한국음향학회지 제41권 제6호 pp. 655~661 (2022)
The Journal of the Acoustical Society of Korea Vol.41, No.6 (2022)
https://doi.org/10.7776/ASK.2022.41.6.655

pISSN : 1225-4428
eISSN : 2287-3775

파킨슨병 환자에 대한 효과적인 음성인식 시스템

Effective speech recognition system for patients with Parkinson's disease

박희용,[1] 김률,[2†] 이상민[1†]

(Huiyong Bak,[1] Ryul Kim,[2†] and Sangmin Lee[1†])

[1]인하대학교 전기컴퓨터공학과, [2]인하대병원 신경과

(Received August 30, 2022; revised October 25, 2022; accepted November 9, 2022)

초 록: 파킨슨병 환자에게는 언어 장애가 만연하기 때문에 이러한 환자에게 적합한 음성인식 시스템이 필요하다. 본 논문에서는 파킨슨병 환자의 음성을 효과적으로 인식하는 음성인식 시스템을 제안한다. 음성인식 시스템은 먼저 건강한 사람의 음성 데이터를 사용하여 Globalformer를 사전 학습한 다음 상대적으로 매우 작은 양의 파킨슨병 환자의 음성 데이터를 사용하여 Globalformer를 미세 조정한다. 실험에는 AI 허브에서 구축한 건강한 사람의 음성 데이터셋과 인하대병원에서 수집한 파킨슨병 환자의 음성 데이터셋이 사용되어졌다. 실험 결과 제안된 음성인식 시스템은 22.15 %의 Character Error Rate(CER)으로 파킨슨병 환자의 음성을 인식하였으며, 다른 방법에 비해 우수한 인식률을 보였다.

핵심용어: 음성인식, 파킨슨병, Globalformer, 신호 처리

ABSTRACT: Since speech impairment is prevalent in patients with Parkinson's disease (PD), speech recognition systems suitable for these patients are needed. In this paper, we propose a speech recognition system that effectively recognizes the speech of patients with PD. The speech recognition system is firstly pre-trained with the Globalformer using the speech data from healthy people, and then fine-tuned using relatively small amount of speech data from the patient with PD. For this analysis, we used the speech dataset of healthy people built by AI hub and that of patients with PD collected at Inha University Hospital. As a result of the experiment, the proposed speech recognition system recognized the speech of patients with PD with Character Error Rate (CER) of 22.15 %, which was a better result compared to other methods.

Keywords: Speech recognition, Parkinson's disease, Globalformer, Signal processing

PACS numbers: 43.72.Ne, 43.30.Zk

I. 서 론

최근의 음성인식 시스템은 딥러닝 알고리즘을 적용하고 대규모 음성 데이터를 사용하면서 인식율이 크게 향상되었으며, 인공지능 비서, 대화로봇, 동시통역 등 다양한 애플리케이션에 활용되고 있다. 그러나 현재 활용되는 음성인식 시스템은 정상인의 음성을 사용하여 만들어진 음성인식 시스템이기 때문에 정상인의 음성과 다른 경우에 인식률이 낮을 것으로 추정된다.

†**Corresponding author:** Sangmin Lee (sanglee@inha.ac.kr)
Department of Electrical and Computer Engineering, Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Republic of Korea
(Tel: 82-32-860-7420)

†**Corresponding author:** Ryul Kim (arkrk86@inha.ac.kr)
Department of Neurology, Inha University Hospital, Inha University College of Medicine, 27 Inhang-ro, Jung-gu, Incheon 22332, Republic of Korea
(Tel: 82-32-890-3764)

신경퇴행성질환인 파킨슨병은 60세 이상 노인의 1 %에서 발병하며 파킨슨병 환자의 75 %는 언어장애를 가지고 있다. 파킨슨병 환자의 언어 장애는 중추 혹은 말초 신경계의 손상으로 인해 발음과 관련된 근육이 약화되어 발생한다. 약화되는 근육은 폐와 후두와 인두 및 비인두와 연구개와 조음기관(입술, 혀, 치아 밑 턱)이 포함된다. 이러한 근육의 약화로 인하여 파킨슨병 환자는 발음이 부정확하고, 목소리의 높낮이와 강세가 부족하고 불분명하며, 쉰 목소리를 가지게 된다. 언어장애로 인하여 파킨슨병 환자들은 다른 사람들과의 의사소통에 제한을 받으며 삶의 질이 떨어진다. 따라서 파킨슨병 환자의 음성인식 시스템을 개발하여 파킨슨병 환자의 언어를 명확하게 전달하고 언어 재활 치료에 도움을 주는 것이 필요하다.[1,2]

파킨슨병 환자를 위한 음성인식 시스템 연구는 파킨슨병 환자의 삶의 질을 향상시키기 위해 연구되어 졌다. Moro-Velazquez *et al*.[3]은 파킨슨병 환자의 음성을 인식하기 위하여 HMM/DNN을 건강한 사람의 음성으로 사전 학습시킨 후 파킨슨병 환자의 음성으로 미세 조정하여 음성인식 시스템을 구축하였다. Yu *et al*.[4]은 LSTM을 건강한 사람의 음성으로 사전 학습시킨 후 파킨슨병 환자의 음성으로 미세조정 하였다. 또한 주파수 마스킹 증강 기법을 사용하여 성능을 향상시켰다. 이와 같이 이전의 연구들은 건강한 사람의 음성으로 음성인식 모델을 사전학습 시킨 후 사전 학습된 음성인식 모델을 파킨슨병 환자의 음성으로 미세 조정하였다. 이러한 방법은 디지털화된 음성 데이터가 적은 파킨슨병 환자의 음성을 인식하는 효과적인 방법이다.

파킨슨병 환자들은 근육 그룹이 약화된 정도에 따라 발음이 조금씩 다르다.[5] 이러한 특징 때문에, 개인 맞춤형 음성인식 시스템 이전에 파킨슨병 환자의 음성의 지역 정보를 줄이고 전역 정보를 활용하는 것이 파킨슨병 환자 음성인식에 효과적일 것이다. 지역정보를 활용하게 되면 파킨슨병 환자들의 발음이 조금씩 다르기 때문에 음성인식 시스템이 음성 패턴을 알아내는데 많은 데이터가 필요하다. 하지만 파킨슨병 환자의 음성데이터는 적은 상황이다. 따라서 사람마다 발음이 조금씩 다르고 데이터가 부족할 때 광역정보를 활용하는 것이 효과적이다.

파킨슨병 환자의 음성의 전역 정보를 효과적으로 활용하여 파킨슨병 환자의 음성을 인식하기 위하여 Transformer에 squeeze and excitation module를 추가한 Globalformer와 Globalformer를 건강한 사람의 음성으로 사전학습 시킨 후 파킨슨병 환자의 음성으로 사전 학습된 Globalformer를 미세조정 하는 음성인식 시스템을 제안한다.

II. 재료 및 방법

2장에서는 본 논문에서 제시하는 파킨슨 환자 음성인식 시스템에 대한 전반적인 과정을 다룬다. 2.1절에서는 제안된 Globalformer에 사용된 Transformer와 squeeze and excitation module을 설명한다. 2.2절에서는 Globalformer를 제안한다. 2.3절에서는 Globalformer를 활용한 음성인식 시스템을 제안한다. 2.4절에서는 실험에 사용된 설정들을 소개하며, 2.5절에서는 실험에 사용된 데이터 셋과 음성인식 성능을 평가하기 위한 객관적인 평가 지표를 설명한다.

2.1 Transformer and squeeze and excitation module

Transformer는 RNN과 달리 attention만으로 구성된 모델이며, 시퀀스를 한 번에 입력 받아 self attention으로 전역 상호 작용을 캡처한다. 또한 Transformer는 병렬처리가 가능하기 때문에 학습 속도가 RNN에 비하여 빠르다.[6] Transformer는 시퀀스를 입력 받아 query와 key와 value로 임베딩 시킨다. 그 후 query와 key를 scaled dot-product attention 한다. 이때 입력 시퀀스 사이의 연관도가 구해진다. 구해진 연관도와 value를 곱함으로써 상호 연관도의 정보가 value에 반영된다. 이를 통해 Transformer는 입력 시퀀스의 전역 상호 작용을 캡처 할 수 있다. Transformer는 전역 상호 작용을 캡처할 수 있기 때문에 음성인식에서 널리 사용되고 있다.[7]

Squeeze and excitation module은 입력의 상대적 중요도를 캡처 할 수 있기 때문에 전역 정보를 활용하기 위해 사용되고 있다.[8,9] Squeeze and excitation module

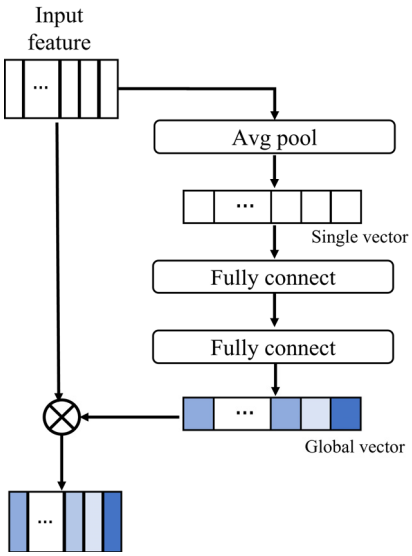

Fig. 1. (Color available online) Squeeze and excitation module.

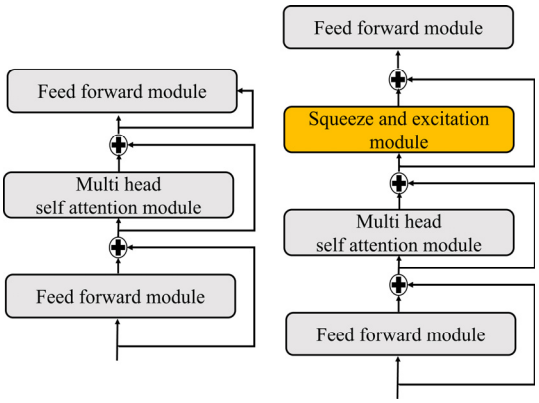

Fig. 2. (Color available online) Model architecture: (a) Transformer (b) globalformer.

는 Fig. 1과 같이 동작한다. 입력 특징은 avg pool로 단일 벡터로 변환된다. 단일 벡터는 sigmoid를 통해 단일 벡터의 상대적 중요도를 가지는 전역 벡터로 변환된다. 전역 벡터는 입력 특징에 곱해진다. 생성된 전역 벡터가 입력 특징에 곱해짐으로써, 입력특징의 전역정보가 캡처 되어진다.

2.2 Globalformer

파킨슨병 환자의 음성의 전역 정보를 효과적으로 활용하여 파킨슨병 환자의 음성을 인식하기 위하여 Fig. 2의 (a)인 Transformer구조에 squeeze and excitation module을 추가한 Fig. 2의 (b)인 Globalformer를 제안한다. Transformer는 입력되는 음성의 장거리 상호 작용을 캡처 할 수 있으며 squeeze and excitation module는 입력의 상대적 중요도를 캡처 할 수 있다. 따라서 Transformer와 squeeze and excitation module을 결합하면 음성의 전역 정보를 효과적으로 활용할 수 있다.

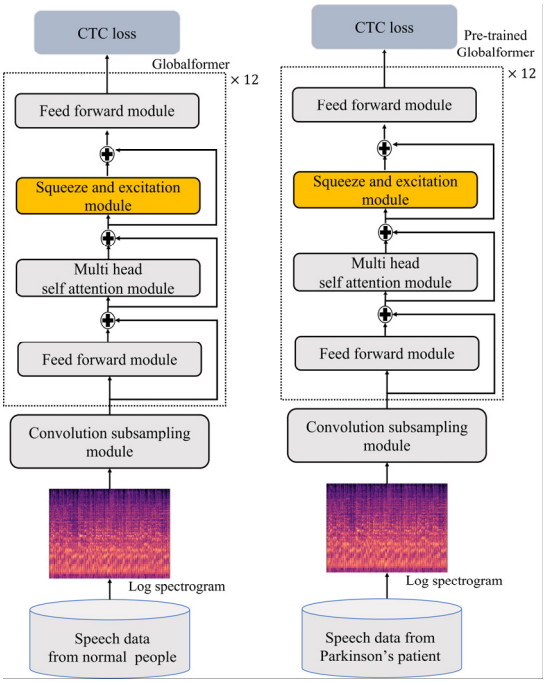

Fig. 3. (Color available online) Architecture of the proposed speech recognition system for patients with Parkinson's disease.

시퀀스가 Globalformer에 입력되면 시퀀스는 multi head self attention module를 통과한다. 이때 시퀀스에 상호 연관도의 정보가 반영된다. 그 후 시퀀스는 squeeze and excitation module을 통과한다. 이 때 시퀀스에 상대적 중요도 정보가 반영된다. Transformer와 squeeze and excitation module의 구조는 Anmol Gulati *et al.*과 Wei Han *et al.*이 제안한 구조를 따른다.[9,10]

2.3 제안하는 파킨슨병 환자를 위한 음성인식 시스템

파킨슨병 환자의 음성을 효과적으로 인식하기 위하여, Fig. 3과 같은 파킨슨병 환자를 위한 음성인식 시스템을 제안한다. 제안하는 음성인식 시스템은 사전학습 과정과 미세조정 과정으로 나뉜다. 사전학습 과정은 다음과 같다. 건강한 사람의 음성 데이터는

Table 1. Datasets used in this study.

Dataset	Type	Hours	No. utterances	No. speakers (Male/Female)
Parkinson's	Train	10.39	14,375	105 (62/43)
	Test	1.00	1,162	
Ksponspeech	Train	965.2	620,000	2000 (923/1,077)
	Test	3.9	2,545	1,348 (619/729)

log spectrogram으로 변환된다. log spectrogram은 convolution subsampling module에 입력되어 주파수와 시간 축의 해상도가 줄어든다. 줄어든 log spectrogram은 Globalformer에 입력된다. 그 후 CTC loss로 Globalformer가 학습된다. 미세조정 과정은 파킨슨병 환자의 음성데이터를 입력 받아 학습된 Globalformer를 CTC loss로 미세조정 한다. 파킨슨병 환자의 음성 데이터가 부족하기 때문에 제안하는 음성인식 시스템은 먼저 건강한 사람의 음성으로 Globalformer를 사전학습 시킨 후 파킨슨병 환자의 음성으로 사전 훈련된 Globalformer를 미세조정 한다. 또한 제안하는 음성인식 시스템은 Globalformer를 사용하기 때문에 환자마다 발음이 조금씩 다른 파킨슨병 환자의 음성을 효과적으로 인식할 수 있다.

2.4 실험 설정

Globalformer은 12개의 block과 8개의 attention head와 176의 model dimension 로 설정하였다. Spectrogram은 25 ms의 윈도우 크 기와 10 ms 윈도우 시프트로 구현되었다. 모델은 adam으로 최적화되었으며, warming up the learning rate으로 learning rate을 업데이트 시켰다.

제안하는 음성인식시스템은 네이버 clova에서 공개한 2003개의 한국어 음절 사전을 사용하였다.[10] 음향모델의 성능만을 측정하기 위하여 본 실험에서는 내부와 외부 언어모델을 사용하지 않는다.

2.5 데이터셋 과 평가 지표

인하대병원에서 수집된 파킨슨병 환자의 음성 데이터가 연구에 사용되었다. 데이터는 환자에게 1,054개의 문장을 랜덤하게 제시하여 읽게 한 후 음성을 녹음하여 수집되었다. 사용된 1,054개의 문장은 일상생활에서 사용되는 문장이다(e.g. 너 행복한 모습 보여줘야지 인스타로, 요즘에 유튜브 어떤 유튜브 많이 보니). Table 1과 같이 파킨슨병 환자 음성 데이터 셋은 105명의 환자에게서 수집된 15,537개의 발화로 구성되어 있다. 동일한 문장이 학습과 테스트에 사용되면 정확한 성능을 평가할 수 없기 때문에 train은 989개의 문장으로 test는 65개의 문장으로 나뉘었다. 그 결과 Table 1과 같이 15,537개의 발화 중 train은 14,375개 test는 1,162개로 나뉘어졌다.

Ksponspeech는 언어장애가 없는 2,000명의 한국인의 대화를 실내에서 녹음한 데이터 셋이며 2018년에 AI hub에서 공개되어졌다.[11]

평가 지표로는 Character Error Rate과 Word Error Rate(WER)이 사용되어졌다. CER은 문장의 음절 오류율 측정하는 평가지표이다. WER은 문장의 단어 오류율을 측정하는 평가지표이다. CER과 WER은 수식 1 로 계산되며, X, Y는 예측 및 정답 스크립트, D는 X, Y 사이의 Levenshtein 거리, 길이 L은 정답 Y의 길이이다.

$$CER(\%) = \frac{D}{L} * 100,\ D = Distance_{Lev}(X, Y). \quad (1)$$

III. 결과 및 고찰

3.1절에서는 실험의 결과를 비교한다. 3.2절에서는 제안하는 파킨슨병 환자를 위한 음성인식 시스템과 이전 연구를 비교한다.

3.1 실험결과 비교

본 연구에서는 네 가지 음성인식 모델로 네 가지 실험이 이루어졌다. 실험에 사용된 음성인식 모델은 Transformer와 Globalformer와 Conformer 와 Con-Globalformer이다. Conformer는 Anmol Gulati *et al.*이 제안한 모델로서, Transformer와 CNN을 결합하여 음성의 전역 정보와 지역 정보를 활용하는 모델이다. Conformer

Table 2. Results of experiment 1.

Experiment	Model	Train data	Test data	CER
1	Conformer	Ksponspeech train	Ksponspeech test	11.49 %
	Con-globalformer			12.10 %
	Transformer			11.80 %
	Globalformer			11.82 %
	Conformer		Parkinson's test	41.09 %
	Con-globalformer			44.04 %
	Transformer			40.90 %
	Globalformer			43.07 %

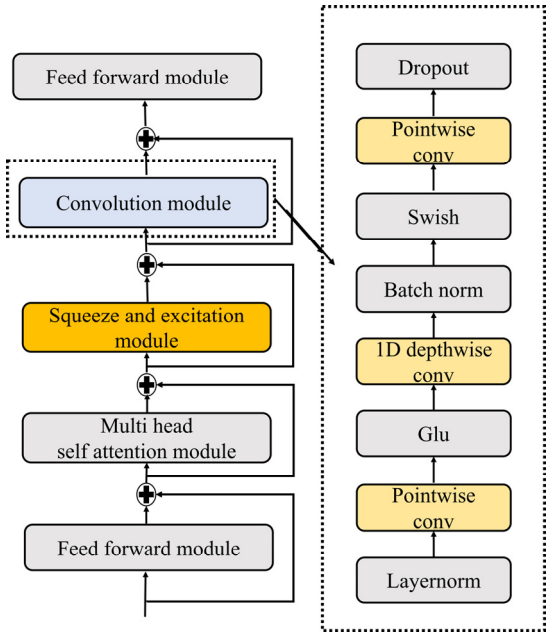

Fig. 4. (Color available online) Con-globalformer architecture.

는 음성인식에서 우수한 성능을 보여 활발히 사용되고 있다.[12] Con-globalformer는 Fig. 4와 같이 Globalformer에 convolution module을 결합한 모델이다. Con-Globalformer는 Globalformer와 CNN이 결합되기 때문에 음성의 전역정보와 지역정보를 활용한다. Conformer와 Con-globalformer는 음성의 지역 정보를 활용하기 때문에, 파킨슨병 환자의 음성의 지역 정보가 인식률에 어떠한 영향을 미치는지 알 수 있다. Con-globalformer에서 사용된 Convolution modul의 구조는 Anmol Gulati *et al.*이 제안한 구조를 따른다.[12]

네 가지 실험은 다음과 같다. 첫번째 실험은 건강한 사람의 음성인식 시스템으로 파킨슨병 환자의 음성을 인식하였다. 두번째 실험은 네 가지 음성인식 모델을 파킨슨병 환자의 음성으로만 학습 후 파킨슨병 환자의 음성을 인식하였다. 세번째 실험은 제안하는 음성인식 시스템과 같이 네 가지 음성인식 모델을 사전학습 및 미세조정 하여 파킨슨병 환자의 음성을 인식하였다. 네번째 실험은 제안하는 음성인식시스템의 성능을 검증하기 위하여, 파킨슨병 환자의 음성 데이터셋의 train과 test를 다르게 나누어 실험을 진행하였다.

Table 3. Results of experiment 2.

Experiment	Model	CER
2	Conformer	70.79 %
	Con-globalformer	70.29 %
	Transformer	66.17 %
	Globalformer (proposed)	**61.57 %**

실험 1로써, 네 가지 음성인식모델을 건강한 사람의 음성으로 학습시킨 후에 건강한 사람의 음성과 파킨슨 환자의 음성을 인식하였다. 실험 결과는 Table 2와 같이 건강한 사람의 음성으로 구축된 음성인식시스템은 파킨슨병 환자의 음성에 대해 인식률이 낮은 것을 알려준다.

실험 2로써, 네 가지 음성인식 모델을 파킨슨병 환자의 음성으로만 학습시킨 후 파킨슨병 환자의 음성을 인식하였다. 이 결과 Table 3와 같이 Globalformer가 61.57 %의 CER로 가장 좋은 인식률을 가졌다. Table 3을 통해 Globalformer가 Transformer에 비해 약 7 %의 성능향상이 있으며, 음성인식 모델에 convolution module이 결합되면 인식률이 하락하는 것을 알려준다. 성능 향상의 이유로는 Transformer의 광역정보 활

Table 4. Results of experiment 3.

Experiment	Model	CER
3	Conformer	22.43 %
	Con-globalformer	22.45 %
	Transformer	22.26 %
	Globalformer (proposed system)	**22.15 %**

Table 5. Results of experiment 4.

Model	CER (Split 1)	CER (Split 2)	CER (Split 3)
Conformer	73.10 %	72.79 %	73.21 %
Con-globalformer	89.85 %	86.48 %	67.38 %
Transformer	68.47 %	65.89 %	66.01 %
Globalformer (proposed)	**60.78 %**	**64.46 %**	**59.43 %**

용능력을 강화했기 때문이며 성능 하락의 이유로는 환자마다 발음이 조금씩 다른 파킨슨병 환자 음성을 지역정보가 학습하기에는 데이터가 부족했기 때문이다.

또한 실험 2의 결과는 파킨슨병 환자의 음성데이터로만 학습하면 데이터가 부족하기 때문에 음성인식시스템의 인식률이 낮은 것을 알려준다. 따라서 기존의 파킨슨병 환자에 대한 음성인식시스템 연구와 같이 정상인의 음성으로 사전학습 할 필요가 있다.

실험 3으로써, 네 가지 음성인식모델을 건강한 사람의 음성으로 사전학습 시킨 후 파킨슨병 환자의 음성으로 미세조정 하여 파킨슨병 환자의 음성을 인식하였다. 실험 결과는 Table 4와 같이 Conformer는 22.43 %의 CER, Con-globalformer는 22.45 %의 CER, Transformer는 22.26 %의 CER, Globalformer는 22.15 %의 CER로 파킨슨병 환자의 음성을 인식하였다. 이 결과는 제안하는 음성인식 시스템이 가장 낮은 CER을 보이는 것을 알려준다.

실험 4로써, 제안하는 음성인식시스템의 성능을 검증하기 위하여, 파킨슨병 환자의 음성 데이터셋은 train과 test를 다르게 3가지로 나뉘었다. 실험은 파킨슨병 환자의 음성 데이터를 네 가지 음성인식 모델을 학습하여 이루어졌다. 건강한 사람의 음성이 사전학습 되면 환자마다 발음이 조금씩 다른 파킨슨병 환자의 음성의 광역정보를 효과적으로 사용하는 것의 성능을 정확하게 확인할 수 없기 때문에, 파킨슨병 환자의 음성만을 사용하여 성능을 확인하였다. 실험 결과는 Table 5와 같이 제안하는 Globalformer가 파킨슨병 환자의 음성을 가장 잘 인식하는 것을 알려준다.

Table 6. Comparison of results with previous studies.

Researcher	Patient's language	CER	WER
Ref. [3]	Spanish	-	47.00 %
Ref. [4]	English	28.40 %	43.00 %
Our group	Korean	**22.15 %**	**42.21 %**

3.2 이전연구와의 비교

Table 6와 같이 본 연구의 실험결과와 이전의 연구들의 실험결과를 비교하였다. 이전의 연구에서 사용된 파킨슨병 환자의 언어가 다르기 때문에 본 연구의 실험 결과와는 정확한 비교가 어렵다. 하지만 제안하는 음성인식 시스템의 인식률이 이전의 연구들에 비해 우수하기 때문에 제안하는 음성인식 시스템이 이전의 연구의 음성인식 시스템 보다 더 우수할 가능성이 매우 높을 것으로 추정된다.

IV. 결 론

파킨슨병 환자를 위한 음성인식 시스템은 파킨슨병 환자의 의사소통 능력을 향상시키기 때문에 파킨슨병 환자를 위한 음성인식 시스템이 필요하다. 본 논문에서는 효과적으로 파킨슨병 환자의 음성을 인식하는 음성인식 시스템을 제안한다. 제안하는 음성인식 시스템은 Globalformer를 건강한 사람의 음성 데이터로 사전학습 시킨 후 훈련된 Globalformer를 파킨슨병 환자의 음성으로 미세조정 한다. 제안하는 음성인식 시스템은 파킨슨병 환자의 음성을 22.15 %의 CER로 인식하여 다른 방법 보다 우수한 성능을 보였다. 본 연구에서는 파킨슨병 환자의 음성을 인식하기 위하여 음성인식 모델의 전역정보 캡처 능력을 강화하였다. 전역정보 캡처 능력이 강화되면 실시간 인식 성능이 하락될 것으로 예상되지만 음성을 한 번에 입력 받는 명령어 인식의 성능은 더 효과적인 결과를 보일 것으로 기대된다.

감사의 글

이 연구는 한국연구재단 기초연구지원사업의 지원으로 수행됨(과제번호, NRF-2020R1A2C2004624와 NRF-2021R1C1C1011822).

References

1. A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, “Speech impairment in a large sample of patients with Parkinson’s disease,” Behav Neurol. **11**, 131-137 (1999).
2. A. Kain, X. Niu, J.-P. Hosom, Q. Miao, and J. van Santen, “Formant re-synthesis of dysarthric speech,” Proc. ISCA Workshop on SSW5, 25-30 (2004).
3. L. Moro-Velazquez, J. Cho, S. Watanabe, M. A. Hasegawa-Johnson, O. Scharenborg, H. Kim, and N. Dehak, “Study of the performance of automatic speech recognition systems in speakers with Parkinson’s disease,” Proc. 20th Interspeech, 3875-3879 (2019).
4. Q. Yu, Y. Ma, and Y. Li, “Enhancing speech recognition for parkinson’s disease patient using transfer learning technique,” J. Shanghai Jiaotong Univ. (Science), **27**, 90-98 (2022).
5. S. O. Caballero-Morales and F. Trujillo-Romero, “Evolutionary approach for integration of multiple pronunciation patterns for enhancement of dysarthric speech recognition,” Expert Syst. Appl. **41**, 841-852 (2014).
6. A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” Proc. Advances in NIPS, 1-11 (2017).
7. L. Dong, S. Xu, and B. Xu, “Speech transformer: a no-recurrence sequence-to sequence model for speech recognition,” Proc. IEEE ICASSP, 5884-5888 (2018).
8. J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” Proc. the IEEE conf. CVPR, 7132-7141 (2018).
9. W. Han, Z. Zhang, Y. Zhang, J. Yu, C. C. Chiu, J. Qin, A. Gulati, R. Pang, and Y. Wu, “ContextNet: Improving convolutional neural networks for automatic speech recognition with global context,” Proc. Interspeech, 3610-3614 (2020).
10. J. W. Ha, K. Nam, J. Kang, S. W. Lee, S. Yang, H. Jung, E. Kim, H. Kim, S. Kim, H. A. Kim, K. Doh, C. K. Lee, N. K. Sung, and S. Kim, “ClovaCall: Korean goal-oriented dialog speech corpus for automatic speech recognition of contact centers,” Proc. Interspeech, 409-413 (2020).
11. J.-U. Bang, S. Yun, S. H. Kim, M. Y. Choi, M. K. Lee, Y. J. Kim, D. H. Kim, J. Park, Y. J. Lee, and S. H. Kim, “KsponSpeech: Korean spontaneous speech corpus for automatic speech recognition,” Appl. Sci. **10**, 6936 (2020).
12. 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,” Proc. Interspeech, 5036-5040 (2020).

저자 약력

▸ 박 희 용 (Huiyong Bak)

2021년 2월 : 인하대학교 메카트로닉스공학과 학사
2021년 3월 ~ 현재 : 인하대학교 전기컴퓨터공학과 석사

▸ 김 률 (Ryul Kim)

2011년 2월 : 제주대학교 의학과 학사
2020년 2월 : 서울대학교 의학과 석사
2022년 3월 ~ 현재 : 인하대병원 신경과 조교수

▸ 이 상 민 (Sangmin Lee)

1987년 : 인하대학교 전자공학과 학사
1989년 : 인하대학교 전자공학과 석사
2000년 : 인하대학교 전자공학과 박사
2006년 ~ 현재 : 인하대학교 전자공학과 교수