JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 28 No. 1, pp. 87-92, January 2023
https://doi.org/10.9708/jksci.2023.28.01.087

Noise Elimination Using Improved MFCC and Gaussian Noise Deviation Estimation

Sang-Yeob Oh*

*Professor, Dept. of Computer Engineering, Gachon University, Gyunggido, Korea

[Abstract]

With the continuous development of the speech recognition system, the recognition rate for speech has developed rapidly, but it has a disadvantage in that it cannot accurately recognize the voice due to the noise generated by mixing various voices with the noise in the use environment. In order to increase the vocabulary recognition rate when processing speech with environmental noise, noise must be removed. Even in the existing HMM, CHMM, GMM, and DNN applied with AI models, unexpected noise occurs or quantization noise is basically added to the digital signal. When this happens, the source signal is altered or corrupted, which lowers the recognition rate. To solve this problem, each voice In order to efficiently extract the features of the speech signal for the frame, the MFCC was improved and processed. To remove the noise from the speech signal, the noise removal method using the Gaussian model applied noise deviation estimation was improved and applied. The performance evaluation of the proposed model was processed using a cross-correlation coefficient to evaluate the accuracy of speech. As a result of evaluating the recognition rate of the proposed method, it was confirmed that the difference in the average value of the correlation coefficient was improved by 0.53 dB.

▸ **Key words**: Voice recognition, noise, MFCC, gaussian model, feature extraction

[요 약]

음성 인식 시스템의 지속적인 발전으로 음성에 대한 인식율은 급속도로 발전되었지만 사용 환경에서의 잡음과 여러 음성이 혼합되어 발생하는 잡음으로 정확한 음성을 인식할 수 없는 단점을 가진다. 환경 잡음이 있는 음성을 처리할 때 음성 인식률을 높이기 위해서는 잡음을 제거해야 하며, 기존의 HMM, CHMM, GMM, 그리고 AI 모델이 적용된 DNN에서도 예상치 못한 잡음이 발생하거나 기본적으로 디지털 신호에 양자화 잡음이 추가되면 소스 신호가 변경되거나 손상되어 인식률이 저하된다. 이를 해결하기 위해 각 음성 프레임에 대한 음성 신호의 특징을 효율적으로 추출하기 위해 MFCC를 개선하여 처리하였으며, 음성 신호에 대한 잡음을 제거하기 위해 가우시안 모델을 적용한 잡음 편차 추정을 이용한 잡음 제거 방법을 개선하여 적용하였다. 제안된 모델에 대한 성능 평가는 음성에 대한 정확성 평가를 위해 교차 상관 계수를 사용하여 처리하였으며, 제안하는 방법의 인식률을 평가한 결과 이들에 대한 상관 계수에 대한 평균값 차이는 0.53 dB 개선된 것을 확인하였다.

▸ **주제어**: 음성 인식, 잡음, MFCC, 가우시안 모델, 특징 추출

• First Author: Sang-Yeob Oh, Corresponding Author: Sang-Yeob Oh
*Sang-Yeob Oh (syoh@gachon.ac.kr), Dept. of Computer Engineering, Gachon University
• Received: 2022. 07. 04, Revised: 2022. 10. 20, Accepted: 2022. 10. 20.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

음성 인식 시스템은 AI와 모바일 하드웨어의 발전으로 무선 IP(Internet Protocol) 네트워크에서 다양하게 음성 인식이 지원되고 있으며, 이를 이용하기 위한 다양한 소프트웨어 기술도 함께 보급되어 사용되고 있다. 음성 인식에서 기본적으로 처리되는 음성 신호에 대해 발생되는 잡음 신호를 제거하여 원래의 음성 신호를 가지고 정확하게 음성 인식을 해야 한다[1-7, 14]. 잡음 신호는 기존 음성 신호에 부가되어 유사한 음성 신호 특징을 가질 수 있으며, 기존 음성 신호에 대해 시간이 지나면서 음성 신호에 대한 특성이 변화하여 잡음이 발생할 수 있는 불안정한(unstable) 잡음이 생성되고, 음성 신호의 잡음이 정확하게 분류되지 않을 때에는 음성 신호에 대한 잡음이 일정하지 않는 스레숄드(threshold)로 인해 정확도가 떨어지는 음성 인식을 수행하게 된다. 그러므로 환경 잡음이 있는 음성을 처리할 때 음성 인식률을 높이기 위해서는 잡음 제거 기술을 필요로 하며[16], 이 과정에서 모델 추정 기술을 위한 잡음 제거 및 특징 추출이 활용되어야 한다. 이와 같은 잡음 제거 및 모델 추정 기술에서 가장 중요한 부분은 음성 신호에 대한 부가적인 혼합 잡음을 추정하고 제거하는 것이다. 음성 인식 시스템에서 신호에 예상치 못한 잡음이 나타나거나 기본적으로 디지털 신호에 양자화 잡음이 추가되면 소스 신호가 변경되거나 손상되어 인식률이 저하된다. 소스 신호가 다양한 종류의 잡음과 혼합되어 변환되거나 변경되는 경우 효과적인 잡음 제거를 위해 HMM(Hidden Markov Model), GMMGaussian mixture model), DNN(Deep Neural Network) 등이 사용되었으나 예상치 못한 잡음이 발생하거나 기본적으로 디지털 신호에 양자화 잡음이 추가되면 소스 신호가 변경되거나 손상되어 인식률이 저하된다[17-18]. 또한, 인간의 청취 영역을 처리하기 위한 MFCC는 음성에 대한 필터와 잡음을 모두 고려한 방법이며, 음성 인식을 위한 기본 주파수를 mel-scale에 적용한 필터 뱅크(filter bank)를 이용하여 음성 처리를 MFCC가 지원하며, 이 MFCC에서 특정 음성에 대한 잡음을 분석하여 처리하기 위한 연구가 진행되고 있다.

본 논문에서 제안된 방법은 음성 인식에 대한 인식을 높이기 위해 음성에 대해 잡음이 부가되어 입력되는 음성에서 음성에 대한 특징을 기반으로 추출하기 위해 개선된 음성 특징 추출 기법을 제안한다. 음성에 대한 특징 추출을 수행하여 음성에 대한 추출을 효율적으로 수행하기 위해 인간의 청각적 특성인 mel-scale을 반영하여 잡음에 강한 특징을 추출하는 방법인 MFCC(Mel Frequency Cepstral Coefficient)를 개선하여 특징 추출 과정에서는 불필요한 중복 신호 정보를 없애고 동일 신호들 간의 일관성을 높여 다른 신호와 변별력을 높일 수 있는 정보를 추출하며, MFCC는 이러한 정보를 최대한 잘 나타낼 수 있는 파라미터를 특징파라미터로 사용하여 잡음에 강한 특성을 유도한다. 본 논문에서는 MFCC에 mel-scale filter bank를 적용하고, 주파수 인지 특성에 대한 주파수에 log를 수행하여 음성에 대한 출력, 필터, 캡스트럽의 차수를 가지고 음성에 대한 특징을 효율적으로 개선한 방법을 적용하였다.

또한, 가우시안 인식 모델을 이용한 확률 밀도 함수와 혼합 가중치를 사용하여 잡음 추정을 위한 확률 분포를 사용하여 잡음을 제거하여 잡음에 강인한 음성 인식 모델을 지원하며, 개선된 음성에 대한 잡음 제거를 위해 음성 신호에 대한 음성 잡음 신호 대역의 분산 제곱과 평균을 적용하여 기존 방법을 개선하였다.

제안된 방법을 Aurora 2.0 데이터베이스를 이용하여 분석 및 실험하였으며, 개선된 선택적 음성 특징 추출 실험과 잡음 편차 추정을 이용한 잡음제거에 대한 성능 평가 측정을 위해 서울 시내 지역명 20개, 지하철역명 20개로 구성하였으며, 제안된 모델에 대한 성능 평가에서 음성에 대한 정확성 평가를 위해 교차 상관 계수를 사용하여 처리된 결과를 제안한 방법과 비교하여 인식율을 평가한 결과 이들에 대한 상관계수 평균값 차이에서 0.53 dB 개선된 것을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 언급하고 3장에서는 개선된 음성 특징 추출 실험과 개선된 거우시안 잡음 편차 추정을 이용한 잡음제거에 대한방법에 대해 설명하며, 4장에서는 시스템에 대한 분석과 평가를 통해 제안된 방법을 평가하고 마지막으로 5장에서 결론을 맺는다.

II. Preliminaries

1. Related works

1.1 Voice feature extraction and noise

음성을 위한 특징으로 mel-cepstrum을 사용하며[2], 이는 음성에서 mel-scale로 존재하는 다수 개의 band-pass 필터로 사용되는 필터 뱅크에 통과시켜 각 band에 대한 에너지를 구한다[11]. 음성 신호에서 잡음을 없애기 위해 잡음을 최소화 한 특징을 추출하는 방법으로 음성 인식에 대한 결과를 높인다. 잡음 발생은 여러 환경 요인에 의해 발생하며[8-10], 이와 같은 잡음을 처리하기

위해서 모델링되거나 학습된 임펄스 응답 필터를 사용하며, 이 필터는 시간의 변화에 의해 발생되는 잡음을 제거해 주는 역할을 한다. 음성 인식 시스템을 사용하는 환경에서 음성에 부가된 잡음에 대해 디지털로 변환하여 처리되므로 임펄스 응답을 가지고 수행되는 선형 시변 필터를 가지고 모델링을 수행하며, 적응 필터에 대한 음성 입력 신호 $s(n)$과 임펄스 응답 $f(n,l)$를 가지고, 출력 $y(n)$을 나타내며, 선형 시변 필터에 대한 임펄스 응답 $f(l)$은 지수함수로 처리되어 $f(n,l)=0$, $(l \geq n)$로 표현된다.

$$y(n)=\sum_{l=0}^{+\infty} f(n,l)\,s(n-1)+noise(n) \qquad (1)$$

1.2 Gaussian model

음성 인식에서 사용되는 가우시안 최적화 모델은 음성에 대한 표본 데이터 집합에 대해 사용되는 분포 밀도를 사용하여 확률밀도함수로 모델링하는 방법으로 음성 데이터의 분포를 모델링한다[15]. 이 방법은 단일 가우시안이 적용되지 않는 복수개의 중심점을 사용하는 1차원 데이터와 2차원 데이터를 가지고 처리한다.

확률밀도함수는 가우시안 분포외에 다른 분포에도 적용될 수 있으며, 가우시안 최적화 밀도는 확률밀도함수를 가우시안 분포로 처리하므로 전체 확률밀도함수는 M개의 가우시안 확률밀도함수의 순차저인 구성으로 식(2)와 같이 표현된다.

$$p(x|\Theta)=\sum_{i=1}^{M} p(x|\omega_i,\Theta_i)P(\omega_i) \qquad (2)$$

$p(x|\omega_i,\Theta_i)$는 데이터 x의 ω_i번째 성분 파라미터 Θ_i로 구성된 확률밀도함수이고, $P(\omega_i)$는 혼합 가중치를 나타낸다. 혼합 가중치는 사전확률 α_i라고 하면 다음 식과 같은 제약조건을 가진다.

$$0 \leq \alpha_i \leq 1, \text{ and } \sum_{i=1}^{M} \alpha_i = 1 \qquad (3)$$

확률밀도함수에 가우시안 분포를 적용하는 경우 Θ_i는 다음과 같은 파라미터 집합을 가진다.

$$\Theta_i=(\mu_1,\mu_2,\cdots,\mu_M,\sigma_1,\sigma_2,\cdots,\sigma_M,\alpha_1,\alpha_2,\cdots,\alpha_M,) \qquad (4)$$

가우시안 최적화 모델로 음성 데에 일반적으로 음성 데이터의 분포를 모델링하는 경우에 대한 혼합 성분에 대한 데이터를 충분히 사용하고, 파리미터 값을 적절히 사용하면 음성 데이터에 대한 연속적인 분포를 모델링한다.

III. The Proposed Scheme

음성 특징 추출은 음성에 대한 중복 신호 정보를 제거하고, 동일 신호들 간의 일관성을 높여 다른 신호와 변별력을 높일 수 있는 정보를 추출한다[11-13]. 특징 추출 과정에서는 불필요한 중복 신호 정보를 없애고 동일 신호들 간의 일관성을 높여 다른 신호와 변별력을 높일 수 있는 정보를 추출한다. MFCC는 이러한 정보를 최대한 잘 나타낼 수 있는 파라미터를 특징파라미터로 사용한다. MFCC로 추출한 특징의 차수는 12이며, delta cepstrum을 결합하여 총 13차 특징을 사용한다[11]. MFCC는 인간의 청각적 특성인 mel-scale을 반영하여 잡음에 강한 특징을 추출하는 방법이며, 본 논문에서 제안하는 방법은 다음 그림과 같은 단계로 처리하여 잡음에 강한 특징을 추출한다.

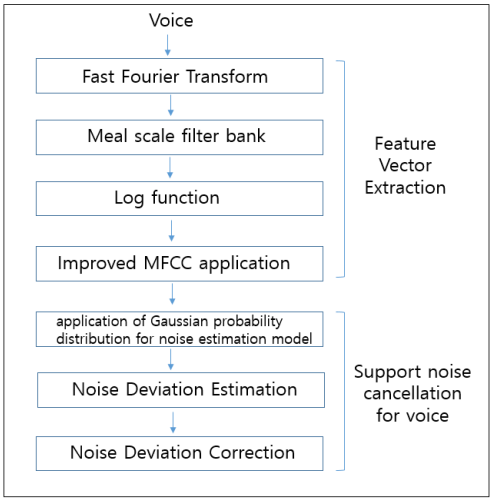

Fig. 1. Proposed MFCC feature Extraction Process and gaussian noise cancellation

환경 잡음 데이터의 효율적 처리를 위해 본 논문에서는 MFCC에서 이러한 정보를 최대한 잘 나타낼 수 있는 파라미터를 특징에 대한 차수로 사용하며, MFCC 특징 추출 과정을 위해 음성 신호에 대한 FFT(Fast Fourier Transform)을 적용하고, FFT 분석에서의 power spectrum은 다음 수식에서 디지털 신호 $x(k)$에 대해 12개를 사용하고, n개의 구간을 가지고 다음과 같이 적용한다.

$$X(k)=\sum_{n=0}^{N-1}x(k)\,Wn^{kn} \quad (5)$$

환경 잡음 데이터는 전처리를 거친 후 FFT을 적용하고 일반적인 주파수를 mel-scale의 주파수로 바꾸는 역할을 하며, 그림 2와 같은 mel-scale filter bank를 통과시킨다.

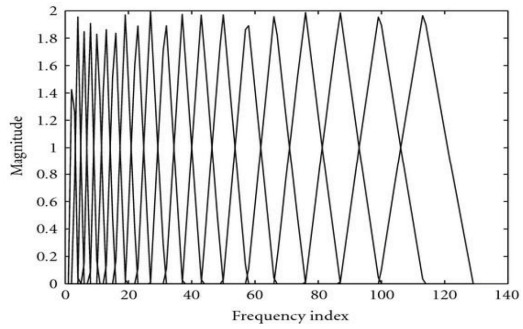

Fig. 2. Meal-scale filter bank

이 식을 mel-scale 필터 뱅크를 사용하여 다음 수식과 같이 표현되며, 입력 주파수를 f에 대한 log 값을 다음과 같이 적용하여 log-spectrum을 구성하여 최종 mel 값을 구한다.

$$mel=2595\cdot\ \log_{10}(1+\frac{f}{700}) \quad (6)$$

이 결과를 가지고, 주파수에 대한 역변환 작업을 수행하고, 주파수 인지 특성에 대한 주파수를 가지고 log를 수행하여 음성에 대한 출력, 필터, cepstrum의 차수를 가지고 음성에 대한 특징을 추출한다.

$$MFCC_m=\sqrt{\frac{2}{n}}\sum_{i=1}^{N}\left\{\log(x[i])\cos\left(\frac{2\pi m}{N}\left(i-\frac{1}{2}\right)\right)\right\} \quad (7)$$

$x[i]$는 mel-scale 필터 뱅크에 대한 출력이며, N은 필터 뱅크에서 사용되는 대역 통과 필터 개수를 나타내며 m은 cepstrum의 차수를 나타내며, 이 과정을 통해 잡음의 영향이 줄어든 특징을 추출하여 음성 시스템의 성능을 높인다.

음성에 대한 잡음 제거를 지원하기 위해 잡음 추정 모델을 위한 가우시안 확률 분포를 다음과 같이 적용한다. 확률 밀도 함수는 M개의 가우시안 확률 밀도 함수의 선형 결합으로 구성되는 좌우 대칭 분포를 가진다.

$$P_x(x)=C\cdot\ \exp\left(-\left|\frac{x}{\alpha}\right|\right)^{\beta} \quad (8)$$

이 식에서 α는 모수로 사용되며, 음성 변환 시에 잡음 신호에 대한 수식은 다음과 같으며, 잡음 편차 추정은 잡음 신호를 log $normal$ 분포인 ln으로 가정하여 추정한다.

$$\sigma_x^2=\ln\left(1+\frac{\sigma_y^2}{\mu_y^2}\right) \quad (9)$$

σ_y^2는 음성 잡음 신호 대역의 분산 제곱이며, μ_y^2은 잡음 신호 대역의 평균의 제곱이다. 이 식을 가지고 잡음으로 추정되는 잡음을 제거하기 위해 적용하고, 잡음 편차를 보정하기 위해 가중 편차를 적용하여 개선한다.

$$\sigma=\omega\sigma_x \quad (10)$$

IV. Simulations

본 연구에서는 유사 음소 인식 처리와 효율적 특징 추출을 이용한 음성 인식 모델 최적화 방법 실험 수행을 하기 위해 Aurora 2 데이터베이스를 사용하였다. Aurora 2 데이터베이스는 음성신호에 인공적으로 부가잡음을 더해주고 채널 왜곡에 대한 잡음 음성으로 구성되어 있으며 국제적으로 공인되어 가장 많이 사용되는 음성데이터 중의 하나이다. Aurora 2.0은 잡음 환경과 잡음 레벨을 가지는 데이터 집합을 포함하여 백색 가우시안 잡음과 혼합 잡음(babble noise)에 대한 음성 신호를 향상하기 위한 성능 검증으로 사용된다. 음성 신호는 잡음 신호의 실시간 처리 지원을 위해 8kHz sampling rate, 16bit를 사용하였으며 FFT 크기는 256 샘플, 음성 단락 현상을 제거하기 위해 1/2 오버래핑(overlapping) 구간을 이용하고 신호 왜곡을 줄이기 위해 해밍 원도우(Hamming Window)를 사용하였다. 앞의 3장에서 설명한 MFCC를 개선한 방법과 음성에 대한 잡음 제거를 위해 가우시안 분포의 음성 잡음 신호 대역의 분산 제곱과 평균을 적용하여 개선한 방법에 대해 채널 간의 유사성을 판단하기 위해 교차 상관 계수 방법과 제안한 방법을 비교하였다. 음성 신호에 대한 비교 분석을 위해 음성에 대한 잡음 처리는 워너 필터를 이용하였으며, 음성 인식 실험을 위해 서울 시내 지역명 20개, 지하철역명 20개로 구성하였다. 인식 실험에서는 실험에 참가한 화자가 음성 인식 목록을 5회 발음하여 총 100단어를 적용하여 실험을 수행하였으며, 음성에 대한 각 채널 별 상관계수의 연산 과정을 수행한 결과 데이터를 각각 비교하였다. 음성에 대한 정확성 평가를 위해 각 음성에 대한 교차 상관 계수를 사용하여 처리된 결과를 제안한 방법과 비교

하여 다음 Table 1에 나타내었으며, 음성에 대한 평가는 SDR(Signal to Distotion Ratio)[16]을 이용하여 수행하였으며, 이에 대한 식은 다음과 같다.

$$SDR(dB) = 10\log_{10}\frac{\sum_{n=1}^{N}[x(n)-\hat{x}(n)]^2}{\sum_{n=1}^{N}x^2(n)} \quad (11)$$

이 식에서 $x(n)$는 잡음이 반영되지 않은 음성을 나타내고 , $\hat{x}(n)$는 잡음이 부가된 음성 신호를 의미하며, n은 시간 인덱이다. 식 (11)을 이용한 왜곡도를 나타내었다.

Table 1. Compare with cross correlation coefficient and proposed method and distortion ratio

Separate Voice	Cross correlation	Proposed method	Distortion Ratio
1	2.21	2.13	0.21
2	2.31	1.51	0.73
3	5.66	1.99	2.87
4	6.14	1.67	3.61
5	1.68	1.79	-0.31
6	2.27	2.38	-0.67
7	1.35	1.83	0.27
8	7.01	6.14	0.67
9	2.19	2.16	-0.37
10	2.36	2.60	-0.23
11	1.51	1.59	-0.17
12	2.47	2.29	0.14
13	2.51	2.50	-0.03
14	2.31	2.14	-0.06
15	2.27	2.11	0.11
16	3.01	2.79	-0.06
17	2.31	2.11	0.27
18	3.10	2.77	0.11
19	2.07	2.00	0.05
20	2.37	2.10	-0.03
Average	2.86	2.33	0.36

왜곡도에 대한 값이 작으면 비교되는 음성의 유사도가 높은 것을 의미하며, Table 1에서 음성 신호에 대한 상관계수와 제안한 방법을 적용한 방법의 차이가 높은 3번째, 4번째 음성 신호에서 높지만 제안 방법은 상호 연관성이 고르게 나타내고 있으며, 8번째 음성 신호만 높게 나타난 것을 확인할 수 있으며, 제안하는 방법의 인식률을 평가한 결과 이들에 대한 상관 계수에 대한 평균값 차이는 0.53 dB 개선된 것을 확인하였다. 또한, 기존의 MFCC 방법과 본 논문의 개선된 MFCC 방법을 비교하기 위해 왜곡도를 비교하여 성능 평가를 실시하였으며, 성능 평가를 위해 기존의 MFCC 방법과 제안된 MFCC 개선 방법의 잡음 제거를 수행하여 측정한 결과를 Table 2에 제시하였으며, 왜곡도에서 성능이 향상됨을 확인할 수 있었다.

Table 2. Distortion Ratio Evaluation of Noise Removal

Separate Voice	Distortion Ratio of MFCC	Distortion Ratio of proposed method
1	0.31	0.21
2	0.85	0.73
3	3.37	2.87
4	3.97	3.61
5	-0.28	-0.31
6	-0.67	-0.67
7	0.47	0.27
8	0.81	0.67
9	-0.31	-0.37
10	-0.28	-0.23
11	-0.19	-0.17
12	0.29	0.14
13	-0.08	-0.03
14	-1.14	-0.06
15	0.17	0.11
16	-0.11	-0.06
17	0.39	0.27
18	0.25	0.11
19	0.14	0.05
20	-0.08	-0.03
Average	0.39	0.36

V. Conclusions

본 논문에서는 음성 특징 추출과 잡음 편차 추정을 이용한 잡음 제거 방법을 사용하여 음성 인식 모델에 대한 성능을 평가하였다. 여러 환경에 대한 잡음을 가지는 현실에서의 잡음 환경에서 SNR이 적은 음성에 대해서는 잡음 신호에 영향이 크기 때문에 음성 검출의 성능이 낮아진다.

각 음성 프레임에 대한 음성 신호의 특징을 효율적으로 추출하기 위해 MFCC를 이용하여 처리하였으며, 음성 신호에 대한 잡음을 제거하기 위해 가우시안 모델을 적용한 잡음 편차 추정을 이용한 잡음 제거 방법을 적용하였다. 개선된 선택적 음성 특징 추출 실험과 잡음 편차 추정을 이용한 잡음제거에 대한 성능 평가 측정을 위해 제안된 방법을 Aurora 2.0 데이터베이스를 이용하여 분석 및 실험하였으며, 성능 평가 측정을 위해 서울 시내 지역명 20개, 지하철역명 20개로 구성하였으며, 제안된 모델에 대한 성능 평가에서 음성에 대한 정확성 평가를 위해 교차 상관계수를 사용하여 처리된 결과를 제안한 방법과 비교하여 인식율을 평가한 결과 이들에 대한 상관계수에 대한 평균값 차이는 0.53 dB 개선된 것을 확인하였다.

REFERENCES

[1] Sang-Yeob Oh, “Speech Recognition Performance Improvement using a convergence of GMM Phoneme Unit parameter and Vocabulary Clustering”, Journal of Convergence for Information Technology, 10(8), 35-39. 2020,DOI : https://doi.org/10.22156/CS4SMB.2020.10.08.035

[2] Sang-Yeob Oh, “DNN based Robust Speech Feature Extraction and Signal Noise Removal Method Using Improved Average Prediction LMS Filter for Speech Recognition”, Journal of Convergence for Information Technology, 10(8), 35-39. 2021, DOI : https://doi.org/ 10.22156/CS4SMB.2021.11.06.001

[3] J. Homer & I. Mareels, “LS detection guided NLMS estimation of sparse system”. Proceedings of the IEEE 2004 International Conference on Acoustic. Speech, and Signal Processing(ICASSP). Montreal, Quebec, Canada, 2004, DOI: 10.1109/ICASSP.2004.1326394

[4] Berrak Sisman, Junichi Yamagishi, Simon King, and Haizhou Li, “An overview of voice conversion and its challenges: From statistical modeling to deep learning”, IEEE/ACM Transactionson Audio, Speech and Language Processing, 2020, DOI: 10.1109/TASLP.2020.3038524

[5] Wu, B. F., Wang, K. C, “Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments”, IEEE Transactions on Speech and Audio Processing, 13(5), 762-775, 2005, DOI: 10.1109/TSA.2005.851909

[6] Li, Q., Zheng, J., Tsai, A., Zhou, Q, “Robust endpoint detection and energy normalization for real-time speech and speaker recognition”, IEEE Transactions on Speech and Audio Processing, 10(3), 146-157. 2002, DOI: 10.1109/TSA.2002.1001979

[7] Arango, A. P´erez, J. and Poblete, B, “Hate Speech Detection is Not as Easy as You May Think”, A Closer Look at Model Validation. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 45–54. Paris, France: Association for Computing Machinery. 2019, https://doi.org/10.1145/3331184.3331262

[8] Aluru, S. S, Mathew, B., Saha, P. and Mukherjee, A. “Deep Learning Models for Multilingual Hate Speech Detection”, arXiv preprint arXiv:2004.06465, 2020, https://doi.org/10.48550/arXiv.2004.06465

[9] ETSI standard document, “Speech Processing, Transmission and Quality aspects(STQ); Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms”, ETSI ES 202 050 v.1.1.3. 2003

[10] Scart, P., Filho, J. “Speech enhancement based on a priori signal to noise estimation”, Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 629-632. 2002, DOI: 10.1109/ICASSP.1996.543199

[11] K. Chung & S. Y. Oh. “Vocabulary optimization process using similar phoneme recognition and feature extraction”, Cluster Computing, 19(3), 1683-1690. 2016, DOI 10.1007/s10586-016-0619-0

[12] Kamarth, S., Loizou, P. “A multi-Band spectral subtraction method for enhancing speech corrupted by colored noise,” Proc. IEEE Int, Conf, Acoustic Speech Signal Process., 101-111. 2002, DOI: 10.1109/ICASSP.2002.5745591

[13] Yi Hu and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement“, IEEE Trans. ASLP, 16, 229–238. 2008, DOI: 10.1109/TASL.2007.911054

[14] G. Hu, D. Wang. “A Tandem Algorithm for Pitch Estimation and Voiced Speech Segregation”, IEEE Trans, Audio, Speech and Language Processing. Vol. 18, No. 8, pp. 2067-2079. 2010, DOI: 10.1109/TASL.2010.2041110

[15] S. Y. Cho, D. M. Sun, Z. D. Qiu. “A Spearman correlation coefficient ranking for matching-score fusion on speaker recognition”, Proc. TENCON Conf. pp. 736-741. 2011, DOI: 10.1109/TENCON.2010.5686608

[16] Sang-Yeob Oh “Selective Speech Feature Extraction Using Channel Similarity in CHMM Vocabulary Recognition”, Journal of Digital Policy & Management, Vol. 11, No. 10, pp. 453-458. 2013, https://doi.org/10.14400/JDPM.2013.11.10.453

[17] A. Srinivasan. “Speech Recognition Using Hidden Markov Model”, Applied Mathematical Sciences, vol.5, no.79, pp. 3943-3948, 2011. 2011

[18] S. Rangachari, P. C. Loizou. “A noise estimation algorithm for highly non-stationary environments”, Speech Communication, Vol. 48, No. 2, pp. 220-231. 2006, https://doi.org/10.1016/j.specom.2005.08.005

Authors

Sang-Yeob Oh received the B.S. degrees in Computer engineering from Gachon University, M.S. and Ph.D. degrees in Computer engineering from KwangWoon University, Korea.

Dr. Oh joined the faculty of the Department of Computer Engineering at Gachon University in 1992. He is currently a Professor in the Department of Computer Engineering, Gachon University. He is interested in voice recognition, noise detect, and voice feature extraction.