한국음향학회지 제38권 제5호 pp. 601~606 (2019)
The Journal of the Acoustical Society of Korea Vol.38, No.5 (2019)
https://doi.org/10.7776/ASK.2019.38.5.601

pISSN : 1225-4428
eISSN : 2287-3775

문자소 기반의 한국어 음성인식

Korean speech recognition based on grapheme

이문학,[1] 장준혁[2†]

(Mun-hak Lee[1] and Joon-Hyuk Chang[2†])

[1]한양대학교 전자컴퓨터통신공학과, [2]한양대학교 융합전자공학부

(Received July 1, 2019; revised August 26, 2019; accepted September 3, 2019)

초 록: 본 논문에서는 한국어 음성인식기 음향모델의 출력단위로 문자소를 제안한다. 제안하는 음성인식 모델은 한글을 G2P(Grapheme to Phoneme)과정 없이 초성, 중성, 종성 단위의 문자소로 분해하여 음향모델의 출력단위로 사용하며, 특별한 발음 정보를 주지 않고도 딥러닝 기반의 음향모델이 한국어 발음규정을 충분히 학습해 낼 수 있음을 보인다. 또한 기존의 음소기반 음성인식 모델과의 성능을 비교 평가하여 DB가 충분한 상황에서 문자소 기반 모델이 상대적으로 뛰어난 성능을 가진다는 것을 보인다.
핵심용어: 음성인식, 딥러닝, 발음사전, Kaldi

ABSTRACT: This paper is a study on speech recognition in the Korean using grapheme unit (Cho-sumg [onset], Jung-sung [nucleus], Jong-sung [coda]). Here we make ASR (Automatic speech recognition) system without G2P (Grapheme to Phoneme) process and show that Deep learning based ASR systems can learn Korean pronunciation rules without G2P process. The proposed model is shown to reduce the word error rate in the presence of sufficient training data.
Keywords: Automatic speech recognition, Deep learning, Lexicon, Kaldi

PACS numbers: 43.72.Bs, 43.72.Ne

I. 서 론

한글은 표음문자지만 단어의 철자와 발음이 항상 일치하지는 않는다. 따라서 한국어 음성인식 시스템을 구축하기 위해서는 단어와 단어에 대응되는 발음열이 명시된 발음사전을 제작하여 사용하는 것이 일반적이다. 한국어 발음사전을 제작하기 위해서는 발음변이 규칙을 찾고 예외처리를 적용하는 G2P(grapheme to phoneme)과정이 필요하다.[1] 하지만 근래의 음성인식 연구는 이러한 발음변이 규칙과 예외발음의 경우 역시 딥러닝을 통해 학습이 가능함을 보여주고 있다.[2] 본 논문에서는 G2P과정을 거치지 않은 초성, 중성, 종성의 문자소 단위 음성인식 모델을 제안하며 기존의 음소 단위 발음사전을 이용한 모델과 성능을 비교하여 평가한다.

†**Corresponding author:** Joon-Hyuk Chang (jchang@hanyang.ac.kr)
Department of Electronic Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea
(Tel: 82-2-2220-0355, Fax: 82-2-2291-0357)

II. 기존연구

한국어 문자는 앞과 뒤에 이어지는 문자에 따라 다양한 발음을 갖는다. 한 예로 살구[살구], 볼살[볼쌀], 살의[사릐]의 '살'은 동일한 문자로 표기되나 모두 다른 발음을 갖는다. 한국어 음성인식에서는 이러한 문자표기와 실제 발음 사이의 불일치를 극복하기 위해 발음사전을 이용한다. 발음사전을 제작하기 위해서는 한국어의 자소-음소간, 음소-변이음간 변환규칙, 예외 발음에 대해 일관적으로 적용할 수 있는 발음변이 규정이 필요하다.[1] 국내 여러 연구기관별로 발음변이를 규정하는 규정집을 보유하고 있으며, 이 규정집을 이용하여 문자표기를 음소열로 치

환할 수 있다.[3] 한국어의 특정 단어는 한 개 이상의 발음을 갖기도 하는데, 이러한 다중발음 정보를 발음사전에 추가하여 음성 인식률의 향상을 도모하는 연구가 진행되었다.[4] 기존의 연구에서는 수작업을 통해 한국어 발음 규정이 적용된 발음사전을 제작하였으나 이는 전문적인 노동력의 소요가 큰 작업이다. 따라서 이러한 과정을 자동화하기 위한 연구가 이어졌다.[5]

음소단위 발음열을 이용한 발음사전은 문장을 어절단위로 분절하여 음성인식 모델을 구축하는 경우 효과적이다. 하지만 어절 이외 형태소나 BPE(Byte Pair Encoding) 알고리즘 기반의 subword 단위[6]로 문장을 분절하여 음성인식을 진행할 경우 앞선 subword와 이어지는 subword 사이 발음 변이 모델링이 어렵다는 문제가 존재하며 이를 극복하기 위한 연구가 이어졌다.[7]

앞서 설명한 발음사전 제작과정은 복잡할 뿐 아니라 유지보수를 위해 전문적인 노동력의 소요가 지속적으로 발생한다. 이러한 한계를 극복하기 위해 발음사전이 필요하지 않은 End-to-End 음성인식에 대한 연구가 이루어졌다.[8] End-to-End 모델은 딥러닝의 출력으로 음소 이외 문자소, subword unit 등을 이용하며, attention 기반의 End-to-End 음성인식 모델을 이용한 Reference [2]에서는 모델의 출력으로 문자소를 이용하는 것이 음소를 이용하는 것 보다 높은 성능을 보임을 확인했다.

이러한 근래의 연구 결과는 딥러닝을 이용한 음성인식 모델이 문자표기와 실제 발음간 불일치를 학습할 수 있다는 사실을 보여준다. 한국어의 경우 영어 대비 모음의 숫자가 많고 발음규정이 명확하여 딥러닝을 통한 학습이 용이할 것으로 예상되며 따라서 이에 대한 연구가 필요하다.

III. 문자소 기반 음성인식

3.1 음소기반 음성인식

음소기반 음성인식은 음향모델의 출력 단위로 음소를 이용한다. 따라서 음소기반 음성인식의 음향모델은 음성의 특정 구간을 입력 받아 음소들의 존재 확률 분포를 출력한다. 출력된 음소존재 확률 분포

그 사이에 문을 열고 밖으로 나갔다
그 사이에 무늘 열고 바끄로 나갇따

Fig. 1. Example for spelling transcription & pronunciation transcription.

Table 1. Symbol for vowel.

	Phoneme for vowel
Single vowel	ㅣ (i) ㅖ (e) ㅐ (E) ㅜ (u) ㅗ (o) ㅏ (a) ㅡ (U) ㅓ (v) ㅚ (O) ㅟ (Y)
Diphthong	ㅟ (wi) ㅞ (we) ㅙ (wE) ㅝ (wv) ㅘ (wa) ㅖ (je) ㅒ (jE) ㅑ (ja) ㅕ (jv) ㅛ (jo) ㅠ (ju) ㅢ (xi)

Table 2. Symbol for consonant.

	Bilabial	Alveolar consonant	Alveo-palatal	Velar	Glottals
Implosive	ㅂ b ㅍ p ㅃ B	ㄷ d ㅌ t ㄸ D		ㄱ g ㅋ k ㄲ G	
Fricative		ㅅ s ㅆ S			ㅎ h
Affricate			ㅈ z ㅊ c ㅉ Z		
Nasal	ㅁ m	ㄴ n		ㅇ N	
Liguids		ㄹ l/r			

를 이용하여 음성인식의 최종 출력인 문장을 얻기 위해서는 음소열과 단어의 철자를 연결해주는 매개가 필요하며 이러한 매개체로 발음사전이 이용된다. 음소를 기반으로 하는 한국어 발음사전을 제작하는 과정은 다음과 같다. 1. 텍스트 전처리 2. tokenizing 3. G2P(grapheme to phoneme) 4. 예외 규정 적용. 본 연구에서는 문자소 기반 모델의 대조군으로 음소 기반 음성인식 모델을 이용하며 SiTEC에서 제공하는 발음사전을 이용해 이를 학습을 진행하였다. SiTEC 발음 사전은 다음과 같은 과정을 통해 제작되었다. 1. 숫자, 문자, 특수기호를 발음에 대응되는 한글로 치환 2. 띄어쓰기를 기준으로 어절 단위 tokenizing 3. 철자전사 된 어절에 일대일 대응되는 발음전사 어절 생성(Fig. 1) 4. G2P 테이블(Tables 1과 2)을 이용하여 규칙기반 발음열 생성 5. 모음 탈락 등 한글을 이용한 전사로 표현하기 어려운 어절의 발음열 생성 6. 제작된 음소열에 대응하는 단어와 조합하여 발음사전 구축. 제작된 발음사전은 다중발음이 고려되어 35,697

가스 [GasU]
가스 [gasU]
가스 [gasx]

Fig. 2. Example for mutiple pronunciation.

개의 단어에 대해 68,539개의 음소열이 등록되어 있으며 다중발음 사례는 다음과 같다(Fig. 2).

다중발음 정보를 이용하는 것은 인식률 향상에 도움이 되나 지나치게 많은 다중발음을 추가하는 것은 인식과정에 혼란을 가중시켜 인식률 하락의 원인이 되기도 한다.[9] 이러한 인식률 하락을 막기 위해 본 논문에서는 음성 인식 결과를 이용해 발음사전의 크기를 제한했다. 실험에서 이용된 발음사전 크기 제한 방법은 다음과 같다. 최초 모든 다중발음이 포함된 발음사전을 이용해 인식기를 학습한다. 학습된 인식기로 학습 데이터 베이스 전체를 디코딩하며 디코딩 결과에 가장 높은 빈도수로 등장한 한 개의 발음열 이외 다중 발음열은 모델의 혼잡도를 높이는 불필요한 발음열로 보아 삭제하였다. 또한 학습 결과에 등장하지 않았더라도 OOV(Out of Vocabulary) 문제 방지를 위해 각 단어 당 한 개의 발음열이 존재할 수 있게 하였다. 제한된 발음사전의 크기는 35,697개로 32,842개의 다중 발음이 제거되었다.

3.2 문자소 기반 음성인식

문자기반 음성인식의 경우 앞선 음소기반 음성인식과 같은 G2P 과정을 거치지 않으며, 따라서 문자를 그대로 음향모델의 출력으로 이용한다. 하지만 한글에는 11,172개에 달하는 문자가 존재하며 이를 음향모델의 출력단위로 이용하는 것에는 두 가지 문제가 있다. 첫째로 학습 DB내 존재하지 않는 문자가 발생하며, 둘째로 적은 빈도수로 등장하는 문자의 경우 학습이 어렵다.

따라서 본 연구에서는 한글 문자를 구성하는 기본 단위인 초성, 중성, 종성을 출력단위로 하는 음성인식 모델을 제안한다. 실험은 한국어 초성 19개와 중성 21개 종성 27개 중 학습 및 평가 DB내 등장하지 않는 세 종류의 종성 'ㅱ', 'ㆄ', 'ㆅ'을 제외한 총 64개의 문자소를 이용하여 진행되었다(Fig. 3).

초성(Cho-sung) :
'ㄱ','ㄲ','ㄴ','ㄷ','ㄸ','ㄹ','ㅁ','ㅂ','ㅃ','ㅅ','ㅆ',
'ㅇ','ㅈ','ㅉ','ㅊ','ㅋ','ㅌ','ㅍ','ㅎ'

중성(Jung-sung) :
'ㅏ','ㅐ','ㅑ','ㅒ','ㅓ','ㅔ','ㅕ','ㅖ','ㅗ','ㅘ','ㅙ',
'ㅚ','ㅛ','ㅜ','ㅝ','ㅞ','ㅟ','ㅠ','ㅡ','ㅢ','ㅣ'

종성(Jong-sung) :
'ㄱ','ㄲ','ㄳ','ㄴ','ㄵ','ㄶ','ㄷ','ㄹ','ㄺ','ㄻ','ㄼ','ㄽ',
'ㅁ','ㅂ','ㅄ','ㅅ','ㅆ','ㅇ','ㅈ','ㅊ','ㅋ','ㅌ','ㅍ','ㅎ'

Fig. 3. Korean grapehme.

3.3 문맥의존 문자소 이용

트라이폰(tri-phone) 모델링은 특정 음소의 앞 뒤로 이어지는 음소정보를 고려하여 음소의 종류를 확장하는 방법이다. 한국어의 경우 대개 앞선 문자의 종성과 이어지는 문자의 초성이 결합하여 발음변이가 발생한다. 따라서 앞 뒤 문자소를 고려한 문맥의존 문자소를 음향모델의 출력단위로 이용할 경우 발음 전사를 직접 제작하는 과정 없이 발음변이의 모델링이 가능하다. 트라이폰을 생성하는 방법은 크게 두 가지가 있다. 첫째는 학습 DB로부터 군집화를 통해 트라이폰을 추출하는 data driven 방법이고 둘째는 음향학적 결정 트리(phonetic decision tree)를 이용하여 트라이폰을 생성하는 방법이다.[10] 본 연구에서는 tree기반의 방법을 이용한다. 실험에 이용한 음향학적 결정 트리는 바이너리 트리로 루트 노드에 모노폰 HMM-state 정보, 리프 노드에 트라이폰 HMM-state 정보, 내부 노드에 앞뒤로 등장하는 폰 정보를 담고 있으며 각 트라이폰 HMM-state의 유사도를 평가하여 하향식 군집화를 진행한다. 이러한 군집화는 한계 트라이폰 HMM-state의 개수에 도달하거나 트라이폰 HMM-state의 개수를 늘리는 것이 모델의 성능에 악영향을 미칠 때까지 반복된다. 본 연구에서는 앞서 설명한 트라이폰 생성 방법과 동일한 방법으로 문맥의존 문자소를 생성하며 최종적으로 생성된 모델은 총 2,392개의 문맥의존 문자소 HMM-state을 가진다.

Table 3. Training DB information.

	Dict01	Dict01 + Dict02
# corpus	41,666	84,103
# word	8,666	35,694
Words appeared less than twice (Percentage of total words)	1,308 (15 %)	17,327 (48 %)
Words appeared more than 10 times (Percentage of total words)	4,961 (57 %)	11,306 (32 %)

IV. 실험 및 결과

4.1 데이터 베이스(DB)

학습 데이터로 SiTEC에서 제작된 Dict01과 Dict02를 이용했으며 데이터셋 별 남, 여 화자 각 200명씩, 총 84,103개의 낭독체 문장으로 구성되어 있다. Dict01과 Dict02를 통합한 전체 학습 데이터 베이스의 발화 시간은 약 100시간이다. 또한 숫자, 특수문자, 문장기호는 모두 제거하거나 발음에 대응되는 한글로 치환하였다.

평가는 Dict01과 Dict02 데이터셋 별 나누어 진행하였다. Dict01은 8,666개 단어로 이루어진 41,666개 발화로 구성되어 있으며 Dict02는 33,256개 단어로 이루어진 42,437개 발화로 구성되어 있다(Table 3).

4.2 데이터 베이스(DB) 증폭

학습 데이터 양에 따른 인식률의 변화를 살피기 위해 학습 DB를 증폭하였다. 잡음환경 모델링을 위해 배경 소음과 전경 소음을 학습 DB에 섞었다. 전경 소음의 경우 학습 데이터의 발화 중간 중간 임의 발생하게 하였으며, 배경 소음의 경우 발화 전체에 걸쳐 발생하도록 하였다. 신호대 잡음비는 0 dB, 5 dB, 10 dB, 15 dB, 20 dB 중 임의 선정하였다. 또한 잔향환경 모델링을 위해 너비와 폭 1 m ~ 50 m, 높이 2 m ~ 5 m 가량의 공간 정보가 들어있는 RIR(Room Impulse Response)를 임의로 생성하여 합성하였다. 또한 RIR을 합성할 때 잡음과 음성간 발생 위치를 임의 설정하여 모델링 하였다. 실험에 이용된 RIR, Noise DB는 Reference [11]의 공개된 데이터를 이용하였다. 학습 이외 평가에도 증폭된 DB를 이용하였으며 실험결과에 Noise셋으로 표기하였다. 또한 실환경에서의 성능 검증을 위해 마우스 시뮬레이터로 3 m 거리에서 재녹음한 DB를 평가에 이용하였으며 실험결과에 Distance셋으로 표기하였다. 마지막으로 학습 및 평가 데이터 베이스 내 OOV문제가 발생하지 않도록 모든 단어를 발음사전에 등록 하였다.

4.3 실험 환경

본 연구는 TDNN-HMM(Time Delay Neural Network-Hidden Markov Model) 기반의 하이브리드 음성인식 모델을 이용해 진행하였다. TDNN-HMM 기반 하이브리드 모델은 GMM(Gaussian Mixture Model)을 이용해 생성된 음성구간별 alignment 정보를 신경망(TDNN)을 이용해 재학습한 모델로 딥러닝 기반의 음향모델의 한국어의 발음변이 학습 가능 여부를 확인하는 본 연구의 목적에 적합하다. TDNN 모델은 7개의 은닉층으로 구성되어 있으며 각 은닉층 별 625개의 hidden node를 갖는다. 음향모델의 입력은 MFCC(Mel Frequency Cepstral Coefficient)를 이용하며 특정 순간의 앞 뒤 11개 프레임의 피쳐를 연쇄하여 TDNN의 입력으로 사용하였다. 언어모델은 SRILM toolkit[12]의 3-gram 모델을 이용하였으며 학습과 디코딩을 비롯한 대부분의 실험은 Kaldi[13]를 이용하여 진행하였다.

4.4 실험 결과

두 종류의 상황을 상정한 실험을 진행했다. 첫 번째는 인식하고자 하는 단어들에 대해 충분한 양의 데이터가 확보된 상황이며, SiTEC Dict01 데이터를 이용해 학습 및 평가를 진행하였다. 두번째는 인식하고자 하는 단어들에 대해 충분하지 못한 데이터가 존재하는 상황이며 SiTEC Dict01 데이터셋과 SiTEC Dict02 데이터셋을 이용해 학습 및 평가를 진행하였다. Dict01 데이터셋 의 경우 데이터셋 내 2번 이하 등장한 단어의 비율이 15 %로 대부분의 단어들에 대해 학습 가능한 양의 음성데이터가 존재한다. 반면 Dict01과 Dict02를 함께 사용한 두 번째 상황의 경우 2번 이하 등장한 단어의 비율이 48 %로 많은 단어가 데이터셋 내 희소하게 등장한다는 특징을 갖는다(Table 3).

Table 4. Word error rate phoneme & grapheme training DB : Dict01.

WER (%)	Phoneme-1 (Before prunning)	Phoneme-2 (After prunning)	Grapheme
Clean-Dict01	0.30	0.24	0.23
Distance-Dict01	1.07	0.94	0.72
Noise-Dict01	7.87	5.19	5.09

Table 5. Word error rate phoneme & grapheme training DB : Dict01 + Dict02.

WER (%)	Phoneme-1 (Before prunning)	Phoneme-2 (After prunning)	Grapheme
Clean-Dict01	0.25	0.24	0.25
Clean-Dict02	0.89	0.88	0.89
Distance-Dict01	0.52	0.66	0.8
Distance-Dict02	11.95	5.28	9.05
Noise-Dict01	4.34	4.11	6.31
Noise-Dict02	3.00	2.61	4.12

첫 번째 실험은 DB가 충분한 상황을 상정하여 Dict 01 데이터셋만을 이용해 진행되었다. 문자소 기반 모델과 다중발음이 고려된 음소기반 모델, 제한된 다중발음만이 고려된 음소기반 모델의 총 세 가지 모델의 성능을 비교하였으며 실험을 통해 Clean/Noise/Distance의 모든 평가셋에 대해 문자소 기반 음성인식 모델이 음소 기반 음성인식 모델 대비 높은 인식 성능을 획득하였다. 따라서 데이터가 충분한 경우 딥러닝 기반 음향모델이 한국어 발음 변이를 훌륭히 학습함을 확인할 수 있었다. 또한 기존 연구에서의 결과와 마찬가지로 다중발음을 제한하여 혼잡도를 낮춤으로서 음성인식 성능을 향상시킬 수 있음을 확인하였다(Table 4).

두 번째 실험은 DB가 부족한 상황을 상정하여 Dict 01과 Dict02 데이터셋을 함께 이용하였다. 앞선 실험과 마찬가지로 문자소 기반 모델과 다중발음이 고려된 음소기반 모델, 제한된 다중발음만이 고려된 음소기반 모델의 세 가지 모델의 성능을 비교하였다. 실험 결과 음소 기반 모델이 문자소 기반 모델대비 Distance셋과 Noise셋에 대해 높은 성능을 보여주었으며 따라서 데이터가 충분하지 못한 경우 음소기반 모델의 일반화 성능이 문자소 기반 모델 대비 뛰어남을 알 수 있었다(Table 5).

Table 6. Word error rate phoneme & grapheme training DB : Dict01 + Dict02 (Augmented).

WER (%)	Phoneme-2 (After Prunning)	Grapheme
Clean-Dict01	0.25	0.21
Clean-Dict02	0.86	0.86
Distance-Dict01	0.28	0.24
Distance-Dict02	1.45	1.32
Noise-Dict01	0.29	0.27
Noise-Dict02	0.92	0.88

세 번째 실험은 data augmentation 기법을 통해 데이터의 양을 두 배로 증폭시켰으며, 두 번째 실험과 마찬가지로 Dict01과 Dict02 데이터셋을 모두 이용하였다. data augmentation 결과 Clean/Noise/Distance 셋 모두에 대해 기존 모델 대비 성능이 증진 하였다. 증진 폭은 문자소 기반 모델이 음소기반 모델 대비 컸으며 결과적으로 문자소 기반 모델의 인식성능이 음소 기반 모델 대비 뛰어났다. 세 번째 실험을 통해 데이터의 부족으로 인한 문자소 기반 모델의 인식률 하락 문제가 data augmentation 방법을 통해 극복 가능함을 확인하였다(Table 6).

V. 결 론

본 논문에서는 문자소를 기반으로 하는 음성인식 모델을 제안한다. 제안하는 모델은 딥러닝을 기반으로 하는 음향모델과 문맥의존 문자소를 이용하며, 학습 데이터가 충분한 상황에서 기존의 음소기반 음성인식 대비 높은 인식 성능을 보인다(Table 4). 하지만 데이터가 충분하지 못한 상황에서의 일반화 성능이 음소기반 모델 대비 낮다는 단점을 가지고 있으며(Table 5), 본 논문에서는 이를 극복하기 위해 data augmentation 기법을 제안한다. data augmentation 기법을 통해 DB를 증폭하는 경우 앞선 데이터 부족으로 인한 문자소 기반 모델의 일반화 성능 저하 문제가 해결되었으며(Table 6), 음소기반 모델 대비 높은 성능을 획득하였다.

감사의 글

본 연구는 방위사업청 및 국방과학연구소에 의해 설립된 신호정보 특화연구센터의 지원을 받아 수행되었음.

References

1. J. W. Yoo, “A study on method of constructing pronunciation unit for continuous speech recognition,” Hankuk University of Foreign Studies Rep., 1995.
2. K. Irie, R. Prabhavalkar, A. Kannan, A. Bruguier, D. Rybach, and P. Nguyen, “Model unit exploration for sequence-to-sequence speech recognition,” arXiv:1902.01955 (2019).
3. H. Hong and J. M. Hwa, *Phonetics-based design of phoneme - like units for Korean speech recognition,* (Master's degree, Seoul University graduate school, 2009).
4. L. G. Nim and J. M. Hwa, “Pronunciation dictionary for continuous speech recognition” (in Korean), Proc. KIISE. Conf. 197-199 (2000).
5. M. -S. Na and M. H. Chung, “Assistive program for automatic speech transcription based on G2P conversion and speech recognition” (in Korean), Proc. KSSS, 131-132 (2016).
6. M. Schuster and K. Nakajima, ”Japanese and Korean voice search,” Proc. IEEE ICASSP, 5149-5152 (2012).
7. J. -U. Bang, S. -H. Kim, and O. -W. Kwon, “Performance of speech recognition unit considering morphological pronunciation variation,” Phonetics and Speech Sciences, **10**, 111-119 (2018).
8. W. Chan, N. Jaitly, Q. Le, and O. Vinyals “Listen, attend and spell: A neural network for large vocabulary conversational speech recognition,” Proc. IEEE ICASSP, 4960-4964 (2016).
9. G. N. Lee and M. H. Jeong, “Pronunciation lexicon modeling and design for Korean large vocabulary continuous speech recognition,” Proc. Interspeech, 4-8 (2004).
10. S. J. Young, J. J. Odell, and P. C. Woodland, “Tree-based state tying for high accuracy acoustic modelling,” Proc. the ARPA Human Language Technology Workshop, 307-312 (1994).
11. T. Ko, V. Peddinti,, D. Povey, M. L. Seltzer, and S. Khudanpur, “A study on data augmentation of reverberant speech for robust speech recognition,” Proc. IEEE ICASSP, 5220-5224 (2017).
12. A. Stolcke, “SRILM an extensible language modeling toolkit,” Proc. ICSLP, 5220-5224 (2002).
13. D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, “The Kaldi speech recognition toolkit,” IEEE Workshop on Automatic Speech Recognition and Understanding (2011).

저자 약력

▸ 이 문 학 (Mun-hak Lee)

2018년 2월 : 한양대학교 기계공학과 학사
2019년 6월 ~ 현재 : 한양대학교 전자컴퓨터통신공학부 석박사통합과정 재학 중

▸ 장 준 혁 (Joon-Hyuk Chang)

2004년 2월 : 서울대학교 전기컴퓨터공학부 박사
2000년 3월 ~ 2005년 4월 : (주)넷더스 연구소장
2004년 5월 ~ 2005년 4월 : 캘리포니아 주립대학 산타바바라(UCSB) 박사후 연구원
2005년 5월 ~ 2005년 8월 : 한국과학기술연구원(KIST) 연구원
2005년 9월 ~ 2011년 2월 : 인하대학교 전자공학부 조교수
2011년 3월 ~ 2017년 3월 : 한양대학교 융합전자공학부 부교수
2017년 3월 ~ 현재 : 한양대학교 융합전자공학부 정교수