피쳐 퓨전 모듈을 이용한 콘포머 기반의 노인 음성 인식

이민식°, 김지희
동국대학교 인공지능학과
paullee257@dgu.ac.kr, jihie.kim@dgu.edu

Conformer-based Elderly Speech Recognition using Feature Fusion Module

Minsik Lee°, Jihie Kim
Department of Artificial Intelligence, Dongguk University

요 약

자동 음성 인식(Automatic Speech Recognition, ASR)은 컴퓨터가 인간의 음성을 텍스트로 변환하는 기술이다. 자동 음성 인식 시스템은 다양한 응용 분야에서 사용되며, 음성 명령 및 제어, 음성 검색, 텍스트 트랜스크립션, 자동 음성 번역 등 다양한 작업을 목적으로 한다. 자동 음성 인식의 노력에도 불구하고 노인 음성 인식(Elderly Speech Recognition, ESR)에 대한 어려움은 줄어들지 않고 있다. 본 연구는 노인 음성 인식에 콘포머(Conformer)와 피쳐 퓨전 모듈(Features Fusion Module, FFM)기반 노인 음성 인식 모델을 제안한다. 학습, 평가는 VOTE400(Voide Of The Elderly 400 Hours) 데이터셋으로 한다. 본 연구는 그동안 잘 이뤄지지 않았던 콘포머와 퓨전피쳐를 사용해 노인 음성 인식을 위한 딥러닝 모델을 제시하였다는데 큰 의미가 있다. 또한 콘포머 모델보다 높은 수준의 정확도를 보임으로써 노인 음성 인식을 위한 딥러닝 모델 연구에 기여했다.

주제어: 합성곱신경망, 콘포머, 노인 음성 인식, 피쳐 퓨전

1. 서론

자동 음성 인식(Automatic Speech Recognition, ASR) [1] 기술은 현대 디지털 시대에서 음성과 텍스트 간의 원활한 상호 작용을 가능하게 하는 중요한 기술 중 하나로, 다양한 응용 분야에서 큰 역할을 하고 있다. 음성 명령 및 제어, 음성 검색, 텍스트 트랜스크립션, 자동 음성 번역 등의 분야에서 자동 음성 인식 기술은 쓰임새가 무궁무진하다. 그러나, 정확한 노인의 음성을 인식하는 것은 여전히 어려운 과제 중 하나이다.

노인 음성 인식(Elderly Speech Recognition, ESR) [2, 3]에는 어렵고 다양한 도전 과제가 존재한다. 나이로 인한 음성 변화, 발음의 불명확성, 배경 소음 등이 이러한 어려움을 증가시키는 요인 중 일부이다.

본 논문의 기여는 위에서 설명한 노인 음성 인식의 어려움을 극복하고 정확도를 향상시키기 위한 노력의 일환으로 음성 향상(Speech Enhancement, SE) 분야에서 연구가 진행중인 피쳐 퓨전(Features Fusion, FF) [4, 5]과 콘포머(Conformer) [6] 모델을 활용한 자동 음성 인식 모델을 제안하고자 한다.

이 연구에서 우리는 콘포머 아키텍처와 피쳐 퓨전 모듈 모듈(Features Fusion Module, FFM)을 기반으로 한 노인 음성 인식 모델을 소개한다. 피쳐 퓨전 모듈은 합성곱신경망(Convolutional Neural Networks, CNN), 깊이별 합성곱(Depthwise Convolution, DC), 높이별 합성곱(Pointwise Convolution, PC), 정류 선형 유닛(Rectified Linear Unit, ReLU) 등을 결합하여 개발했다. 보다 정확한 노인 음성 인식을 위해, 본 실험에서는 VOTE400 [7] 데이터셋을 활용하여 모델을 훈련하고 평가 하였다. 이 데이터셋은 노인 음성 인식 분야의 연구를 위해 전국에서 수집된 데이터로, 한국어 자연어 처리 기술과 콘포머 모델의 하이퍼파라미터를 조정함으로써 최적의 성능을 달성하였다. 이번 연구는 피쳐 퓨전과 콘포머 아키텍처를 결합하여 한국어 노인 음성 인식을 위한 딥러닝 모델을 제안한 첫 시도 중 하나이다. 실험 결과, 모델은 29.55%에서 30.67%의 정확도를 보여주었으며, 이러한 성과는 앞으로 노인 음성 인식 딥러닝 모델의 개발과 적용에 있어서 큰 잠재력을 지닌 것으로 보여준다. 이 연구는 노인 음성 인식 분야에서 딥러닝 모델의 새로운 가능성을 제시하며, 노인들의 음성 인식 기술 개선에 기여할 것으로 기대된다. 이어지는 장들에서는 연구 방법론, 실험 과정, 결과 및 결론에 대해 더 자세히 다룰 것이다.

2. 음성 인식

자동 음성 인식 기술의 초기 연구는 음향 모델, 언어 모델 및 발음 모델을 기반으로 하는 전통적인 음성 인식 시스템에 중점을 두었다. 이러한 시스템은 히든마르코프모델(Hidden Markov Models, HMM)과 가우시안혼합모델(Gaussian Mixture Models, GMM)를 사용하여 음성을 모델링하고, 주로 통계적인 접근 방식을 사용했다. 그러나 전통적인 모델은 복잡한 언어 현상을 다루기에는 한계가 있다. 딥러닝의 발전으로 자동 음성 인식 분야에서 혁신적인 변화가 일어났었으며 딥러닝을 활용한 자동 음성 인식 모델은 합성곱 신경망, 순환 신경망(Recurrent Neural Networks, RNN), 장단기 신경망(Long Short-Term Memory, LSTM), 게이트 순환 유닛(Gated Recurrent

Unit, GRU)와 같은 신경망 아키텍처를 사용하여 음성을 모델링한다. 이러한 딥러닝 모델은 대량의 데이터에서 학습하고, 더 정확한 음성 인식 결과를 제공하며 다양한 언어 및 환경에서 적용되어왔다.

엔드-투-엔드(End-to-End) [8] 자동 음성 인식 접근 방식은 전통적인 자동 음성 인식 시스템과 달리 중간 단계의 음성 특징 추출(feature extraction) 및 발음 모델링 과정을 제거하고, 하나의 딥러닝 모델로 직접 음성을 텍스트로 변환한다. 이러한 방식은 모델의 간소화와 성능 향상을 가져왔다.

최근에는 트랜스포머(Transformer) [9], 콘포머와 같은 어텐션(Attention Mechanism) [10] 아키텍처가 사용되고 있다. 이러한 관련 연구들은 자동 음성 인식 기술의 진보와 다양한 응용 분야에 대한 적용 가능성을 확장시키고 있으며, 향후 더 나은 음성 인식 기술을 개발하는 데 중요한 역할을 해오고 있다.

3. 노인 음성 인식

노인 음성 특성 분석은 노인 음성 인식 연구의 초기 단계에서는 노인의 음성 특성과 변화를 분석하는 데 중점이 두어졌다. [11]이러한 연구는 노화로 인한 음성 변화, 발음 불명확성, 음성 속도 및 음성 강도 등 [12, 13]과 같은 노인 음성의 특징을 이해하는 데 기여하며 노 음성 인식에서 노인의 음성을 더 잘 이해하고 인식하기 위한 특수한 음성 인식 모델의 개발이 진행되고 있다. 이러한 모델은 노인 음성 특성을 고려하여 설계되며, 딥러닝 모델과 전통적인 음성 인식 기술을 결합하여 정확도를 향상시키고 있다. 노인 음성 인식 연구를 위해 노인 음성 데이터셋이 수집되고 공유되고 있고, 이러한 데이터셋은 노인 음성의 다양한 특성을 포함하고 있으며 연구자들은 이를 활용하여 모델을 훈련하고 평가한다.

관련 연구들은 노인들의 음성 인식 기술을 개선하고 노인들의 일상 생활을 더 나아지게 만드는 데 기여하며, 노인 음성 인식 분야에서의 미래 연구 방향과 혁신을 지속적으로 모색하고 있다.

4. 실험

4.1 데이터셋

VOTE400 데이터셋은 65세 이상의 노인 분들의 음성 녹음 데이터를 기반으로 구축한 데이터셋이다. 이 데이터셋은 정확한 음성 인식 모델을 개발하기 위해 수집되었으며, 마인즈랩과 ETRI에서 구축하여 고령자와 로봇 간 원활한 음성 교류를 가능하게 하는 음성 인식 시스템을 개발하기 위하여 수집한 대용량 한국어 노인 음성 데이터이다. 발화자는 다양한 지역별로 분포되어 있으며, 일상생활에서 흔히 사용하는 문장들을 읽어 녹음하였다. 총 300시간으로 구성되어 있다. 음성 데이터는 16 bit 단일채널 PCM, 11,050 Hz 44,100 Hz 등 다양한 Sampling rate로 저장되어 있다. 대화체 및 낭독체 음성 데이터셋으로 구성되어 있는데, 대화체 음성 데이터는 두 사람의 연속적인 대화를 녹음한 음성이며, 낭독체 음성 데이터는 더욱 고품질의 선별 문장에 해당하는 음성이 녹음되었다. 본 실험을 위해 낭독체 음성 데이터를 활용하였다. 데이터셋에는 총 7,832개의 고유한 문장이 포함되어 있고, 데이터는 EUC-KR로 인코딩된 전사 텍스트 파일과 함께 제공된다. 총 104명이다. 낭독체 음성 데이터셋의 100시간에 대해 실험하였으며, 전체 데이터 111,814개 가운데 학습 데이터로 와 검증용 데이터로 89,451개, 그리고 테스트 데이터로 22,363개를 랜덤하게 나누어 제안된 평가를 수행하였다. 수집된 데이터셋은 음성 인식 모델의 평가 및 개발에 활용될 수 있으며, 노인 분들의 음성을 기반으로 한 언어 모델의 성능 향상을 위한 중요한 자원이다.

표 1. Training and Test data Split

	학습 데이터	평가 데이터	총 합계
데이터 수	89,451	22,363	111,814

표1데이터셋의 111,814개 샘플을 나눌 때, 전체 데이터셋의 80%를 학습(Train) 데이터셋으로 할당하고 나머지 20%를 평가(Test) 데이터셋으로 할당하는 방법을 사용했다. 이에 따라 학습 데이터셋은 약 89,451개의 샘플로 구성되며, 평가 데이터셋은 약 22,363개의 샘플로 구성된다.

4.2 실험환경

우리는 OPENSPEECH Toolkit [14]을 이용하여 실험을 진행했으며, Pytorch 모델을 상속받아 FFM를 설계해 실험을 진행하였다. 데이터는 80차원 스펙토그램으로 처리하여 실험을 진행했다. 인코더(Encoder)네트워크는 총 12개의 콘포머 레이어(Layer)를 포함하며, 각 잔차 블록(residual block)에는 정규화(Normalization)를 위해 드롭 아웃(dropout) [15]을 0.1로 설정했다. 디코더(Decoder)는 1개의 LSTM 레이어, CTC(Connectionist Temporal Classification)Loss [16], Adam(Adaptive moment estimation) [17] 옵티마이저(Optimizer)($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon =$ 1e-8)로 설정 했다. 평가 지표(Metrics)로는 문자열에서 글자별 정답률(Character Error Rate, CER)로 지정한다. 인식된 문자열(한 글자)과 정답 문자열(한 글자) 사이의 문자 오류 비율을 나타내는 지표이다. 또한 에포크(Epoch)를 50으로 실험한다. 배치 사이즈(Batch Size)는 64로 설정하면서 평가 데이터 사용비율(Validation Split)을 0.2로 모델의 성능을 확인하였다. 실행시 성능변화 폭을 최소화하기 위해 시드값(Seed)을 고정하였다. 또한 실험을 위해서 본 논문은 RAM 187GB, Intel(R) Xeon(R) Gold 6226 CPU @ 2.70GHz, 그리고 NVIDIA 2080TI GPU

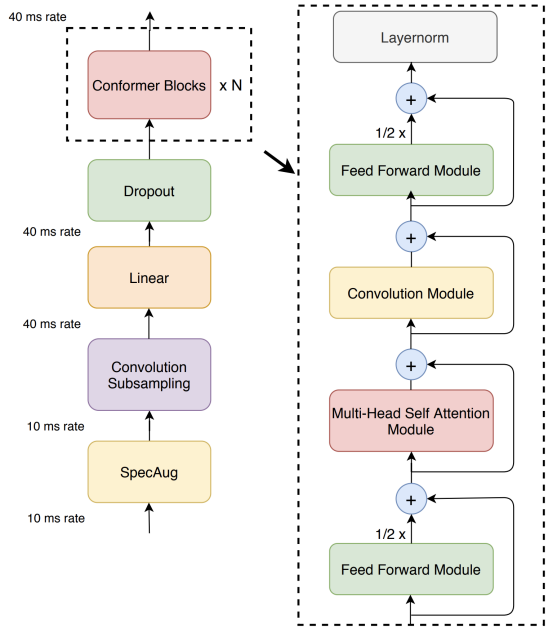

그림 1. Conformer

에 해당하는 컴퓨터 자원을 사용했다.

4.3 모델

4.3.1 Conformer

그림1과 같이 콘포머는 음성 및 텍스트 처리에 사용되는 딥러닝 아키텍처로, 트랜스포머 기반의 모델이다. 입력 데이터를 효과적으로 추상화하고 관계를 모델링하기 위해 셀프 어텐션, 멀티 헤드 어텐션, 컨볼루션 레이어 등을 사용한다. 주로 음성 인식 및 자연어 처리에 적용되며, 시간 및 공간적인 패턴을 모두 고려하여 높은 성능을 발휘해 왔다.

4.3.2 Feature Fusion Module

그림2과 같이 데이터 소스 및 특징을 융합하는 FFM의 강점은 다방면에서 빛을 발한다. 먼저, 다양한 데이터 소스 및 특징을 통합함으로써 정보의 다차원적 활용이 가능하다. 예를 들어, 노인 음성 인식 분야에서 발음의 부정확성, 노이즈, 음성의 변조와 같은 다양한 어려움에 직면할 수 있다. 특히 늘어지는 음성, 부정확한 발음이 많은, 즉 노이즈를 포함하는 부분에서 노인 음성 인식에서 다차원적 가중치 학습은 특히 좋은 성능을 발휘할 것이다. 시간축에 따라서 음성이 분포하게 되는데 이것을 각각의 합성곱 연산으로 특징을 잡아내어 학습하면 길게 늘어진 음성일지라도 높은 인식률을 얻을 수 있다고 가정하였다. 또한 이러한 학습 방법은 노이즈가 많은 음성 특징을 잡아내는 대도 효과적일 것이라 예상 되었다. 합성곱 연산의 강력한 점은 각각의 필터를 통해 특징을 추출하는데 특화 되어 있기 때문이다. 노이즈가 많은 부분의 음성을 합성곱연산을 통해 노이즈를 완화하고 인식률을 높이는데 도움이 되었을 것이라 예상했다. 이렇게 다양한 측면을 학습한 특징들을 함께 활용하면 모델은 보다 강력한 음성 인식과 이해를 구현할 수 있다. 게다가, 피쳐 퓨전 모듈은 서로 다른 합성곱을 활용하여 다양한 입력 유형을 효율적으로 다룰 수 있는 능력을 제공한다. 이는 각 입력 유형에 맞춤화된 가중치를 조절하고 적용할 수 있으며 서로 다른 방식으로 학습된 특징을 결합한다. 이는 모델의 안정성을 높이고 성능을 향상시킬 수 있다. 피쳐를 학습하면서 각각의 합성곱을 통해 더 깊은 특징을 추출하게 된다. 최종적으로 두 가지 유형의 정보를 결합, 학습된 피쳐와 원본 피쳐를 요소별곱(Element-wise product)으로 통합하여 모델은 더 다양하고 풍부한 정보를 활용할 수 있다. 모델에 들어가는 입력은 다음과 같다. R은 음성 인식 모델, S는 원본, F는 모듈을 거쳐서나온 입력, 요소별곱 한다.

$$\hat{y} = R(\underset{speech}{S} \otimes \underset{Fusion}{F}) \tag{1}$$

요약하면, FFM에서 나온 융합된 피쳐가 다양한 레이어에서 학습되며 앞서 언급된 일부 문제를 해결하는데 도움을 주고 효율적으로 모델을 개발할 수 있는 기술이다. 노인 음성 인식의 학습 과정에서 어려운 노인 음성 인식을하는데 도움이 된다.

모델의 음성 인식 학습을 위해 CTC Loss를 사용한다. CTC는 입력되는 하나의 음성 프레임에 하나의 문자가 나오도록 모델링하는 방식이다. 음성의 특성상 중복된 문자가 발생하게 되는데 CTC는 Blank를 사용하여 중복 문자를 없애 이를 해결한 방법이다. 이러한 CTC는 forward와 backward 계산시 최적화되며 학습 시 0번째 Time step 에서 n번째 Time step 까지 모든 가능한 경로의 확률의 합을 나타내며 확률의 합을 증가시키는 방식으로 학습하게 된다.

$$Loss_{\mathcal{L}(S)} = -ln\Pi_{(}y, \hat{y} \in s)P(z|x) \tag{2}$$

4.4 결과

4.4.1 글자 인식 결과

표2 는 VOTE400 테스트 데이터에 대한 우리 모델의 결과를 콘포머 모델과 비교한다. 모든 평가 결과는 3자리에서 반올림했다. 평가 / 테스트에서 기존의 콘포머보다 성능이 뛰어났다. 우리 모델은 기존 콘포머 모델이 실험한 것들 중 가장 낮은 글자 오류율을 기록한것보다 낮은 글자 오류율을 달성했다. 이는 콘포머에 피쳐 퓨전 모듈을 결합한것이 효율적이라는 노인 음성 인식을 결과를 더욱 명확하게 보여준다. 실험 결과에서 보여지듯 29.55%에서 30.67% 의 정확도를 달성하였다. 노인의 음성 특성에 따라 접근한 피쳐 퓨전 모듈 방식을 통해 모델의 성능이 향상 되었다. 노인의 일상 환경과 상황을 고려한 데이터셋을

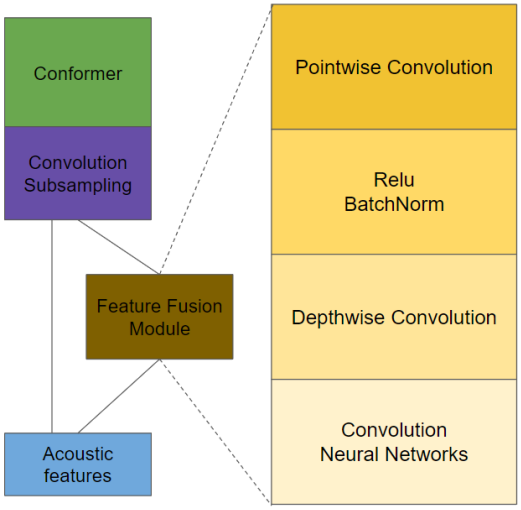

그림 2. Feature Fusion Conformer

표 2. Character Error Rate

모델	평가 데이터	테스트 데이터
Conformer	32.23%	34.35%
FFM-Conformer	29.55%	30.67%

통해 학습과 평가가 되었음으로, 모델이 다양한 환경에서도 안정적으로 동작할 것이라 예상된다.

4.5 Ablation Study

4.5.1 커널 사이즈

표3 에서는 깊이별 컨볼루션에서 커널 크기의 효과를 연구하기 위해 우리는 각기다른 크기의 10, 12, 15에서 커널 크기를 실험 했다. 커널의 크기가 커지면서 성능이 향상되었다.

4.5.2 어텐션 헤드

표4 에서는 어텐션에서는 각 어텐션 헤드가 입력의 서로 다른 부분에 집중하는 방법을 학습하여 단순 가중 평균을 뛰어넘는 예측을 개선할 수 있었다. 우리는 레이어에서 어텐션 헤드 수를 4에서 8까지 변경하는 효과를 연구하기 위한 실험을 수행했다. 표시된 것처럼 관심 헤드를 8개까지 늘리렸을 경우 정확도가 향상된다는 것을 알 수 있었다.

5. 결론

이 연구는 노인 음성 인식 분야에서 새로운 기술과 모델을 개발하고 노인들의 음성을 더 잘 인식하는 방법을 탐구하였다. 노인 음성 인식은 그들의 생활 품질을 향상시키고 의료 및 응용 분야 등 다양한 분야에서 혁신적인 솔루션을 제공할 수 있는 중요한 주제다.

표 3. kernel size

kernel size	평가 데이터	테스트 데이터
11	37.27%	38.38%
13	36.55%	37.67%
15	34.75%	35.57%

표 4. Conformer Attention Heads

Attention Heads	평가 데이터	테스트 데이터
4	35.67%	37.54%
6	34.77%	36.57%
8	33.55%	35.67%

우리의 연구에서는 콘포머와 피쳐 퓨전 모듈을 결합한 모델을 제안하였다. 이 모델은 다양한 딥러닝 기술을 활용하여 노인 음성 인식을 더 정확하게 할 수 있도록 설계하였다. 실험 결과, 우리의 모델은 기존 모델을 뛰어넘은 정확도를 보이며, 특히 노인 음성 인식에 대한 몇가지 문제점에 대한 해결책으로 합성곱 신경망에 대한 깊은 접근을 통해 더 나은 성능을 보였다는 것이 핵심이다.

또한, 이 연구에서는 노인 음성 인식 기술의 다양한 응용 분야에 대한 가능성을 탐구하였다. 이러한 응용 분야는 노인들의 건강 모니터링, 의료 진단, 스마트 홈 시스템, 음성 기반 상담, 커뮤니케이션 도구 등을 포함하는 모든 부분에 적용될 수 있는 가능성이 있음을 알게 되었다. 피쳐를 다양한 방법에서 학습해 노인 음성 인식 실험을 통해 입증하였다. 노인 음성 인식 분야에서의 연구는 노인들의 삶의 질을 향상시키고 그들의 의료 및 일상 생활에 도움을 제공할 수 있는 가능성을 제시하는 데 중요한 역할을 한다. 더 나아가, 우리의 연구는 음성 인식 기술의 미래 연구 방향과 혁신을 지속적으로 모색하며, 노인들의 일상을 더 나아지게 만드는 데 기여할 것이다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업(IITP-2023-2020-0-01789)(50%)과 인공지능융합혁신인재양성사업(IITP-2023-RS-2023-00254592)(50%)의 연구결과로 수행되었음

본 결과물은 ㈜마인즈랩과 한국전자통신연구원이 연구 과제 수행을 통해 구축 공개한 고령자 음성데이터 VOTE400 데이터셋을 사용함. 해당 연구 과제는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발 사업의 일환으로 수행하였음. [2017-0-00162, 고령 사회에 대응하기 위한 실환

경 휴먼케어 로봇 기술 개발]

참고문헌

[1] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach, "An overview of noise-robust automatic speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 4, pp. 745–777, 2014.

[2] J. Yu, X. Xie, S. Liu, S. Hu, M. W. Lam, X. Wu, K. H. Wong, X. Liu, and H. Meng, "Development of the cuhk dysarthric speech recognition system for the ua speech corpus." *Interspeech*, pp. 2938–2942, 2018.

[3] 문정현, 강준서, 김기웅, 배종빈, 이현준, and 임창원, "치매 환자를 포함한 한국 노인 음성 데이터 딥러닝 기반 음성인식," *응용통계연구*, Vol. 36, No. 1, pp. 33–48, 2023.

[4] Y. Hu, N. Hou, C. Chen, and E. S. Chng, "Interactive feature fusion for end-to-end noise-robust speech recognition," 2022.

[5] Z.-H. Lai, T.-H. Zhang, Q. Liu, X. Qian, L.-F. Wei, S.-L. Chen, F. Chen, and X.-C. Yin, "Interformer: Interactive local and global features fusion for automatic speech recognition," 2023.

[6] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang, "Conformer: Convolution-augmented transformer for speech recognition," 2020.

[7] M. Jang, S. Seo, D. Kim, J. Lee, J. Kim, and J.-H. Ahn, "Vote400(voide of the elderly 400 hours): A speech dataset to study voice interface for elderly-care," *ArXiv*, Vol. abs/2101.11469, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:231718957

[8] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," *International conference on machine learning*, pp. 1764–1772, 2014.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017. [Online]. Available: https://arxiv.org/pdf/1706.03762.pdf

[10] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, Vol. 452, pp. 48–62, 2021.

[11] 박지웅, 이승준, and 권순일, "노인음성인식을 위한 전처리에 관한 연구," *한국정보처리학회 학술대회논문집*, Vol. 20, No. 2, pp. 1646–1648, 2013.

[12] 박현, 신혜정, and 손명동, "노인들의 언어 문제와 언어 재활 인식에 관한 기초 조사," *언어치료연구*, Vol. 21, No. 4, pp. 227–247, 2012.

[13] 허명진 and 신명선, "노인 음성의 음향학적 특성," *언어치료연구*, Vol. 19, No. 2, pp. 41–51, 2010.

[14] S. Kim, S. Bae, and C. Won, "Kospeech: Open-source toolkit for end-to-end korean speech recognition," *arXiv preprint arXiv:2009.03092*, 2020.

[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, Vol. 15, No. 1, pp. 1929–1958, 2014.

[16] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," *Proceedings of the 23rd international conference on Machine learning*, pp. 369–376, 2006.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.