해 텍스트화 할 수 있는지 검토하고자 하였다(Fig. 1).

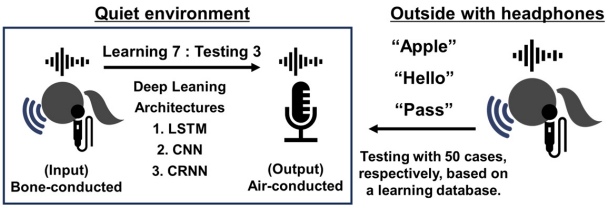


Fig. 1 Concise overview of experimental data processing

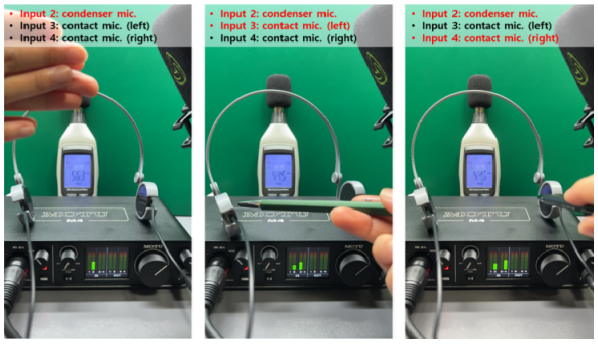


Fig. 2 Assessing the detection statuses of the signal based on specified conditions

2. 실험 방법

2.1 데이터 추출 조건

여러 소음이 존재하는 외부 환경 조건을 구현하기 위하여 스피커(Creative PEBBLE V2; Creative, 싱가포르)를 책상 위에 배치하였으며, 구현된 소음의 크기를 파악하기 위하여 실시간 데시벨 데이터를 소음 측정기(CEM DT95; 찬주테크㈜, 한국)로 확보하였다. 진동하는 파장 데이터만을 추출할 수 있다고 알려진 압전 소자 센서(piezoelectric sensor)가 탑재된 콘택트 마이크(contract microphone) AD-20 (Adeline, 중국)가 외부의 박수 소리를 녹음할 수 없다는 사실을 파악한 후 헤드폰과 같이 왼쪽과 오른쪽이 구별되어 있는 스테레오 구조로 작동할 수 있도록 하였다(Fig. 2). 공기전도 음성 신호의 경우 콘덴서 마이크 LCT440 Pure (LEWITT, 오스트리아)를 입으로부터 약 10 cm 떨어진 거리에 배치한 상태로 확보되었으며, 입력값과 출력값이 될 골전도

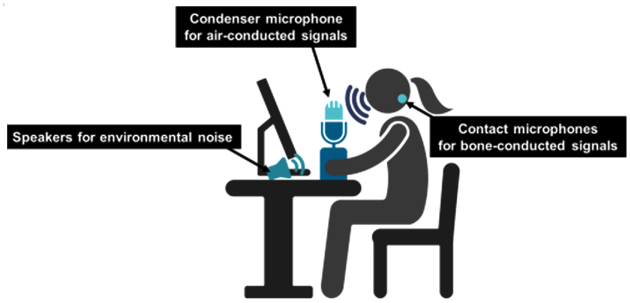


Fig. 3 Experimental environmental condition

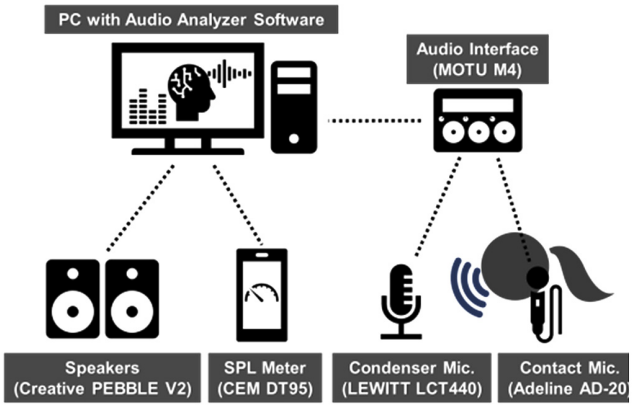


Fig. 4 Experimental apparatus configuration

음성 신호와 공기전도 음성 신호를 동시 녹음할 수 있도록 4채널 오디오 인터페이스(M4; Mark of The Unicorn, 미국)를 사용하였다. 녹음은 미국 Adobe 사의 Audition 프로그램을 이용하여 진행하였으며, 17 dB 소리 증폭과 기본 노이즈 제거 효과를 1회 실시하여 데이터를 정돈하였다. 최종적으로 Fig. 3과 같은 환경 조건에서 Fig. 4와 같은 구조로 실험이 진행되었다.

음성 데이터 확보 과정에는 20대 여성 1명과 20대 남성 1명이 참가하였으며 데이터의 비중은 각 조건 별 50개씩으로 동등하였다.

먼저, 딥러닝 모델이 단어를 구별할 수 있는지 확인하기 위하여 두 음절인 영어 단어 apple과 hello를 조건으로 설정하였다. 그리고 성대가 울리지 않는 무성음은 골전도 음성 신호로 추출할 수 없다고 하여[12], 무성음이 포함된 hello를 비롯하여 pass를 조건으로 추가 선정한 후 녹음 및 인식 가능 여부를 검토하였다.

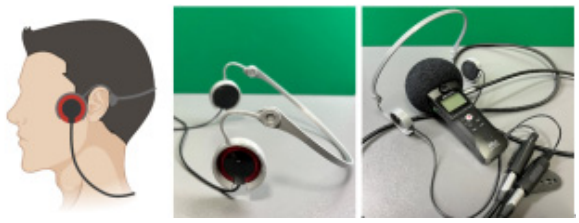


Fig. 5 Visual representation of the designed product concept and realized prototype production outcome. The representation was created with BioRender.com

2.2 헤드폰 제작 방안

후두 앞에서 추출한 골전도 음성 신호가 제일 명확하지만[5], 목을 압박하는 형태는 사용자가 불편함을 느낄 수밖에 없도록 한다. 또한, 기존의 골전도 헤드폰은 귀에 걸린 채 골전도 스피커가 귀 앞턱뼈가지(rami) 쪽에 밀착된 형태를 보이고 있기 때문에 골전도 마이크를 후두 쪽에 배치하는 데에는 어려움이 있다. 따라서 본 연구에서는 골전도 마이크가 기존의 골전도 헤드폰 같은 구조에서 역할을 수행할 수 있도록 골전도 마이크 고정 제품을 설계하였다. FDM 기반의 3D 프린터(Creator pro 2; Flashforge, 중국)에 PLA 필라멘트(Ender 1.75 mm PLA; Creality, 중국)를 적용하여 제품을 제작하였으며, Fig. 5와 같이 골전도 마이크가 잘 고정되었음을 확인되었다. 이후, 실제 헤드폰이 사용되는 조건에서 골전도 음성 신호 데이터를 확보하고자 하였기에 휴대용 오디오 인터페이스(H1n-VP;ZOOM, 일본)에 골전도 마이크를 연결한 후 실외에서 녹음을 진행하였다

2.3 딥러닝 모델

구글의 Colaboratory를 사용하여 딥러닝 모델을 설계 및 구동하였다. 영어 단어 조건 3개에 대해서 공기전도 음성 신호와 골전도 음성 신호를 100개씩 준비하였으며, 코딩을 통해 각 조건마다 70개의 데이터가 무작위로 추출되어 입력값과 출력값을 설정할 수 있도록 하였다. 모든 모델의 학습 횟수(epoch)는 30번으로 설정되었으며, 남은 30개의 데이터로는 학습 결과를 검토하여 검증 손실(validation loss)이 0.617미만, 검증 정확도(validation accuracy)가 97.8% 이상이 되었음을 확인하였다. 골전도 헤드폰 형태로 조건 당 100개씩 추출된 골전도 음성 신호는 무작위로 50개씩 추출되어 구축된 데이터베이스에 입력되었다.

Table 1, 2, 3은 본 연구에서 활용된 LSTM, CNN, CRNN 모델의 구조를 각각 나타낸 것이다. 세 모델 모두 Librosa 라이브러리를 활용하여 wav 파일을 불러온 후, 음성 신호의 단기 전력 스펙트럼을 나타내는 계수인 MFCC[13] 데이터를 추출하여 음성 인식에 활용하였다. 이때, 확보된 음성 데이터의 길이가 0.8~1.2초로 상이하여 최댓값인 1.2초를 기준으로 “0”값을 채워 모든 데이터가 같은 크기를 가질 수 있도록 하였다. 이후, 입력값을 구축되어있는 LSTM 모델과 CNN 모델에 입력될 수 있는 적합한 형태로 변환한 후 학습을 각각 시켰다. CRNN 모델의 경우 CNN 모델이 지닌 특징 추출 기능과 LSTM 모델이 지닌 RNN의 시퀀싱 모델링 기능을 결합한 구조를 지니었으므로, 입력값을 CNN 모델에서 사용되는 형태로 변환한 후 학습에 활용되었다.

Table 1 Audio classification using an LSTM model with MFCC features

Section	Description	Main Functions / Components
Import libraries	Import necessary libraries for file operations, audio processing, ML modeling, etc.	os, random, librosa, numpy, tensorflow, matplotlib, seaborn
MFCC extraction	Function to extract MFCC features from audio files.	librosa.load, librosa.feature.mfcc
Data splitting	Split data into training, validation, and test sets.	random.seed, random.shuffle
Data loading	Mount Google drive, define categories, and load data.	google.colab.drive.mount, os.listdir
Data processing	Process audio files to extract features and split into sets.	extract_mfcc, split_data
Data padding	Pad data sequences for uniform input size.	numpy.pad
Model building	Construct an LSTM model for classification.	Sequential, LSTM, Dense
Model compilation	Compile the model with loss function and optimizer.	model.compile
Model training	Train the model on the training data.	model.fit
Model evaluation	Evaluate the model performance on test data.	model.evaluate
Prediction	Predict and classify on the test set.	model.predict

Table 2 Key functions in building and training a CNN for audio classification. The training, evaluation, and prediction processes are excluded for simplicity in the table

Section	Description	Main Functions / Components
Import libraries	Loads essential Python libraries for processing and model building.	os, random, librosa, numpy, tensorflow, matplotlib.pyplot, sklearn.metrics, seaborn
MFCC extraction	Converts audio files into MFCC features for analysis.	librosa.load, librosa.feature.mfcc
Data splitting	Divides the dataset into training, validation, and test sets.	random.shuffle, List slicing
Data loading and processing	Loads audio files, extracts features, and prepares dataset.	os.listdir, extract_mfcc, split_data
Padding sequences and reshaping	Ensures uniform data shape for CNN input.	np.pad, np.newaxis
Building the CNN model	Constructs the neural network with convolutional and dense layers.	Sequential, Conv2D, MaxPooling2D, Flatten, Dense

Table 3 Comprehensive summary of essential functions and libraries utilized in CRNN model development for audio feature extraction and processing. The training, evaluation, and prediction processes are excluded for simplicity in the table

Section	Description	Main Functions / Components
Libraries import	Import necessary Python libraries and modules.	os, random, librosa, numpy, tensorflow, matplotlib.pyplot, sklearn, seaborn
MFCC extraction	Extracts MFCC features from audio files.	librosa.load, librosa.feature.mfcc
Data splitting	Splits a list of files into training, validation, and test sets based on given proportions.	random.seed, random.shuffle
Data loading	Process and prepare data for the model.	os.listdir, os.path.join
Padding sequences	Pads the MFCC sequences to ensure uniform length.	numpy.pad, numpy.array
Reshaping for CNN input	Reshapes data for compatibility with the CNN layers.	.reshape
Model building (CRNN)	Constructs a model with convolutional (CNN) and recurrent (LSTM) layers.	Sequential, Conv2D, MaxPooling2D, Reshape, LSTM, Dense
Model compilation	Prepares the model for training by setting the loss function, optimizer, and metrics.	.compile

3. 실험 결과

3.1 콘덴서 마이크와 콘택트 마이크

파형은 시간 경과에 따른 신호의 진폭을 나타내어 음파로 인한 기압의 변화를 나타내며[14], 스펙트럼 주파수(spectral frequency)는 오디오 신호에 존재하는 주파수 분포를 나타낸다[15]. 그리고 앞서 언급된 MFFCC는 화자 식별, 음성 인식 등에 활용된다[13]. 이 점을 참고하여 지하철 이용 시 발생하는 소리를 스피커를 통해 재생시켜 실제 소음 환경인 70~80 dB에 노출된 상태[16]에서 공기전도 음성 신호와 골전도 음성 신호를 동시에 확보한 후 파형, 스펙트럼 주파수, MFCC를 Fig. 6과 같이 확인하였다. 영어 단어 apple을 2~3초 간격으로 6번 말한 공기전도 음성 신호의 파형 그래프에

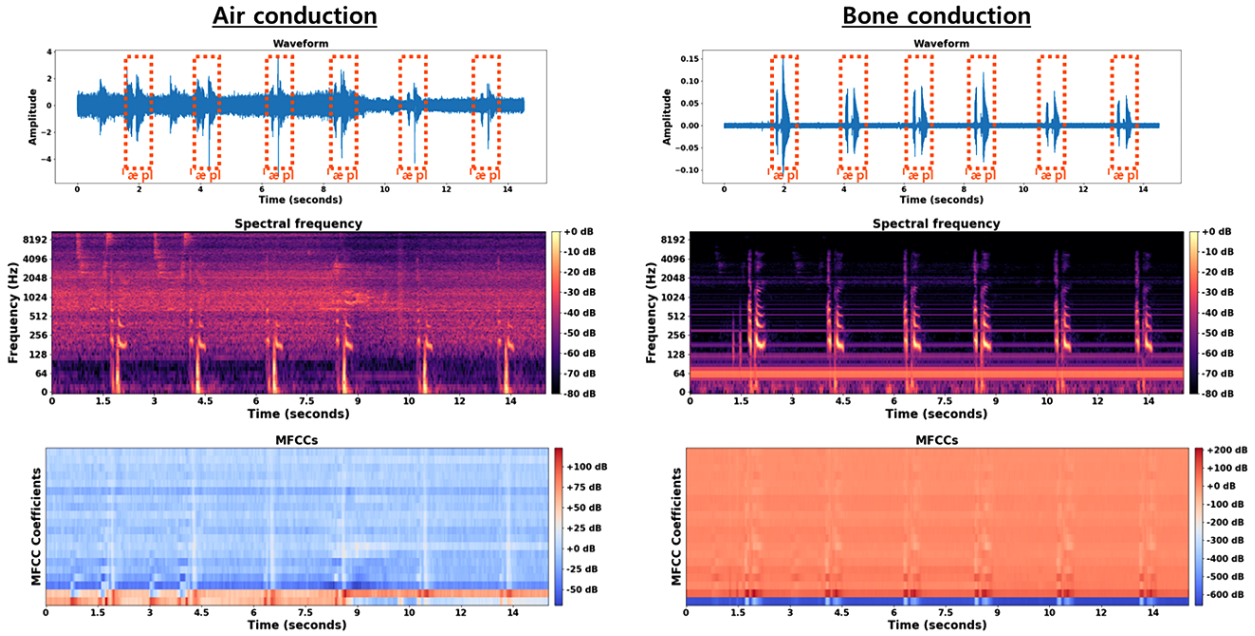


Fig. 6 Differences between the recorded air-conducted and bone-conducted speech signals in a subway noise-exposed environment

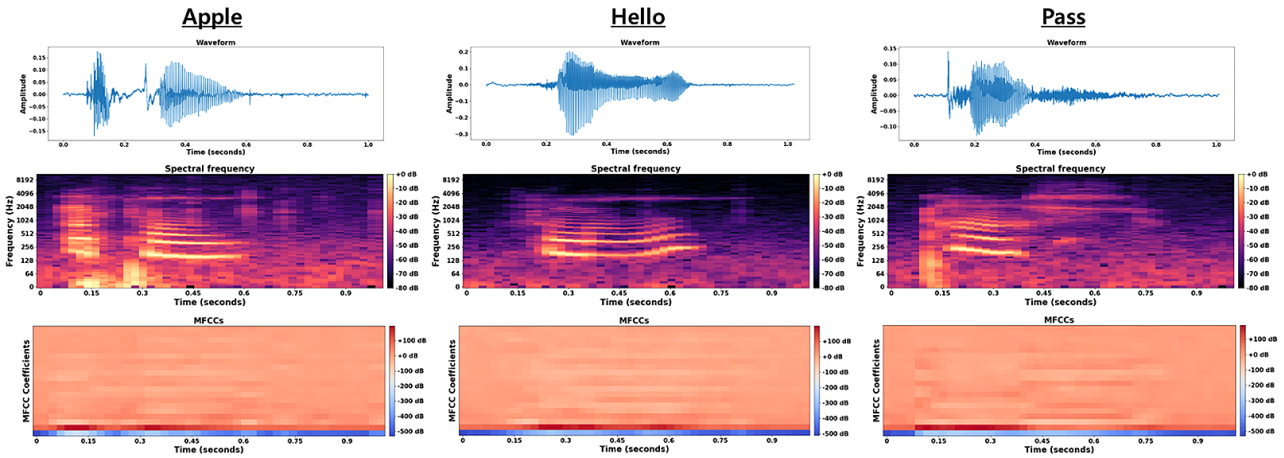


Fig. 7 Acoustic properties of three English words (apple, hello, pass) captured utilizing the portable bone conduction microphones

서는 0~2초 사이와 2~4초 사이에서 지하철 안내 방송이 함께 인식된 모습을 관찰할 수 있다. 또한, 공기전도 신호의 스펙트럼 주파수와 MFCC 결과는 주변의 잡음이 함께 녹음되어 음성 데이터의 명료도가 크게 저하되었음을 보여준다. 반면, 공기전도 신호에서 관찰된 지하철 소음이 골전도 신호에서는 전혀 녹음되지 않은 채 apple의 두 음절만이 명확하게 녹음된 모습을 파형 그래프에서 확인할 수 있다. 실제로, 골전도 신호의 스펙트럼 주파수와 MFCC 그래프에서도 음성 신호의 특징이 매우 뚜렷하게 나타났다. 따라서 골전도 마이크인 콘택트 마이크는 주변의 잡음으로부터 자유로운 채 음성 신호를 추출할 수 있으므로, 음성 신호 확보 과정에 있어 공기전도 방식인 콘덴서 마이크보다 더 우수하다고 판단할 수 있다.

3.2 단어에 따른 발음 변화

골전도 헤드폰 형태로 만든 골전도 마이크 고정 제품을 이용하여 야외에서 apple, hello, pass를 모두 잘 녹음할 수 있는지 검토하였다(Fig. 7). 먼저 apple의 파형에서 ap-ple로 두 음절이 완벽히 나뉘어진 채 녹음이 된 모습을 관찰할 수 있으며, 스펙트럼 주파수에서는 발음 /æ/ 앞에 악센트가 들어가 첫 부분이 또렷하게 나타난 것을 볼 수 있다. 단어 hello에서는 hel-lo로 나누어 살펴볼 수 있는데, 앞의 /hə/가 강조되고 이어지는 /loʊ/가 약한 소리로 길게 늘어지는 특징을 보인다. 무성 성문 반찰음인 /h/는 성대의 울림을 동반하지 않음에도 불구하고[12], 골전도 데이터를 재생시키면 발음 /h/가 제대로 인식할 수 있을 정도로 들렸기 때문에 골전도 마이크인 콘택트 마이크는 무성음을 문제없이 녹음할 수 있다는 사실을

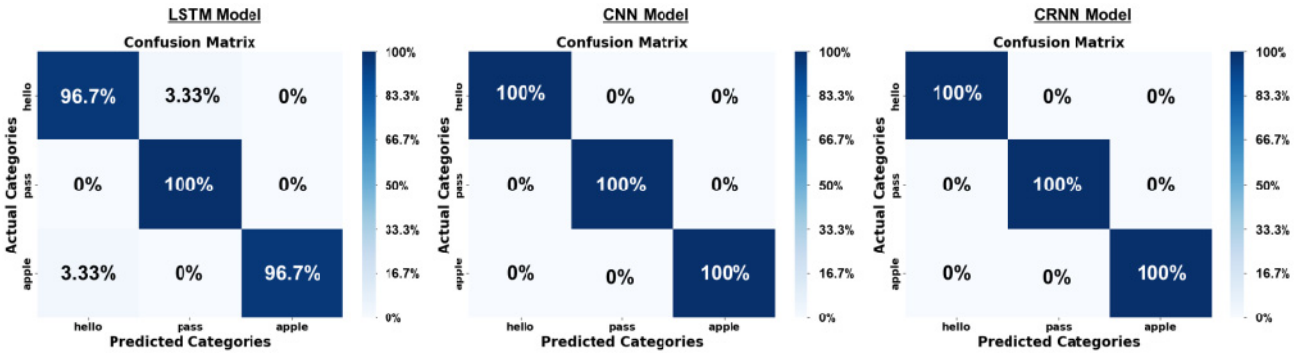


Fig. 8 Learning results of deep learning models using bone-conducted speech signals and air-conducted speech signals recorded in a quiet environment

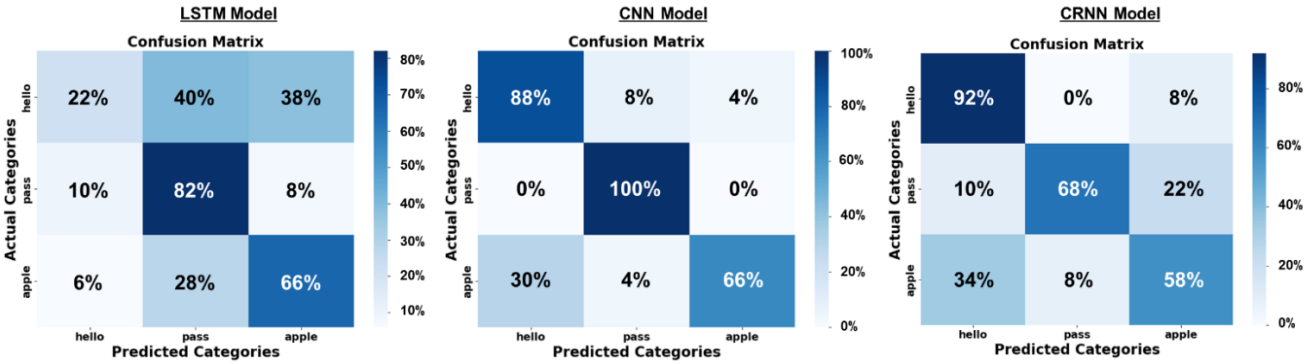


Fig. 9 Prediction results when the bone conducted speech signals recorded with the portable bone conduction microphones were inputted to each deep learning model

파악할 수 있었다. 단어 pass의 경우, 앞의 /p/ 발음이 짧고 강하게 나타났다가 점점 작은 소리로 길게 늘어지는 /s/로 끝맺게 된다. 마찬가지로 성대가 울리지 않는 무성음 /p/와 /s/가 단어 pass의 경우, 앞의 /p/ 발음이 짧고 강하게 나타났다가 점점 작은 소리로 길게 늘어지는 /s/로 끝맺게 된다. 마찬가지로 성대가 울리지 않는 무성음 /p/와 /s/가 제대로 녹음이 되었으며, 짧아서 파형에서는 구별할 수 없었던 hello의 /h/와 달리 pass의 /s/는 길게 발음되어 충분히 육안으로 식별할 수 있었다. 골전도 마이크는 피부와 접촉된 면적이 작아질수록 녹음되는 음성 신호의 상태가 매우 달라질 수 있으므로, 본 연구에서는 왼쪽에서 녹음된 데이터와 오른쪽에서 녹음된 데이터를 각각 단일 채널 데이터로 변환하여 실효성을 입증하고자 하였다.

3.3 딥러닝 모델의 적용 결과

다양한 종류의 딥러닝 모델을 이용하여 조용한 환경에서 동시 녹음된 골전도 음성 신호를 공기전도 음성 신호에 매칭하였다. 학습에 동반된 데이터는 70개로, 30개의 데이터로 모델의 일치 확률을 검사했을 때 LSTM, CNN, CRNN 모델 모두 매우 높은 일치율을 보였다(Fig. 8). 성별이 다른 참여자가 성량의 크기와 어투를 조절해가며 녹음을 진행했음에도 불구하고 완벽한 일치율을 보인 딥러닝 모델은 모두 과적합 상태라고 판별되었다. 딥러닝 모델 특성상 적합 확률이 과하게 높을 경우 실제 사례에 적용할 시 오차가 커진다는 문제점이 있다[17]. 이를 감안하여 구축된 딥러닝 모델이 실제로 잘 활용될 수 있는지 검토하기 위하여, 야외에서 휴대 상태로 추출된 골전도 데이터 50개씩을 학습된 딥러닝 모델에 입력하였다. 이때, 학습 과정과 동일하게 검증 과정 역시 100개의 데이터에서 무작위 추출되었다. 결과적으로 Fig. 9와 같이 LSTM 모델, CRNN 모델, CNN 모델 순으로 인식율이 증가함을 확인할 수 있었는데, 이는 LSTM과 CRNN 모델이 언어와 말투를 비롯해 화자를 구별하는 데에 적합한 정보를 인식하는 데에 특화된 순환 신경망을 내포하고 있어 다양한 매개변수를 인식하기 때문인 것으로 추측된다[18,19]. 실제로 오디오의 패턴만을 파악하여 매칭하는 CNN 모델이 제일 적은 오차범위로 단어를 인식하였고, 이를 범용해서 적용한 CRNN 모델에서는 LSTM 모델보다는 더 나은 결과값을 보였다.

4. 결 론

본 연구에서는 공기전도 마이크 없이 골전도 마이크만을 사용할 경우 자동 음성 인식이 가능한지를 확인하고자 하였다. 현재까지

개발된 자동 음성 인식 데이터베이스는 공기전도 신호를 기반으로 하고[20], 공기전도 음성 신호와 골전도 음성 신호의 특성이 상이하다는 점을 고려하여[7] 기존의 데이터베이스를 활용할 수 방향으로 연구를 진행하게 되었다. 조용한 환경에서 녹음된 골전도 음성신호를 바탕으로 공기전도 음성 신호를 도출해낼 수 있도록 다양한 딥러닝 모델을 활용하였으며, 일상 소음이 함께 녹음되는 공기 전도 마이크와 소리가 작게 녹음되는 골전도 마이크의 특징을 고려하여 별도의 데이터 가공 과정을 거쳐 학습을 진행시켰다. 그리고 구축된 데이터베이스가 실제로 활용될 수 있는지 확인하기 위해 외부 환경에서 확보한 골전도 음성 신호를 입력한 결과, CNN 모델이 가장 이상적임을 확인할 수 있었다. 이러한 사실을 바탕으로 콘택트 마이크에 적용된 피에조 센서를 골전도 헤드폰에 탑재한 후 딥러닝 모델을 활용하게 된다면 외부 간섭 요소로부터 자유롭고 보다 정확한 자동 음성 인식 기술을 구현할 수 있을 것이라고 결론 내릴 수 있다.

후 기

본 논문은 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과임 (2022 RIS-005).

References

[1] De Sousa, K. C., Swanepoel, D. W., Moore, D. R., Myburgh, H. C., Smits, C., 2020, Improving Sensitivity of the Digits-in-noise Test using Antiphasic Stimuli, Ear Hear., 41:2 442-450, https://doi.org/10.1097/AUD.0000000000000775.

[2] Møller, A. R., 2012, Hearing: Anatomy, Physiology, and Disorders of the Auditory system, Plural Publishing, San Diego, CA.

[3] Freeman, S., Sichel, J. Y., Sohmer, H., 2000, Bone Conduction Experiments in Animals-evidence for a Non-osseous Mechanism, Hear. Res., 146:1-2 72-80, https://doi.org/10.1016/s0378-5955(00)00098-8.

[4] Zhang, L., Tan, S., Wang, Z., Ren, Y., Wang, Z., Yang, J., 2020, Viblive: A Continuous liveness Detection for Secure Voice User Interface in IoT Environment, ACSAC '20: Proceedings of the 36th Annual Computer Security Applications Conference, 884-896, https://doi.org/10.1145/3427228.3427281.

[5] McBride, M., Tran, P., Letowski, T., Patrick, R., 2011, The Effect of Bone Conduction Microphone Locations on Speech Intelligibility and Sound Quality, Appl. Ergon., 42:3 495-502, https://doi.org/10.1016/j.apergo.2010.09.004.

[6] Zhou, Y., Chen, Y., Ma, Y., Liu, H., 2020, A Real-time Dual-microphone Speech Enhancement Algorithm Assisted by Bone Conduction Sensor, Sensors, 20:18 5050, https://doi.org/10.3390/s20185050.

[7] Yu, C., Hung, K. H., Wang, S. S., Tsao, Y., Hung, J. W., 2020, Time-domain Multi-modal Bone/Air Conducted Speech Enhancement, IEEE Signal Process. Lett., 27 1035-1039, https://doi.org/10.1109/LSP.2020.3000968.

[8] Putta, V. S., Selwin Mich Priyadharson, A., Sundramurthy, V. P., 2022, Regional Language Speech Recognition from Bone-Conducted Speech Signals through Different Deep Learning Architectures, Comput. Intell. Neurosci., 2022 4473952, https://doi.org/10.1155/2022/4473952.

[9] Sak, H., Senior, A. W., Beaufays, F., 2014, Long Short-term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling, Interspeech, 2014 338-342, https://doi.org/10.21437/Interspeech.2014-80.

[10] Soliman, A., Mohamed, S., Abdelrahman, I. A., 2021, Isolated Word Speech Recognition using Convolutional Neural Network, 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 1-6, https://doi.org/10.1109/ICCCEEE49695.2021.9429684.

[11] Alashban, A. A., Qamhan, M. A., Meftah, A. H., Alotaibi, Y. A., 2022, Spoken Language Identification System Using Convolutional Recurrent Neural Network, Appl. Sci., 12:18 9181, https://doi.org/10.3390/app12189181.

[12] Murakami, Y., Kurita, H., 2022, Robust Phonetic Features in Bone-Conducted Speech Communication, ICIC Express Letters, 16:5 513-520, https://doi.org/10.24507/icicel.16.05.513.

[13] Ayvaz, U., Gürüler, H., Khan, F., Ahmed, N., Whangbo, T., Bobomirzaevich, A., 2022, Automatic Speaker Recognition Using Mel-frequency Cepstral Coefficients through Machine Learning, CMC-Comput. Mat. Contin., 71:3 5511-5521, https://doi.org/10.32604/cmc.2022.023278.

[14] Defossez, A., Synnaeve, G., Adi, Y., 2020, Real Time Speech Enhancement in the Waveform Domain, Interspeech, 2020 3291-3295, https://doi.org/10.21437/Interspeech.2020-2409.

[15] Kabal, P., Ramachandran, R. P., 1986, The Computation of Line Spectral Frequencies Using Chebyshev Polynomials, IEEE Transactions on Acoustics, Speech, and Signal Processing, 34:6 1419-1426, https://doi.org/10.1109/TASSP.1986.1164983.

[16] Lee, D., Kim, G., Han, W., 2017, Analysis of Subway Interior

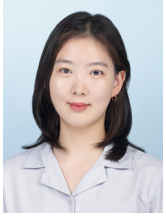
Noise at Peak Commuter Time, J. Audiol. Otol., 21:2 61-65, https://doi.org/10.7874/jao.2017.21.2.61.

[17] Belkin, M., Ma, S., Mandal, S., 2018, To Understand Deep Learning We Need to Understand Kernel Learning, Proceedings of the 35th International Conference on Machine Learning, 80 541-549.

[18] Gers, F. A., Schmidhuber, E., 2001, LSTM Recurrent Networks Learn Simple Context-free and Context-sensitive Languages, IEEE Trans. Neural Netw., 12:6 1333-1340, https://doi.org/10.1109/72.963769.

[19] Bai, Z., Zhang, X. L., 2021, Speaker Recognition Based on Deep Learning: An Overview, Neural Netw., 140 65-99, https://doi.org/10.1016/j.neunet.2021.03.004.

[20] Vincent, E., Watanabe, S., Nugraha, A. A., Barker, J., Marxer, R., 2017, An Analysis of Environment, Microphone and Data Simulation Mismatches in Robust Speech Recognition, Comput. Speech Lang., 46 535-557, https://doi.org/10.1016/j.csl.2016.11.005.



Heeju Song
MS Candidate in the Department of Smart Health Science and Technology, Kangwon National University.
Her research interest is Organ-on-a-chip and 3D Bioprinting.
E-mail: s.canary462@gmail.com

Seona Yu
MS Student in the Department of Smart Health Science and Technology, Kangwon National University.
Her research interest is Nanobio Engineering.
E-mail: 89seona@kangwon.ac.kr

Shikang Sun
Ph. D Student in the Department of Smart Health Science and Technology, Kangwon National University.
His research interest is the Sociology of Sports and Leisure.
E-mail: ssk960911@gmail.com

Woong Ki Jang
Ph.D in the Department of of Smart Health Science and Technology, Kangwon National University.
His research interest is Micro/Nanoscale Surface Texturing Technologies and the Design of Medical Devices and AI Application System Design.
E-mail: wkddndrl@kangwon.ac.kr

Hyang-Hee Hwang
Professor in the Department of Smart Health Science and Technology, Kangwon National University.
Her research interest is the Sociology of Sports and Leisure.
E-mail: phyhee@kangwon.ac.kr

Hyun-Ouk Kim
Professor in the Department of Smart Health Science and Technology, Kangwon National University.
His research interest is Nanobio Engineering.
E-mail: kimhoman@kangwon.ac.kr

Byeong-Hee Kim
Professor in the Department of Smart Health Science and Technology, Kangwon National University.
His research interest is Micro and Nano System Design, the Precision Control of Machine Tools, the Design of Medical Devices, and AI Application System Design.
E-mail: kbh@kangwon.ac.kr

Hyungseok Lee
Professor in the Department of Smart Health Science and Technology, Kangwon National University.
His research interest is 3D Bioprinting, Tissue Engineering, and Wearable Devices.
E-mail: ahl@kangwon.ac.kr