한국음향학회지 제41권 제3호 pp. 319~325 (2022)
The Journal of the Acoustical Society of Korea Vol.41, No.3 (2022)
https://doi.org/10.7776/ASK.2022.41.3.319

pISSN : 1225-4428
eISSN : 2287-3775

가변 길이 입력 발성에서의 화자 인증 성능 향상을 위한 통합된 수용 영역 다양화 기법

Integrated receptive field diversification method for improving speaker verification performance for variable-length utterances

신현서,[1] 김주호,[1] 허정우,[1] 심혜진,[1] 유하진[1†]

(Hyun-seo Shin,[1] Ju-ho Kim,[1] Jungwoo Heo,[1] Hye-jin Shim,[1] and Ha-Jin Yu[1†])

[1]서울시립대학교 컴퓨터과학과

(Received March 16, 2022; accepted May 4, 2022)

초 록: 화자 인증 시스템에서 입력 발성 길이의 변화는 성능을 하락시킬 수 있는 대표적인 요인이다. 이러한 문제점을 개선하기 위해, 몇몇 연구에서는 시스템 내부의 특징 가공 과정을 여러가지 서로 다른 경로에서 수행하거나 서로 다른 수용 영역(Receptive Field)을 가진 합성곱 계층을 활용하여 다양한 화자 특징을 추출하였다. 이러한 연구에 착안하여, 본 연구에서는 가변 길이 입력 발성을 처리하기 위해 보다 다양한 수용 영역에서 화자 정보를 추출하고 이를 선택적으로 통합하는 통합된 수용 영역 다양화 기법을 제안한다. 제안한 통합 기법은 입력된 특징을 여러가지 서로 다른 경로에서 다른 수용 영역을 가진 합성곱 계층으로 가공하며, 가공된 특징을 입력 발성의 길이에 따라 동적으로 통합하여 화자 특징을 추출한다. 본 연구의 심층신경망은 VoxCeleb2 데이터세트로 학습되었으며, 가변 길이 입력 발성에 대한 성능을 확인하기 위해 VoxCeleb1 평가 데이터 세트를 1 s, 2 s, 5 s 길이로 자른 발성과 전체 길이 발성에 대해 각각 평가를 수행하였다. 실험 결과, 통합된 수용 영역 다양화 기법이 베이스라인 대비 동일 오류율을 평균적으로 19.7 % 감소시켜, 제안한 기법이 가변 길이 입력 발성에 의한 성능 저하를 개선할 수 있음을 확인하였다.

핵심용어: 화자 인증, 가변 길이 발성, 심층신경망, 수용 영역

ABSTRACT: The variation of utterance lengths is a representative factor that can degrade the performance of speaker verification systems. To handle this issue, previous studies had attempted to extract speaker features from various branches or to use convolution layers with different receptive fields. Combining the advantages of the previous two approaches for variable-length input, this paper proposes integrated receptive field diversification that extracts speaker features through more diverse receptive field. The proposed method processes the input features by convolutional layers with different receptive fields at multiple time-axis branches, and extracts speaker embedding by dynamically aggregating the processed features according to the lengths of input utterances. The deep neural networks in this study were trained on the VoxCeleb2 dataset and tested on the VoxCeleb1 evaluation dataset that divided into 1 s, 2 s, 5 s, and full-length. Experimental results demonstrated that the proposed method reduces the equal error rate by 19.7 % compared to the baseline.

Keywords: Speaker verification, Variable-length utterance, Deep neural network, Receptive field

PACS numbers: 43.60.Bf, 43.72.Fx

†**Corresponding author:** Ha-Jin Yu (hjyu@uos.ac.kr)
School of Computer Science, College of Engineering, University of Seoul, 163 Siripdae-ro, DongDaemun-gu, Seoul 02504, Republic of Korea
(Tel: 82-2-6490-5697, Fax: 82-2-6490-2444)

Copyright ⓒ2022 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.

I. 서 론

화자 인증(speaker verification)은 입력된 발성의 화자가 사전에 등록된 화자와 일치하는지 판단하는 과제로, 화자 인증 시스템은 다음과 같은 절차를 통해 수행된다. 먼저, 시스템은 화자 특징 추출기를 사용하여 입력된 발성과 사전 등록된 화자의 발성으로부터 화자 특징이 포함된 고정 차원의 특징 벡터를 추출한다. 이후, 시스템은 두 특징 벡터 간의 유사도 점수를 계산하여 사전 정의한 임계값을 기준으로 동일 화자의 여부를 결정한다. 최근 심층 학습의 발전에 힘입어, 화자 인증 분야에서는 심층신경망 기반의 화자 특징 추출기를 활용한 시스템이 우수한 성능을 보였다.[1,2]

그럼에도 불구하고 대다수의 심층신경망 기반 화자 인증 시스템은 입력 발성의 길이가 학습에 사용된 발성 길이와 크게 다를 경우 급격한 성능 저하가 발생한다.[3] 이는 짧은 발성이 긴 발성에 비하여 상대적으로 활용할 수 있는 화자 정보량이 제한적이기 때문에 발성 길이에 따라 화자 정보량이 달라서 성능 저하를 유발하는 것이다. 따라서 화자 인증 시스템은 발성 길이의 변화로 인한 성능 하락을 완화하기 위해 발성 길이 별로 적합한 화자 정보를 추출할 수 있도록 구축되어야 한다.

Kim *et al*.[4]은 다양한 특징 지도를 추출하고 이를 선택적으로 집계함으로써 가변 길이 발성 문제를 해결하고자 하였다. 이에 따라 제안된 확장된 동적 스케일링 정책(Extended Dynamic Scaling Policy, EDSP) 기법은 심층신경망 내부에 특징 지도를 확대 또는 축소하는 해상도 경로를 추가하여 화자 특징을 추출하는 기법이다. 특징 지도의 해상도 변경으로 인해 변환된 특징 지도는 동일한 합성곱 계층이 적용되어도 기존 특징 지도와 다른 화자 특징을 추출하게 된다.

앞서 제시한 연구를 통해 서로 다른 방식으로 화자 특징을 추출하는 것은 가변 길이 발성에서 성능 하락 완화에 유효함을 확인할 수 있다. 본 연구는 EDSP를 개선하고자 다양한 수용 영역을 통해 서로 다른 방식으로 특징 지도를 추출하고자 한다. 다양한 수용 영역은 서로 상이한 특징 지도를 생성할 수 있기 때문에 다양한 화자 특징을 추출할 수 있다. 이에 따라 다양한 수용 영역에서 추출된 특징 지도를 입력 발성의 길이에 따라 동적으로 통합하여 화자 특징을 추출하는 통합된 수용 영역 다양화 기법을 제안한다. 제안한 기법은 각 해상도 경로에서 서로 다른 합성곱 계층을 통하여 서로 다른 수용 영역에서 추출된 특징 지도를 확보하며, 각 특징 지도를 softmax attention 기법을 통해 심층신경망이 적합한 특징을 선택할 수 있도록 동작한다. 따라서 제안한 기법은 수용 영역 다양화와 선택적 집계 방식으로 인해 다양한 길이의 발성에 강인한 성능을 보일 수 있다.

본 논문은 심층신경망 학습을 위해 VoxCeleb2 데이터세트[5]를 사용하였으며 Kim *et al*.[4]에서 사용한 ResNeXt[6]을 베이스라인으로 한다. 학습된 시스템의 평가는 VoxCeleb1 데이터세트[7]를 사용하였으며, 가변 길이 발성에 의한 성능 저하를 확인하기 위해 평가 발성을 1 s, 2 s, 5 s로 길이로 자른 발성과 전체 발성에 대해 각각 동일 오류율(Equal Error Rate, EER)을 측정하였다. 실험 결과, 제안한 기법에 의하여 베이스라인 대비 발성 길이 별 동일 오류율이 평균적으로 19.7 % 감소하였다. 본 연구에서는 해당 결과를 통해 제안한 통합 기법이 가변 길이 발성으로 인한 화자 인증 성능 저하를 완화함을 실험적으로 확인하였다.

본 논문의 II장에서는 연구에서 활용한 기존 기술을 설명한다. III장에서는 본 논문에서 제안한 심층신경망의 구조와 특징을 소개한다. IV장에서는 화자 인증 시스템의 실험 설계를 기술하고, 실험 결과를 분석한다. 최종적으로 V장에서는 결론을 서술한다.

II. 기존 기술

2.1 확장된 동적 스케일링 정책(Extended Dynamic Scaling Policy, EDSP)

EDSP는 Fig. 1(a)와 같이 저해상도 경로, 기존 해상도 경로, 고해상도 경로와 각 경로에서 병렬적으로 추출된 화자 특징을 집계하는 게이트 모듈로 구성된다. 저해상도 경로에서는 다운 샘플링 연산을 통해 낮은 해상도의 특징 지도를 생성한 뒤 합성곱 연산을 수행함으로써 긴 발성에 대해 포괄적인 화자 특징 추출이 가능하다. 반면에, 고해상도 경로에서는 업 샘플링 연산을 통해 높은 해상도의 특징 지도를

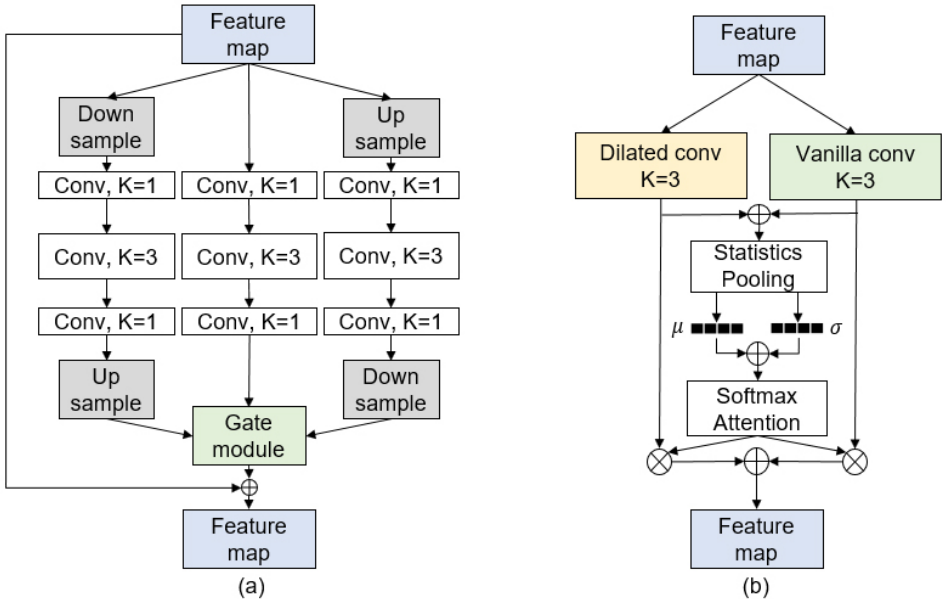


Fig. 1. (Color available online) Illustration of EDSP and ISKConv method. (a) Left: structure of EDSP. (b) Right: structure of ISKConv. K refer to the kernel size of a convolutional layer. μ and σ denote a mean vector and a standard deviation vector ($\oplus$: the element-wise summation, $\otimes$: the element-wise multiplication).

생성한 뒤 합성곱 연산을 수행함으로써 화자 정보가 상대적으로 부족한 짧은 발성에 대해 정교한 화자 특징 추출이 가능하다. 이후, 각 경로에서 추출된 특징 지도는 기존 해상도와 동일한 크기가 되도록 업/다운 샘플링 연산을 통해 길이가 조정되고, 기존 경로의 특징 지도와 함께 게이트 모듈의 입력으로 사용된다. 게이트 모듈에서는 3가지 경로에서 추출된 특징 지도를 softmax attention을 활용하여 선택적으로 집계한다. 따라서, EDSP 기법은 시간 축을 기준으로 변형된 다양한 특징 지도로부터 병렬적으로 화자 정보를 추출하고 집계함으로써 발성 길이에 따라 적합한 특징 지도의 해상도를 선택하여 화자 정보를 추출할 수 있다.

2.2 향상된 선택적 커널 합성곱(Improved Selective Kernel Convolution, ISKConv)

ISKConv[8] 연산은 Fig. 1(b)와 같이 병렬적으로 배치된 팽창된 합성곱(dilated convolution) 계층과 일반 합성곱 계층, statistics pooling 계층, 그리고 softmax attention 계층으로 구성된다. 팽창된 합성곱 계층은 특정 간격으로 시간 축의 입력 값을 건너뛰어 커널 길이보다 큰 수용 영역에서 화자 특징을 추출할 수 있으므로 일반 합성곱 계층과 다른 특징을 추출한다. 두 합성곱 계층에서 추출된 특징은 원소별 덧셈 연산 이후 statistics pooling 계층으로 입력되어 시계열 정보가 압축된 발성 단위(utterence-level) 정보를 가진 평균 벡터와 표준편차 벡터를 생성한다. 이후 두 벡터는 서로 원소별 덧셈 연산에 의해 합쳐지며 softmax attention 계층에 입력되어 화자 정보가 선택적으로 집계된다. 따라서 ISKConv 연산은 서로 다른 수용 영역을 통해 발성 길이의 차이에 따른 특징을 추출하고 이를 선택적으로 집계함으로써 가변 길이 발성에서 성능 저하를 완화할 수 있다.

III. 제안한 기법

EDSP 기법은 다양한 특징 지도에서 화자 정보를 추출 및 집계하여 발성 길이 별로 적합한 화자 특징을 추출하고자 하였다. 이러한 접근법에 착안하여, 본 논문에서는 다양한 수용 영역을 통해 여러 특징 지도를 확보하고, 이를 선택적으로 통합하는 통합된 수용 영역 다양화 기법을 제안한다.

제안하는 기법은 Fig. 2와 같이 EDSP의 합성곱 블록 계층을 ISKConv 연산 계층으로 대체한 구조를 가

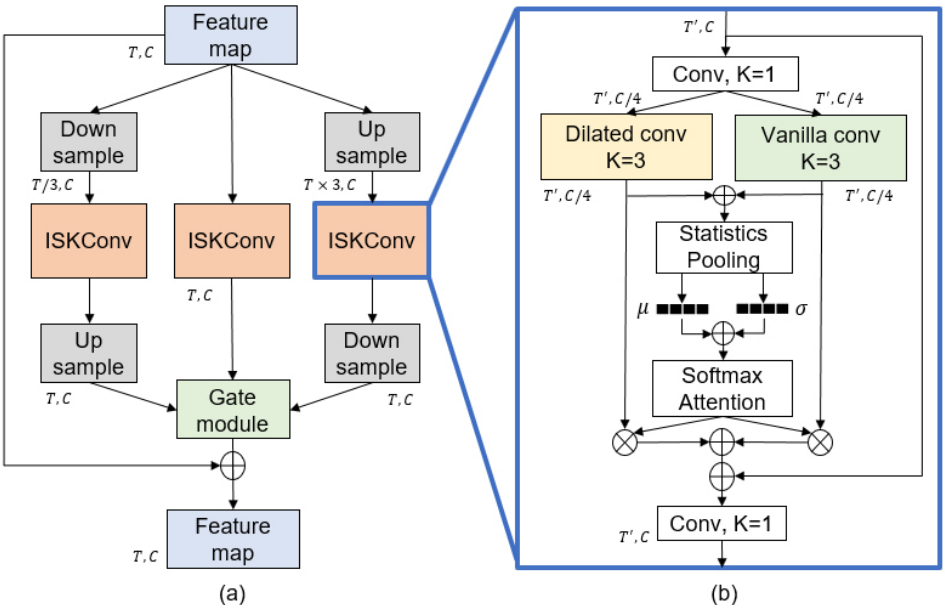


Fig. 2. (Color available online) Illustration of the proposed method. (a) Left: the intergrated layer with EDSP and ISKConv blocks. (b) Right: bottleneck based ISKConv architecture. T and C denote time and channel scale of each feature maps.

Table 1. The temporal size and elements of feature map in each resolution path. T denotes the number of input feature map's time axis elements and x_i means i-th element of the input feature. α, β and γ are parameters of a kernel in transposed convolutional layer.

Path	Temporal size	Elements
Low	$T/3$	$\left\{\frac{x_1+x_2+x_3}{3}, \cdots, \frac{x_{T-2}+x_{T-1}+x_T}{3}\right\}$
Original	T	$\{x_1, x_2, x_3, \cdots, x_T\}$
High	$T\times 3$	$\{\alpha x_1, \beta x_1, \gamma x_1, \cdots, \alpha x_T, \beta x_T, \gamma x_T\}$

지며 동작 과정은 다음과 같다. 먼저, 입력된 특징 지도는 기존 경로와 더불어 다운, 업 샘플링 연산을 통해 저해상도 경로, 고해상도 경로로 분리되어 입력된다. 이때, 다운 샘플링 연산은 average pooling을 사용하고, 업 샘플링 연산은 전치 합성곱을 사용한다. 각 계층의 커널 크기는 3으로 동일하다. Table 1은 경로별 특징 지도의 시간 축 크기 및 각 채널의 원소를 나타내며, 저해상도 및 고해상도 분기에서 특징 지도의 시간 축 크기가 3배로 축소 또는 확장됨을 확인할 수 있다.

이후, 각 해상도 경로에서는 ISKConv 연산을 통해 특징을 추출한다. 제안한 기법은 3개의 해상도 경로에서 일반 합성곱 계층과 팽창된 합성곱 계층을 병렬적 배치함으로써 총 6개의 서로 다른 수용 영역에서 특징 지도를 획득할 수 있다. 사용된 일반 합성곱의 커널 크기는 3이며, 팽창된 합성곱은 5의 커널 크기를 가지지만 커널 내의 변수는 커널 중간을 건너뛰며 3개만 정의된다. Table 2는 입력 특징 지도의 한 채널의 원소가 x_i라고 할 때, 각 해상도 경로별 합성곱 계층의 6가지 수용 영역을 나타내며 각각의 수용 영역이 서로 상이함을 확인할 수 있다. 이를 통해 제안한 기법은 다양한 수용 영역에서 화자 특징 추출이 가능하며, 이는 가변 길이 발성에서 다양하게 분포된 화자 정보를 추출할 수 있음을 의미한다.

ISKConv 연산 내부의 statistics pooling 계층은 시계열 정보를 압축하여 평균 벡터와 표준편차 벡터를 생성하여 프레임 단위 정보를 발성 단위 정보로 변환한다. 제안한 기법에서는 해상도 경로별로 발성 단위 정보가 추출되므로, softmax attention 계층이 각 경로를 통과한 발성 단위 화자 정보를 1차적으로 통합한다. 이후, 각 경로별 ISKConv 연산의 출력은 원

Table 2. The receptive field of each convolutional layers in the proposed method.

Path	Conv	Receptive field
Low	Vanilla	$\left\{\frac{x_{t-4}+x_{t-3}+x_{t-2}}{3}, \frac{x_{t-1}+x_t+x_{t+1}}{3}, \frac{x_{t+2}+x_{t+3}+x_{t+4}}{3}\right\}$
	Dilated	$\left\{\frac{x_{t-7}+x_{t-6}+x_{t-5}}{3}, \frac{x_{t-1}+x_t+x_{t+1}}{3}, \frac{x_{t+5}+x_{t+6}+x_{t+7}}{3}\right\}$
Original	Vanilla	$\{x_{t-1}, x_t, x_{t+1}\}$
	Dilated	$\{x_{t-2}, x_t, x_{t+2}\}$
High	Vanilla	$\{\alpha x_t, \beta x_t, \gamma x_t\}$
	Dilated	$\{\gamma x_{t-1}, \beta x_t, \alpha x_{t+1}\}$

래의 해상도로 복원된 뒤 게이트 모듈을 거쳐 2차적으로 통합되어 단일 특징 지도가 된다. 추가적으로 ISKConv 연산으로 인한 파라미터의 과도한 증가를 방지하기 위해 Fig. 2(b)와 같이 커널 크기가 1인 합성곱 계층을 ISKConv 연산의 입력 부분과 출력 부분에 배치하여 bottleneck 구조를 형성시켰다. 또한 잔차 경로를 ISKConv 연산에 추가하여 증가한 파라미터에 대해 역전파가 효율적으로 일어날 수 있도록 하였다.

이러한 구조로 인하여 통합된 수용 영역의 다양화 기법은 서로 다른 수용 영역을 통해 화자 정보를 추출하며, 이를 해상도별 발성 정보를 기준으로 선택적으로 집계함으로써 심층신경망이 가변 길이 발성에서 강인한 성능을 보일 수 있도록 한다.

IV. 실험 설계 및 결과

4.1 데이터세트

본 연구에서는 6,112명의 화자로부터 수집한 1,128,246개의 발성으로 구성된 VoxCeleb2 데이터세트 전체를 사용하여 심층신경망을 학습하였고, 1,251명의 화자로부터 수집한 153,516개의 발성으로 구성된 VoxCeleb1 데이터세트를 사용하여 학습된 신경망을 평가하였다. 평가 trial은 VoxCeleb1에서 공식적으로 제공하는 기존 평가 시험(original evaluation trial, Vox1-O), 확장된 평가 시험(extended evaluation trial, Vox1-E), 고난도 평가 시험(hard evaluation trial, Vox1,H)을 사용하였다. Vox1-O은 40명의 화자의 발성으로 구성된 37,720개의 trial, Vox1-E은 1,251명의 화자의 발성으로 구성된 579,818개의 trial, Vox1-H은 1,190명의 화자의 발성으로 구성된 550,894개의 trial로 구성된다. 본 논문에서는 다양한 길이의 발성에서 화자 인증 성능을 평가하기 위해 전체 길이 발성뿐만 아니라, Vox1-O trial에 포함된 평가 발성을 1 s, 2 s, 5 s로 잘라서 평가를 수행하였다.

4.2 실험 설계

본 논문에서는 발성의 원시파형을 입력으로 사용하는 ResNeXt 모델을 베이스라인으로 사용하였다. Mini-batch는 심층신경망이 가변 길이 발성을 학습할 수 있도록 하기 위해 화자 당 3.59초의 고정 길이 발성과 1 s~3.59 s의 무작위 길이 발성 두 가지를 추출하여 총 발성의 개수가 320개가 되도록 구성하였다. 학습에는 $1e^{-3}$의 learning rate를 사용하였으며, 이는 80 epoch에 걸쳐 cosine LR scheduler[9]를 통해 $1e^{-7}$까지 감소한다. 학습의 최적화 알고리즘은 AMSGrad[10]를 사용하였으며 $1e^{-4}$ weight decay를 적용하였다.

4.3 실험 결과

Table 3은 기존의 EDSP, ISKConv 기법이 적용된 신경망과 제안한 기법이 적용된 신경망의 가변 길이 입력 발성에 대한 화자 인증 성능 평가 결과를 나타낸다. 실험 결과, 베이스라인은 전체 길이 발성 평가에서 2.185 %의 동일 오류율을 보였지만 발성 길이가 짧아질수록 급격한 성능 저하가 나타나며 1 s 발성 평가에서 6.129 %의 동일 오류율을 보였다. 이에 비하여 EDSP 기법은 베이스라인 대비 모든 평가 조건에서 동일 오류율을 평균적으로 개선하였다. ISKConv 방식을 적용한 신경망은 Vox1-O의 전체 길이 발성 평가와 5 s 발성 평가에서 베이스라인에 비해 화자 인증 성능이 하락되었으나, 1 s, 2 s 발성 평가에서 성능 개선을 보였다. 이는 긴 발성에서 더 좋은 성능을 보이는 신경망이 반드시 짧은 발성에서도 좋은 성능을 보이는 것은 아니라는 것을 시사하며 ISKConv가 가

Table 3. Experimental results of various model on VoxCeleb1 test dataset.

Model	Vox1-O (EER %)				Vox1-E (EER %)	Vox1-H (EER %)
	full	1s	2s	5s		
ResNeXt (Baseline)	2.185	6.129	3.552	2.420	1.952	3.420
ResNeXt + EDSP	2.004	5.281	3.279	2.180	1.812	3.378
ResNeXt + ISKConv	2.259	5.684	3.478	2.468	2.038	3.871
Proposed method	**1.819**	**4.931**	**2.760**	**1.967**	**1.658**	**3.203**

변 길이 발성으로 인한 성능 하락을 완화할 수 있음을 확인할 수 있다. 이러한 결과들을 통해, EDSP 기법과 ISKConv 기법이 가변 길이 발성에 유효한 기법임을 확인하였다.

본 연구에서 제안한 통합된 수용 영역 다양화 기법이 적용된 신경망은 모든 평가에서 가장 우수한 성능을 보였다. 베이스라인 대비 Vox1-O의 발성 길이별 동일 오류율은 전체 길이 발성 평가에서 16.75 % 감소하였으며, 1 s, 2 s, 5 s 발성 평가에서 각각 19.55 %, 22.30 %, 18.72 %가 감소하여 평균적으로 동일 오류율이 19.7 % 감소하였다. 해당 결과는 전체 길이 발성에 비하여 짧은 발성에서 개선율이 향상됨을 통해 제안한 기법이 가변 길이 발성으로 인한 성능 하락을 완화할 수 있음을 알 수 있다. 또한 제안한 기법은 EDSP 및 ISKConv 기법에 비해서도 성능이 개선되었음을 확인하였다. 따라서 제안한 기법이 기존 기법들에 비하여 가변 길이 발성에 강인한 성능을 보일 수 있음을 실험적으로 증명하였다.

추가적으로 본 연구에서는 화자 인증에서의 일반화 성능을 확인하기 위해 Vox1-E, Vox1-H trial을 사용한 평가를 수행하였다. 실험 결과, 제안한 기법이 적용된 신경망은 평가에서 타 시스템 대비 가장 우수한 성능인 1.658 %, 3.203 %의 동일 오류율을 보여주었다. 따라서, 제안한 기법은 가변 길이 발성뿐만 아니라 일반적인 화자 인증에서 우수한 일반화 성능을 가질 수 있음을 확인하였다.

V. 결 론

본 논문에서는 발성 길이 변화로 인한 화자 인증 시스템의 성능 하락을 개선하기 위해 통합된 수용 영역 다양화 기법을 제안하였다. 제안한 통합 기법은 다양한 수용 영역에서 추출된 특징을 softmax 기반 attention 기법을 통해 선택적으로 집계하는 방식으로 동작한다. 실험을 통해 제안한 기법은 Vox1-O의 모든 발성 길이 평가 조건에서 평균적으로 베이스라인 대비 오류율을 19.7 % 감소시킴을 확인하였다. 이로써 통합된 수용 영역 다양화 기법이 적용된 화자 인증 시스템은 가변 길이 입력 발성에 강인하게 동작함을 실험적으로 증명하였다. 향후 연구로는 제안한 통합 기법에 사용된 게이트 모듈과 ISKConv의 softmax attention 기법을 통합할 수 있는 효율적인 화자 정보 집계 방식을 탐구하고자 한다.

감사의 글

이 논문은 2020년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초 연구사업임(2020R1A2C1007081).

References

1. E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. G. Dominguez, “Deep neural networks for small footprint text-dependent speaker verification,” Proc. IEEE ICASSP, 4052-4056 (2014).
2. D. Snyder, D. G. Romero, G. Sell, D. Povey, and S. Khudanpur, “X-vectors: Robust dnn embeddings for speaker recognition,” Proc. IEEE ICASSP, 5329-5333 (2018).
3. C. Zhang and K. Koishida, “End-to-end textindependent speaker verification with triplet loss on short utterances,” Proc. Interspeech, 1487-1491 (2017).
4. J. H. Kim, H. J. Shim, J. W. Heo, and H. J. Yu, “RawNeXt: Speaker verification system for variable-duration utterances with deep layer aggregation and extended dynamic scaling policies,” Proc. IEEE ICASSP, 7647-7651 (2022).
5. J. S. Chung, A. Nagrani, and A. Zisserman, “Voxceleb2:

Deep speaker recognition," Proc. Interspeech, 1086-1090 (2018).
6. S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transfomations for deep neural networks," Proc. CVPR, 1492-1500 (2017).
7. A. Nagrani, J. S. Chung, and A. Zisserman, "Voxceleb: a large-scale speaker identification dataset," Proc. Interspeech, 2616-2620 (2017).
8. Y. Wu, J. Zhao, C. Guo, and J. Xu, "Improving Deep CNN Architectures with Variable-Length Training Samples for Text-Independent Speaker Verification," Proc. Interspeech, 81-85 (2021).
9. I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," Proc. ICLR, (2017).
10. S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of adam and beyond," Proc. ICLR, (2018).

저자 약력

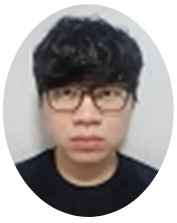
▸ 신 현 서 (Hyun-seo Shin)

2018년 2월 : 서울시립대학교 컴퓨터과학부 학사
2021년 9월 ~ 현재 : 서울시립대학교 컴퓨터과학부 석사과정

▸ 김 주 호 (Ju-ho Kim)

2020년 2월 : 서울시립대학교 컴퓨터과학부 학사
2020년 3월 ~ 현재 : 서울시립대학교 컴퓨터과학부 석·박사통합과정

▸ 허 정 우 (Jungwoo Heo)



2018년 3월 ~ 현재 : 서울시립대학교 컴퓨터과학부 학·석사통합과정

▸ 심 혜 진 (Hye-jin Shim)

2017년 2월 : 광운대학교 동북아통상학부 학사
2019년 2월 : 서울시립대학교 컴퓨터과학부 석사
2020년 3월 ~ 현재 : 서울시립대학교 컴퓨터과학부 박사과정

▸ 유 하 진 (Ha-Jin Yu)

1990년 2월 : KAIST 전산학과 학사
1992년 2월 : KAIST 전산학과 석사
1997년 2월 : KAIST 전산학과 박사
2017년 ~ 2000년 : LG 전자 전자기술원 선임연구원
2000년 ~ 2002년 : SL2(주) 연구소장
2002년 ~ 현재 : 서울시립대학교 컴퓨터과학부 교수