JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 21, No. 12 : 2291~2297 Dec. 2017

효과적인 음성 인식 평가를 위한 심층 신경망 기반의 음성 인식 성능 지표

지승은 · 김우일*

Speech Recognition Accuracy Measure using Deep Neural Network for Effective Evaluation of Speech Recognition Performance

Seung-eun Ji · Wooil Kim*

Department of Computer Science & Engineering, Incheon National University, Incheon 22012, Korea

요 약

본 논문에서는 음성 데이터베이스를 평가하기 위해 여러 가지의 음성 특성 지표 추출 알고리즘을 설명하고 심층 신경망 기반의 새로운 음성 성능 지표 생성 방법을 제안한다. 선행 연구에서는 효과적인 음성 인식 성능 지표를 생성하기 위해 대표적인 음성 인식 성능 지표인 단어 오인식률(Word Error Rate, WER)과 상관도가 높은 여러 가지 음성 특성 지표들을 조합하여 새로운 성능 지표를 생성하였다. 생성된 음성 성능 지표는 다양한 잡음 환경에서 각 음성 특성 지표를 단독으로 사용할 때보다 단어 오인식률과 높은 상관도를 나타내어 음성 인식 성능을 예측하는데 효과적임을 입증 하였다. 본 논문에서는 심층 신경망을 기반으로 한 음성 특성 지표 추출 방법에 대해 설명하며 선행 연구에서 조합에 사용한 GMM(Gaussian Mixture Model) 음향 모델 확률 값을 심층 신경망 학습을 통해 추출한 확률 값으로 대체해 조합함으로써 단어 오인식률과 보다 높은 상관도를 갖는 것을 확인한다.

ABSTRACT

This paper describe to extract speech measure algorithm for evaluating a speech database, and presents generating method of a speech quality measure using DNN(Deep Neural Network). In our previous study, to produce an effective speech quality measure, we propose a combination of various speech measures which are highly correlated with WER(Word Error Rate). The new combination of various types of speech quality measures in this study is more effective to predict the speech recognition performance compared to each speech measure alone. In this paper, we describe the method of extracting measure using DNN, and we change one of the combined measure from GMM(Gaussican Mixture Model) score used in the previous study to DNN score. The combination with DNN score shows a higher correlation with WER compared to the combination with GMM score.

키워드 : 단어 오인식률, 상관 계수, 성능 예측, 음성 인식 성능 지표, 심층 신경망

Key word : WER, Correlation coefficient, Performance prediction, Speech quality measure, DNN

Received 26 July 2017, **Revised** 25 September 2017, **Accepted** 03 October 2017

* **Corresponding Author** Wooil Kim (E-mail: wikim@inu.ac.kr, Tel:+82-32-835-8459)
Department of Computer Science & Engineering, Incheon National University, Incheon 22012, Korea

Open Access https://doi.org/10.6109/jkiice.2017.21.12.2291 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. 서 론

음성 인식 기술은 데이터를 입력하거나 서비스를 제공받기 위해 음성을 이용하는 기술이다. 음성 인식 기술을 이용한 시스템은 여러 가지 데이터 입력 방법 중 사람의 의사 전달과 가장 유사한 방법을 사용하기 때문에 사용이 편리하다는 장점이 있다. 하지만 음성 신호는 마우스나 키보드를 통한 일반적인 입력 신호들과 달리 잡음에 노출된 상태에서 실시간으로 처리되기 때문에 실제 환경에서 시스템의 인식률이 떨어지는 경우가 빈번하다. 따라서 음성 인식 시스템은 시중에 출시되기 이전에 시스템의 인식 성능을 검증하기 위해 대량의 음성 데이터베이스를 이용한 평가 작업이 요구된다. 또한 평가에 필요한 데이터베이스 구축 과정에서 데이터베이스를 구성하는 각 음성 데이터가 음성 인식 성능 평가에 적합한지에 대한 검증 작업이 선행되어야 한다. 그러나 이러한 음성 데이터의 적합성을 판단하기 위해서는 음성 분석 분야 전문가의 자문이 필요하며, 음성 신호에 대한 다양한 분석 자료의 비교가 요구된다. 본 논문에서는 이러한 음성 데이터베이스 구축 문제를 해결하기 위하여 여러 가지 지표들을 조합한 효과적인 음성 성능 지표를 제안한다. 효과적인 음성 성능 지표를 생성하기 위해 다수의 선행 연구가 진행되어 왔다[1-3]. 본 연구에서는 기존의 대표적인 음성 인식 성능 지표인 단어 오인식률(Word Error Rate, WER)과 상관관계(Correlation)가 높은 음성 특성 지표를 분석한 후, 이를 조합하여 새로운 음성 특성 지표를 생성한다. 본 저자의 선행 연구[4, 5]에서는 다양한 특성 지표를 여러 방법으로 조합하여 보다 효과적인 음성 인식 성능 지표를 생성하기 위해 진행 된 실험들을 소개하였다. 본 연구에서는 선행 연구에서 제시하였던 조합 방법을 바탕으로 음성 인식 성능 향상을 위해 심층 신경망 기반의 음성 특성 지표를 추가한 새로운 조합을 제안한다.

본 논문은 다음과 같이 구성된다. II장에서는 음성 인식 성능 실험에 사용된 주요 특성 지표들을 설명하고, III장에서는 선행 연구에서부터 제안해 온 음성 인식 성능 지표 생성 방법에 대해 구체적으로 서술한다. 그리고 IV장에서는 심층 신경망에 대하여 간단히 정의한 후에 본 연구에서 구축한 심층 신경망 모델의 파라미터를 설명한다. V장에서 실험 방법 및 결과를 설명하고 VI장에서 결론을 맺는다.

Ⅱ. 음성 특성 지표

본 장에서는 분석한 여러 가지 음성 특성 지표들 중 본 연구에서 음성 인식 성능 예측 알고리즘 구현을 위해 사용한 음성 특성 지표에 대해 설명한다.

2.1. SNR

SNR(Signal-to-Noise ratio, SNR)[6]은 음질의 성능을 판단할 때 대표적으로 쓰이는 음성 특성 지표로서 주변 잡음 크기에 대하여 듣고자 하는 음성 신호 크기의 상대적인 비율을 표현한 값이다. SNR은 음성 신호의 크기를 잡음 신호의 크기로 나눈 것에 로그를 취한 값으로 SNR이 양수일 경우 잡음 신호 보다 음성 신호의 크기가 큰 경우이다. 따라서 SNR이 큰 음성일수록 음성의 인식 성능이 좋을 것이라 예상할 수 있다. 식 (1)은 SNR의 일반적인 수식으로, P_{signal}은 음성 신호의 크기이며 P_{noise}는 잡음 신호의 크기이다.

$$SNR(dB) = 10\log_{10}(\frac{P_{signal}}{P_{noise}}) \quad (1)$$

2.2. PESQ

사람이 주관적으로 음성을 듣고 1~5 등급 사이로 음성의 품질을 평가하는 방법인 MOS(Mean opinion score)[6] 기법은 주관적인 방법이기 때문에 평가자가 필요하며 같은 음성이라도 사람마다 다른 의견을 표현할 수 있다. 이러한 문제점 때문에 개발된 기법인 PESQ(Perceptual evaluation of speech quality)[6]는 음성 품질 평가를 위해 자동화된 테스트로써 MOS를 자동화 시킨 방법이다. 표 1은 실제 MOS 기법의 평가 기준을 나타낸다.

Table. 1 Standard of mean opinion score(MOS)

Score	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

2.3. GMM 음향 모델 확률 값

가우시안 확률 밀도 함수(Gaussian probability density function, Gaussian pdf)[7]는 평균과 분산으로 확률 분포를 표현하는 모델이다. 가우시안 혼합 모델(Gaussian mixture model, GMM)은 복수 개의 가우시안 확률 밀도 함수를 혼합한 모델이다. 가우시안 혼합 모델은 각 가우시안 요소 별로 가중치를 부여함으로써 식 (2)와 같은 가중 합으로 표현할 수 있다. 식 (2)는 K개의 가우시안 요소를 혼합한 가우시안 혼합 모델에 대한 식으로 w는 가중치, N은 평균 μ, 분산 Σ로 표현되는 단일 가우시안 모델을 뜻한다.

$$p(x) = \sum_{k=1}^{K} w_k N(x : \mu_i, \Sigma_{x,k}) \tag{2}$$

GMM 음향 모델 확률 값은 깨끗한 레퍼런스 데이터베이스로 생성한 GMM 음향 혼합 모델에 대한 우도에 로그를 취한 값으로, 입력된 특징이 훈련된 모델에서 발견될 확률 값이다. 확률인 0~1 사이 값에 로그를 취해서 음수 값으로 나타나게 된다. 입력 된 음성 특징으로 구한 모델 확률 값은 훈련된 이상적인 모델과 유사할수록 큰 값을 갖기 때문에 음성 인식 성능 지표로 쓰일 수 있다. 식 (3)은 GMM 음향 모델 확률 값을 표현한 식으로 평균 μ와 분산 Σ을 갖는 GMM 모델에서 입력된 특징 x의 우도에 로그를 취한 GMM 모델 확률 값의 식이다. 본 실험에서는 깨끗한 음성 데이터베이스를 기준으로 GMM 음향 모델 확률 값을 구하였다.

$$GMMscore = \log P(x_n | g(\mu_n, \Sigma_n)) \tag{3}$$

2.4. MFCC 계수 유사도

MFCC(Mel-frequency cepstral coefficients)[8] 특징 추출 기법은 현재 음성 인식 시스템의 특징 추출 기법으로 가장 널리 사용되고 있는 방법이다. 우선 음성 데이터의 아날로그 신호를 푸리에 변환을 통해 주파수 스펙트럼으로 변환한다. 그 후 인간의 청각 시스템을 모방한 주파수 스케일인 Mel-filter Back 분석을 통해 얻은 계수를 로그를 취한다. 마지막으로 이산 코사인 변환(Discrete Cosine Transform, DCT)을 적용함으로써 Cepstrum 계수로 변환한다. 그림 1은 MFCC 특징 추출 과정을 표현한 다이어그램이다.

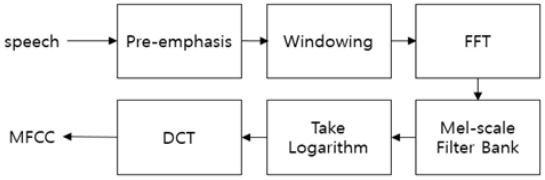

Fig. 1 MFCC feature extraction block diagram

MFCC 계수 유사도[9]는 깨끗한 음성 신호의 MFCC 계수와 오염된 음성 신호의 MFCC 계수의 차이를 측정함으로써 얻을 수 있으며 차이가 작을수록 오염된 음성과 깨끗한 음성 특징이 유사하다고 볼 수 있기 때문에 음성 인식 성능 지표로 쓰일 수 있다. 본 실험에서는 유클리디안 거리 공식을 사용하여 유사도를 구하며 이는 깨끗한 신호와 잡음에 오염된 신호의 MFCC 특징 계수의 거리를 나타낸다. 식 (4)는 MFCC 계수 유사도의 수식이다. f는 오염된 신호의 MFCC 계수이며 f_{clean}은 깨끗한 신호의 계수이다. 본 연구에서는 39차원의 MFCC 특징을 사용한다.

$$MFCC\,distance = \sqrt{\sum_{n=1}^{39} (f_n - f_{clean,n})^2} \tag{4}$$

Ⅲ. 음성 특성 지표를 이용한 새로운 음성 인식 성능 지표 생성 방법

분석한 여러 가지 특성 지표들 중 단어 오인식률과 상관도가 높은 특성 지표들을 채택하고 조합하여 새로운 성능 지표를 제안하였다. 단어 오인식률은 깨끗한 파일의 Transcription과 오염된 파일의 인식 결과를 비교하여 발화한 단어 중 오인식한 단어의 비율을 나타내는 지표이다. 단어 오인식률은 음성 인식의 성능을 직관적으로 표현할 수 있는 대표적인 음성 성능 지표이기 때문에 단어 오인식률과 높은 상관도를 보이는 음성 특성 지표는 음성 성능 지표로 쓰이기 적합하다고 예상하였다. 본 연구에서는 단어 오인식률과 여러 가지 특성 지표들을 대체할 수 있는 효과적인 성능 지표를 생성하기 위해 다양한 음성 특성 지표를 분석하였으며, 그 중 단어 오인식률과 상관도가 높은 GMM 음향 모델 확률 값, SNR, PESQ, MFCC 계수 유사도를 채택하여 새로운 음성 인식 성능 지표를 생성하였다. 새로운 음성 인

식 성능 지표는 식 (5)와 같이 선택된 n개의 음성 특성 지표 중 각 지표 f_i를 0~1의 범위로 정규화하고 단어 오인식률과 상관도에 비례하는 가중치를 적용하여 생성하였다. 아래 식 (6)은 오인식률과의 상관도에 비례하는 가중치 w_i의 계산식을 나타내며, 식 (7)은 새로운 성능 지표 I를 구하는 식이다.

$$f_{norm,i} = \frac{f_i - \min(f)}{\max(f) - \min(f)} \tag{5}$$

$$w_i = \frac{corr(f_i, WER)}{\sum_{j=1}^{n} corr(f_j, WER)} \tag{6}$$

$$I = \sum_{i=1}^{n} (f_{norm,i} * w_i) \tag{7}$$

Ⅳ. 심층 신경망을 이용한 새로운 음성 인식 성능 지표 생성

본 장에서는 심층 신경망을 기반으로 한 음성 특성 지표 추출 방법을 설명하며 이를 위해 기본적인 심층 신경망에 대한 배경 지식을 먼저 소개한다.

4.1. 심층 신경망

심층 신경망(Deep neural network, DNN)[10] 기법은 1940년대에 인간의 신경 세포를 모델링하여 패턴 분류 및 인식에 이용하기 위해 제안된 기법으로 최근 인공지능, 컴퓨터 비전, 음성 인식 등 머신 러닝 분야의 시스템에서 많이 상용되는 기법이다. 그림 2는 실제 인간의 뇌의 신경망을 구성하는 단위인 뉴런이다.

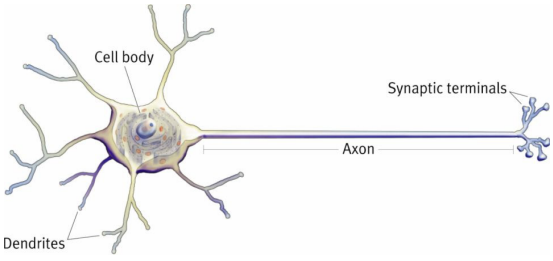

Fig. 2 Structure of a typical neuron[11]

뉴런은 수상돌기(dendrite)로부터 가중이 되는 어떠한 자극을 받아 이 자극이 특정 임계 치를 넘으면 축색돌기(axon)를 통해 다른 뉴런으로 자극을 전달한다. 이러한 뉴런을 수학적으로 모델링 한 것을 퍼셉트론이라 한다. 그림 3은 퍼셉트론의 구조를 나타내며 식(8)은 한 개의 퍼셉트론에 대한 입력과 출력을 표현한다. x_0부터 x_i까지 각 입력마다 가중치를 곱하여 모두 합한 값이 활성 함수 f의 입력 값으로 들어가며 활성 함수의 결과인 0 또는 1이 해당 퍼셉트론의 출력이 된다. 입력 x_0는 바이어스 값이다.

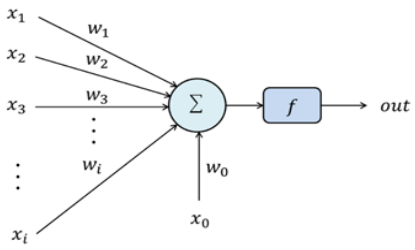

Fig. 3 Structure of a perceptron[11]

$$f(\sum_{n=0}^{i} x_i w_i) = \begin{cases} 1 & \text{if} \quad \sum_{i=0}^{n} x_i w_i > threshold \\ 0 & \text{if} \quad \sum_{i=0}^{n} x_i w_i \le threshold \end{cases} \tag{8}$$

심층 신경망은 수 개의 은닉 층과 각 은닉 층을 구성하는 노드로 구성된다. 각 노드는 퍼셉트론이며 은닉층이 깊어질수록 높은 수준의 추상화 모델을 구축한다. 과거에는 오버 피팅 등의 여러 가지 학습 문제로 인하여 주목받지 못하였으나 이를 해결할 새로운 알고리즘들이 제안되고[10-15] 컴퓨팅 파워와 스토리지 기술이 발전하면서 심층 신경망의 이론이 실제 구현되었다. 그림 4는 심층 신경망의 전체적인 구조를 보여준다.

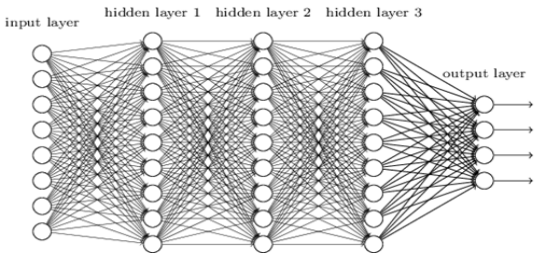

Fig. 4 Structure of a typical deep neural network[14]

심층 신경망 기법은 기존의 패턴 인식 시스템과 달리 특징 추출부터 분류 단, 인식 단이 모두 전체 네트워크 안에 포함된다. 때문에 특징 추출 과정이 필요 없이 층을 거듭할수록 보다 추상적인 특징을 자동으로 추출하고 또 자동으로 구체화하여 인식 과정까지 모두 스스로 진행한다.

4.2. 심층 신경망 기반 음성 특성 지표

3장에서 설명한 음성 특성 지표들을 조합한 지표보다 효과적인 음성 인식 성능 지표를 생성하기 위해 본 연구에서는 심층 신경망을 이용한 음향 모델 확률 값을 제안한다. 제안하는 심층 신경망 모델은 입력으로 음성 프레임 당 39차원의 MFCC 특징을 사용하며 출력은 깨끗한 음성과 잡음에 오염된 음성인 총 두 개의 레이블로 구성된다. 신경망의 은닉 층은 총 4개로, 각 256 – 256 – 128 – 32개의 노드를 갖는다. 그림 5는 제안한 심층 신경망의 구조를 나타낸다.

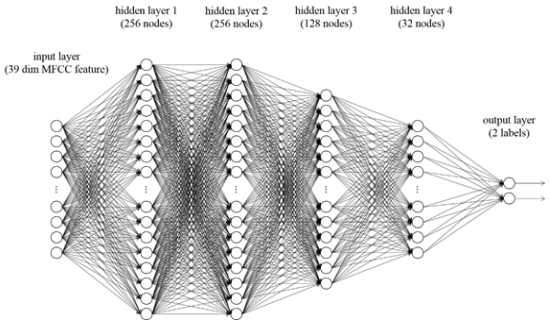

Fig. 5 Proposed structure of deep neural network model

최종으로 계산 되어 출력 되는 값은 심층 신경망 음향 모델의 확률 값이며 각 레이블마다 0부터 1까지 범위의 값으로 산출된다. 두 개의 레이블이 갖는 확률 값 중 조합 지표 실험에는 깨끗한 음향 모델 확률 값인 CLEAN 레이블의 출력 값을 사용한다. 식 (4)는 조합에 보다 효과적으로 사용하기 위해 음향 모델 확률 값 f를 변환하는 수식이다. f는 출력 층에서 출력된 확률 값이며 로그 스케일로 변환하여 조합에 사용된다.

$$f_{\log} = 10\log_{10} f \tag{9}$$

Ⅴ. 실험 및 결과

5.1. 실험 과정

본 실험은 Babble 잡음, 자동차 잡음이 각각 5, 10, 15dB의 SNR로 오염된 TIMIT 음성 데이터베이스[16]를 사용하였다. TIMIT 데이터베이스는 630명의 서로 다른 화자가 발음한 영어 낭독체 문장이 녹음된 6,300개의 음성 샘플로 구성되어 있으며 이는 약 5.6시간 길이의 녹음 분량에 해당된다. TIMIT 음성 데이터베이스 중 훈련에는 깨끗한 음성 신호 파일과 각 잡음 환경 별 10dB의 SNR로 오염된 음성 신호를 포함한 4,620개씩 총 13,860개의 음성 파일을 사용하였으며 1,000epoch 마다 진행되는 검증 과정에 각 잡음 환경 별 5, 10, 15dB로 오염된 음성 신호 1,680개씩 총 11,760개의 음성 파일을 사용하였다. 인식과 평가에 쓰이는 선행 연구와 동일한 720개의 음성 데이터를 사용하였으며 이는 검증 데이터의 일부와 중복된다. 심층 신경망 훈련과 특징 추출은 Linux 환경에서 진행하였으며 BVLC (Berkely Vision and Learning Center)에서 제공하는 심층 신경망 학습 프레임워크인 CAFFE[17, 18]를 사용하였다.

5.2. 실험 결과

표 2는 실험을 위해 채택한 음성 특성 지표들과 단어 오인식률의 상관도를 나타내며, 그림 6은 Babble 잡음 환경에서 생성한 음성 인식 성능 지표와 단어 오인식률의 상관 분포이다. 심층 신경망 음향 모델 확률 값을 포함하여 조합한 음성 인식 성능 지표는 Babble 잡음, 자동차 잡음 환경에서 각각 단어 오인식률과 –0.7838, -0.7758의 상관도를 보인다. GMM 모델 확률 값을 조합하여 지표를 생성한 선행 연구에서는 각각 단어 오인식률과 –0.7766, -0.7691의 상관도를 보였다. 본 실험 결과는 Babble 잡음, 자동차 잡음 환경에서 각 음성 특성 지표를 단독으로 사용할 때보다 단어 오인식률과 높은 상관도를 보이며 선행 연구 결과인 GMM 모델 확률 값을 조합하였을 때보다도 높은 상관도를 보이는 것을 확인할 수 있었다.

Table. 2 Correlation of speech quality measures and WER using GMM and DNN acoustic score

Speech measure	Babble noise	Car noise
SNR	-0.69	-0.69
PESQ	-0.74	-0.74
DNN acoustic score	-0.72	-0.73
MFCC distance	-0.72	-0.71
Combination measure using GMM	-0.7766	-0.7691
Combination measure using DNN	-0.7838	-0.7758

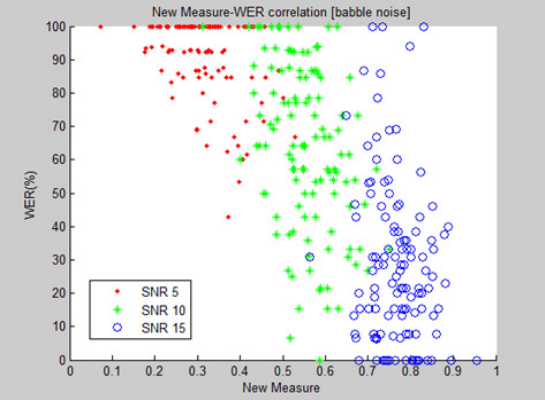

Fig. 6 Correlation of the combination measure and WER in babble noise

Ⅵ. 결 론

본 논문은 음성 인식용 데이터베이스를 평가하기 위해 여러 가지의 음성 특성 지표 추출 알고리즘을 설명하고 새로운 음성 인식 성능 지표 생성 방법을 소개하였다. 선행 연구에서부터 효과적인 음성 인식 성능 지표를 생성하기 위해 대표적인 음성 인식 성능 지표인 단어 오인식률과 상관도가 높은 여러 가지 특성 지표들을 채택, 조합하여 새로운 성능 지표를 제안해왔다. 본 연구에서는 선행 연구에서 제안했던 음성 성능 지표의 인식 성능을 향상시키기 위해 심층 신경망 모델을 사용한 특징 지표를 조합하는 방법을 제안하였다. 조합한 음성 성능 지표는 Babble 잡음, 자동차 잡음 환경에서 모두 각 음성 특성 지표를 단독으로 사용할 때보다 단어 오인식률과 높은 상관도를 나타내었으며 선행 연구에서 제안하였던 조합보다 향상된 상관도를 보여 음성 데이터베이스의 음성 인식 성능을 평가하는데 효과적임을 입증하였다.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2016R1D1A2B03935008).

REFERENCES

[1] S. Yoon, L. Chen, and K. Zechner, “Predicting word accuracy for the automatic speech recognition of non-native speech,” *Interspeech-2010*, pp. 773-776, Jul. 2010.

[2] W. Kim and J. H. L. Hansen, “Phonetic distance based confidence measure,” *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 773-776, Feb. 2010.

[3] H. Park, S. Jee and M. Bae, “Study on the Confidence-Parameter Estimation through Speech Signal,” *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, vol. 6, no. 7, pp. 101-108, Jul. 2016.

[4] S. Ji and W. Kim, “A New Speech Quality Measure for Speech Database Verification System,” *Journal of the Korea Institute of Information and Communication Engineering*, vol. 20, no. 3, pp. 464-470, Mar. 2016.

[5] S. Ji and W. Kim, “Speech Recognition Accuracy Prediction Using Speech Quality Measure,” *Journal of the Korea Institute of Information and Communication Engineering*, vol. 20, no. 3, pp. 471-476, Mar. 2016.

[6] J. R. Deller, J. H. L. Hansen et al., D*iscrete-time processing of speech signals*, Piscataway, NJ: IEEE Press, 1999.

[7] A. L. Garcia, *Probability, Statistics and random processes for electrical engineering*, 3rd ed., Pearson Education, 2008.

[8] Mel frequency cepstral coefficient tutorial. Practical cryptography [Internet]. Available: http://practicalcryptography.com/miscellaneous/machine-learning.

[9] A, S. Thakur, and N. Sahayam, “Speech recognition using euclidean distance,” *International Journal of Emerging*

Technology and Advanced Engineering (IJETAE), vol. 3, no. 3, pp. 587-590, Mar. 2013.

[10] G. Hinton, L. Deng et al., "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, Oct. 2012.

[11] A. K. Jain, J. Mao, K.M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer*, vol. 29, no. 3, pp. 31-44, Mar. 1996.

[12] H. N. Robert, "Theory of the backpropagation neural network," *IEEE International 1989 Joint Conference on Neural Network (IJCNN)*, pp. 593-605, Oct. 1989.

[13] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*, Heidelberg, Dordrecht, London New, York: Springer, pp. 437-478, 2012.

[14] X. L. Zhang, and J. Wu., "Deep neural networks based voice activity detection," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 4, pp. 697-710, Mar. 2013.

[15] M. A. Nielsen, Neural network and deep learning [online]. Available: http://neuralnetworksanddeeplearning.com.

[16] TIMIT database download page. Linguistic Data Consortium [Internet]. Available: http://www.ldc.upenn.edu.

[17] CAFFE deep neural network framework download page. Berkeley Vision and Learning Center [Internet]. Available: http://github.com/BVLC/caffe.

[18] CAFFE deep neural network framework tutorial page. Berkeley Vision and Learning Center [Internet]. Available: http://caffe.berkeleyvision.org.

지승은(Seung-Eun Ji)

2015년 인천대학교 컴퓨터공학부 공학사
2017년 인천대학교 컴퓨터공학부 공학석사
※관심분야 : 패턴인식, 음성인식, 휴먼 컴퓨터 인터페이스

김우일(Wooil Kim)

2003년 고려대학교 전자공학과 공학박사
2004년 ~ 2005년 미국 카네기 멜론 대학교 박사 후 연구원
2005년 ~ 2012년 미국 텍사스 주립대 (University of Texas at Dallas) 연구원 및 연구교수
2012년 ~ 현재 인천대학교 컴퓨터공학부 조교수, 부교수
※관심분야 : 신호처리, 패턴인식, 음성인식, 휴먼 컴퓨터 인터페이스