

프레임 음성구간에 대한 잡음음성의 진폭 스펙트럼의 추정 알고리즘

최재승*

An Algorithm for Estimating the Amplitude Spectrum of Noisy Speech for the Frame Speech Section

Jae-Seung Choi*

요약

최근 음성인식 성능을 개선하기 위하여 높은 잡음환경 하에서 음성구간을 정확하게 검출할 방법과 주파수 영역에서 스펙트럼을 고려한 방법을 실현할 필요가 있다. 따라서 본 논문에서는 잡음의 크기가 변화하는 경우에 그 잡음의 스펙트럼 변화를 추정할 수 있는 최소 2승오차 단시간 진폭스펙트럼 추정 방법을 제안한다. 본 기법은 높은 잡음 환경 하에서 각 프레임 구간에서 음성활동 구간 검출의 한계치를 적용하여, 각 프레임 구간에서 잡음의 변화에 대해서 적절한 잡음 스펙트럼을 추정하는 알고리즘이다. 실험에서는 두 종류의 음성 데이터베이스와 3종류의 유색잡음과 백색잡음을 사용하였으며, 특히 백색잡음에 대해서 입력 잡음이 증가할수록 제안한 알고리즘의 출력 SNR이 최대 9.13(dB)까지 개선되어 본 알고리즘의 우수함을 알 수 있었다.

Abstract

Recently, in order to improve speech recognition performance, it is necessary to implement a method to accurately detect speech sections in high noise environments and a method that considers the spectrum in the frequency domain. Therefore, this paper proposes a minimum square error short-time amplitude spectrum estimation technique that can estimate the spectral change of noise when the noise level changes. This method is an algorithm that estimates an appropriate noise spectrum for changes in noise in each frame by applying a threshold for voice activity detection under a high noise environment. In the experiment, two types of speech databases, three types of colored noise and white noise were used. In particular, it was found that the proposed algorithm is superior because the output SNRs of the proposed algorithm improved up to 9.13 (dB) for white noise as the input noise increases.

Keywords

speech recognition, frequency domain, high noise environments, mmse-stsa, voice activity detection

* 신라대학교 전기전자공학과 교수
- ORCID: <https://orcid.org/0000-0002-5699-9701>

· Received: Jun. 09, 2025, Revised: Jul. 17, 2025, Accepted: Jul. 20, 2025
· Corresponding Author: Jae-Seung Choi
Dept. of Electrical and Electronic Engineering, Silla University, 140
Baegyang-daero(Blvd), 700beon-gil(Rd), Sasang-gu, Busan, 46958 Korea
Tel.: +82-51-999-5608, Email: choijaes7@silla.ac.kr

I. 서 론

근래에 음성인식 등의 처리를 음성과형에 의한 신호처리 작업 및 아날로그 필터뱅크에 의한 음성의 분석 등의 다양한 방법으로 계산기에서 하드웨어를 사용한 연구가 진행된다. 또한 사람이 발성한 음성을 하드웨어 및 컴퓨터 등의 계산기에서 정확하게 인식시키는 음성인식의 연구는 오래 전부터 진행되어 왔으며 또한 하드웨어에 의한 음성인식 장치가 개발된다[1].

실제 환경에서와 같은 자동차, 열차, 항공기 등의 교통소음 등의 고소음의 환경이 증대된다. 소음의 레벨이 높거나 장시간 노출되게 되면 심리적 고통뿐만 아니라 사회적으로도 중대한 문제가 된다. 또한 고소음의 환경은 음성에 의한 통신을 방해하기 때문에 이러한 문제를 해결하기 위해서는 소음의 레벨을 경감, 소음원의 제어 및 차단 등의 연구가 활발히 진행되고 있으며, 잡음에 의한 음성통신의 방해, 음성의 명료성 및 명료도와 소음의 관계에 대해서도 연구 결과가 보고된다[2].

음성인식기술의 진전과 함께 특히 높은 잡음환경에서 음성구간을 정확하게 검출하여 음성인식의 성능을 향상시키는 것이 중요하다[3]. 특히 고소음 환경하에서는 음성구간을 정확하게 검출하는 것이 어렵기 때문에 음성구간을 오검출하게 되면 불필요한 무음구간에 대해서 인식을 진행하게 되기 때문에 오차가 발생하여, 음성인식 성능의 저하와 직결되는 문제가 발생하게 된다. 이러한 음성구간 검출은 소음환경 하에서 음성인식을 실행하는 경우가 많으며 특히 고소음 환경 하에 있어서 성능의 저하가 현저하기 때문에 최근에는 스펙트럼의 형상을 고려하는 방법의 연구가 진행되고 있으며, 심층신경회로망(DNN, Deep Neural Network)에 의한 음성의 구간검출의 유효성을 실현하는 연구도 실현된다[4].

잡음 억압에 있어서 음질의 열화와 잡음 억압에 의해서 새롭게 발생하는 음악 잡음을 얼마나 억제할 수 있는 점이 중요한 연구 과제이다. 따라서 음악 잡음을 억제하기 위해서는 그 주파수성분이 주로 고역에 있다는 것을 이용하여 고역 억압을 중점적으로 실행하게 된다. 다음으로 잡음을 충분히 억압시키기 위해서 음성에 왜곡이 발생하여 음질이

열화하므로 이 열화를 방지하기 위해서는 원음인 음성과 잡음 성분을 어느 정도 비율로 부가하는 방법을 검토하고 있다[5].

논문에서는 잡음의 종류 및 크기가 변화할 때 그 변화에 추정 가능한 잡음 스펙트럼추정의 제안법인 최소 2승오차 단시간 진폭스펙트럼 추정(MMSE-STSA, Minimum Mean-Square Error-Short-Time Spectral Amplitude Estimator)[6][7]법을 제안한다. 제안법에서는 먼저 고소음 환경하에서 각 프레임 구간에서 음성활동 구간 검출(VAD, Voice Activity Detector)[8][9]법의 한계치를 사용하여, 각 프레임의 스펙트럼 전력을 비교함으로써 음성구간 및 잡음구간을 분류한다. 이 후에 각각의 프레임 구간에 있어서 적절한 잡음 스펙트럼추정법을 사용하여 잡음스펙트럼을 추정한다. 실험에서는 두 종류의 데이터베이스에 포함된 음성신호와 자동차 잡음, 전시장내 잡음, 여러 사람이 떠드는 소리 바블잡음, 백색잡음을 사용하여 잡음의 레벨도 시간적으로 변동시킨다. 제안법을 사용함으로써 잡음의 급격한 변화에도 대응이 가능하며, 신호대 잡음비(SNR, Signal to Noise Ratio)의 개선을 실험으로 확인한다.

II. 관련 연구 및 제안한 알고리즘

이동전화, 음성인식 장치를 잡음이 많은 배경 환경 하에서 사용할 경우에 주변의 잡음신호가 음성신호에 중첩되기 때문에 음성품질의 열화 및 음성의 인식을 저화를 일으킨다. 특히 잡음이 혼입된 음성의 스펙트럼과 잡음 스펙트럼의 비율을 잡음혼입음성에 중첩하여 잡음성분을 억제하는 스펙트럼 억압법(SS, Spectral Subtraction)[10], 음성인식 장치에서 음성에 혼입한 배경잡음을 억제하는 목적의 MMSE-STSA[6][7], 적응디지털필터 방법[11] 등이 제안된다. SS법에서는 SNR의 출력이 개선되지만 잡음 스펙트럼의 과도한 감산으로부터 음악잡음이 발생하는 등의 문제가 지적된다. 한편으로 적응디지털필터에 의한 방법은 음성신호의 기본주기에 대응하여 적응필터를 구성하지만 SNR이 낮은 경우에 있어서 음성신호의 기본주기의 정확한 검출이 요구되는 음성강조를 위한 필터법이다.

SS와 MMSE-STSA의 단일 채널을 이용한 비선형 잡음억압 방법은 강력한 잡음억압성을 가지지만, 비선형처리로 인해 비선형 잔류잡음인 음악잡음의 발생이 문제가 된다.

음성신호는 일반적으로 비정상적이며 그 음향적 특징은 변화하기 때문에 음성의 스펙트럼 분석에서는 짧은 시간의 구간에 있어서 음성은 정상상태라고 가정한다. 다음으로 조금씩 프레임의 시간 구간을 이동하면서 창함수를 사용하여 추출한 프레임의 파형 데이터에 대해서 순차적으로 고속 푸리에 변환(FFT, Fast Fourier Transform)을 실행하여 스펙트럼의 시계열 데이터를 구한다.

제안한 논문에서는 잡음이 중첩된 음성을 M 샘플의 프레임으로 분류하여, $2M$ 샘플의 창함수를 사용하여 오버랩시킴으로써 k 번째의 프레임에 대해서 추출된 잡음이 중첩된 음성 $\hat{x}_k(t)$ 은 식 (1)과 같이 나타낼 수 있다.

$$\hat{x}_k(t) = \begin{cases} h(t)x_{k-1}(t) & , 1 \leq k \leq M \\ h(t)x_k(t-M) & , M \leq k \leq 2M \end{cases} \quad (1)$$

시간영역에서의 잡음이 혼입된 음성 $x(t)$ 를 $y(t) = s(t) + n(t)$ 와 같이 정의한다. 여기에서 잡음신호를 $n(t)$, 깨끗한 음성신호를 $s(t)$ 로 정의한다. 잡음환경에서의 음성신호 출력 $y(t)$ 의 스펙트럼은 $Y(\omega) = X(\omega) + N(\omega)$ 로 표현할 수 있다. 이 신호의 k 번째의 프레임에 있어서 주파수 영역에서의 잡음이 중첩된 음성 $y_k(t)$ 의 스펙트럼을 식 (2)와 같이 나타낼 수 있다. 여기에서 ω 는 주파수를 나타낸다.

$$Y_k(\omega) = X_k(\omega) + N_k(\omega) \quad (2)$$

k 번째 프레임에 있어서 진폭스펙트럼과 이득 함수 $G_k(\omega)$ 를 다시 식 (3)과 같이 정의한다.

$$\begin{aligned} |Y_k(\omega)| &= |X_k(\omega)| + |N_k(\omega)|, \\ |\hat{X}_k(\omega)| &= 1 - \frac{G_k(\omega)}{|Y_k(\omega)|}, \\ G_k(\omega) &= 1 - \frac{|\hat{N}_k(\omega)|}{|Y_k(\omega)|} \end{aligned} \quad (3)$$

음성강조기술은 시간영역의 처리와 주파수영역의 처리로 크게 분류할 수 있으며, 최근의 휴대전화에 탑재되어 있는 음성강조 방법의 대부분이 주파수영역의 처리를 고려하고 있으며 다수의 연구성과가 보고된다. 특히 제안한 논문에서는 주파수영역의 처리 방법을 사용한 음성강조기술에 대하여 기술한다.

논문에서 제안하는 MMSE-STSA의 추정음성은 소음 중첩음성의 위상정보를 사용하여 위상이 포함된 추정 음성진폭스펙트럼을 역푸리에 변환(IFFT, Inverse Fast Fourier Transform)에 의하여 구할 수 있다. MMSE-STSA 잡음억압처리는 1채널 입력계의 주파수 영역처리이며 주위잡음이 혼입한 음성입력 음성의 단시간 스펙트럼 진폭성분에 잡음억압 이득(Gain)을 중첩하여 주위잡음을 억압하여 음성을 강조한다. 이 처리는 주파수마다 잡음전력의 추정치에 기초하여 입력음성에 포함되는 잡음성분을 억압하기 위하여 잡음 전력의 추정 정도가 억압 처리 후의 음성품질에 크게 관여한다. 이러한 것은 입력신호에 음성이 포함되는 유성 구간에 있어서 주위잡음이 변동하는 경우에 특히 중요하다.

잡음전력 추정에 있어서 음성구간검출 방법을 사용하여 음성이 존재하지 않은 구간에서 이 구간을 잡음으로 판단하여 잡음전력을 취득하는 방법[8][9]이 제안되어 있다. 그러나 이러한 방법에서는 음성 구간 중의 주위잡음의 변동에 따라갈 수 없는 문제를 포함하고 있다. 이 때문에 VAD를 사용하지 않은 음성구간에서도 연속적으로 잡음 전력을 추정하는 방법으로서 전력 스펙트럼 평균치의 최소치를 각 주파수마다 구하여 현시점의 잡음 전력으로 하는 방법이 이용된다.

논문에서는 잡음의 크기가 급속히 변화하는 경우에 응용할 수 있는 주파수 영역의 MMSE-STSA의 음성향상 알고리즘을 제안한다. 제안한 MMSE-STSA 음성향상 알고리즘은 주파수 영역에서 추정 스펙트럼 잡음을 사용한다. 논문에서는 잡음이 혼합된 잡음음성 신호 중에서 음성신호가 포함되어 있지 않은 잡음구간만을 사용하여 각 프레임에서 잡음의 주파수영역의 스펙트럼을 사용한다. 따라서 제안한 논문에서는 음성이 존재하지 않는 각 프레임 구간에서 문턱치 기반의 VAD를 사용하여 음성신호가 존재하는 가를 결정한 후 묵음구간에서만 잡음 전력을 추

정하는 방법을 사용한다. 기존의 MMSE-STSA에서 사용한 방법[6][7]과 다르게 본 알고리즘은 서로 다른 데이터베이스의 음성과 인간의 음성과 유사한 음성 특성을 가진 바블잡음과 전시장내 잡음 등을 사용하여 각 프레임에서 잡음 레벨에 따라 적절하게 스펙트럼을 추정하는 차이점이 있다. 따라서 음성인식[12][13]에 있어서 어려움이 많은 유색잡음에 대해서도 제안하는 알고리즘이 다양한 잡음 및 잡음 레벨에 대해서 추정이 가능하고, 또한 실험에서 본 알고리즘의 우수성을 확인할 수 있다.

III. 제안 알고리즘의 실험 결과

실험에서는 2 종류의 데이터 셋을 사용하였으며, 16kHz로 표본화된 NOIZEUS 음성 데이터셋과 TIMIT 음성 데이터셋[14]이다. 본 실험에서 사용한 NOIZEUS는 연구 그룹 간에서 음성 강조 알고리즘의 비교를 용이하게 하기 위해서 사용되는 잡음이 포함된 음성코퍼스이며, 실험에서는 NOIZEUS의 남성 음성 파일을 사용하였다. 잡음 신호는 AURORA2 데이터베이스에서 가져왔으며 다양한 장소에서 녹음한 내용을 포함하고 있다. 실험에 사용하는 Aurora2 데이터베이스는 ELRA(European Language Resources Association)에서 배포하였으며 잡음 하의 음성인식의 평가용 데이터베이스이다. 또한 논문에서는 가우시안 백색잡음도 사용하여 실험을 실시하였다.

제안한 논문의 MMSE-STSA의 음성신호 및 잡음신호의 전력 추정에 따른 음성을 개선하는 분석 실험에는 NOIZEUS Database의 깨끗한 남성의 음성신호(M1)와 TIMIT의 깨끗한 남성의 음성신호(M2)를 사용하였으며, 잡음음성으로는 white noise, car noise, babble noise, exhibition noise를 중첩한 잡음혼입 음성신호를 사용하여 실험을 수행한다. 논문에서는 15dB, 10dB, 5dB, 0dB의 입력 SNR이 되도록 백색잡음과 AURORA2 DB의 3종류의 유색잡음을 두 종류의 깨끗한 음성신호에 혼합시켜 잡음이 혼합된 음성신호를 구성하여 실험을 실시하였다. 논문에서는 음성 및 잡음신호를 이용하여 제안한 알고리즘을 MATLAB을 사용하여 성능평가를 하였으며, 출력 SNR를 이용하여 제안한 MMSE-STSA 알고리즘의 성능을 실험에 의해서 평가하였다.

표 1부터 표 4는 실험에 사용된 2 종류의 남성화자에 의한 음성에 백색잡음, 자동차잡음, 바블잡음, 전시회잡음을 각각 부가한 경우에 대해서 입력 SNR을 0dB, 5dB, 10dB, 15dB로 하였을 때, 출력 SNR의 성능을 측정한 실험 결과이다. 표의 결과로부터 알 수 있듯이 입력 잡음이 증가할수록 제안한 알고리즘(Prop.)의 출력 SNR이 향상(Impr.)되고 있으며, 최대 9.13(dB)까지 개선되고 있다(표 1의 백색잡음의 경우). 또한 입력 잡음의 강도가 높은 경우(즉, 입력 SNR이 낮은 경우: 자동차잡음, 바블잡음, 전시회잡음)에는 개선된 양이 비교적 줄어들었지만 전반적으로 본 알고리즘에 의한 개선 성능은 양호하다는 것도 확인할 수 있었다.

표 1. 백색잡음에 대한 SNR 결과
Table 1. SNR results for white noise

Input SNR (dB)	Output SNR results (dB)			
	M1		M2	
	Prop.	Impr.	Prop.	Impr.
15	17.91	2.91	17.08	2.08
10	15.57	5.57	14.56	4.56
5	12.54	7.54	11.54	6.54
0	9.13	9.13	8.29	8.29

표 2. 자동차잡음에 대한 SNR 결과
Table 2. SNR results for car noise

Input SNR (dB)	Output SNR results (dB)			
	M1		M2	
	Prop.	Impr.	Prop.	Impr.
15	18.89	3.89	16.82	1.82
10	14.93	4.93	13.26	3.26
5	10.10	5.10	9.38	4.38
0	5.29	5.29	4.90	4.90

표 3. 바블잡음에 대한 SNR 결과
Table 3. SNR results for babble noise

Input SNR (dB)	Output SNR results (dB)			
	M1		M2	
	Prop.	Impr.	Prop.	Impr.
15	18.95	3.95	16.31	1.31
10	14.75	4.75	12.25	2.25
5	9.79	4.79	7.54	2.54
0	3.67	3.67	2.73	2.73

표 4. 전시회잡음에 대한 SNR 결과

Table 4. SNR results for exhibition noise

Input SNR (dB)	Output SNR results (dB)			
	M1		M2	
	Prop.	Impr.	Prop.	Impr.
15	16.92	1.92	16.35	1.35
10	13.09	3.09	12.85	2.85
5	8.90	3.90	8.74	3.74
0	4.53	4.53	4.30	4.30

표 5는 기존의 다른 방법과의 성능 비교를 위하여 지각 필터뱅크(PF, Perceptual Filterbank) 및 MMSE-STSA(PF+MMSE-STSA)[7]에 의한 방법과 제안한 알고리즘과의 비교 결과를 나타낸다. 표는 M1 음성과 백색잡음에 대한 SNR의 성능 비교 결과이며 최대 2.16dB 개선됨을 알 수 있다. 그림 1은 표 1~4의 SNR 결과를 보다 명확하게 나타내기 위하여 M1 음성의 각 잡음에 대한 출력 SNR 결과를 보여 주고 있다. 그러므로 제안하는 알고리즘을 사용함으로써 잡음의 강도 및 급격한 변화에도 충분히 대응할 수 있었으며, 또한 SNR이 향상됨으로써 제안하는 알고리즘이 우수함을 알 수 있었다.

표 5. SNR에 의한 백색잡음의 성능 비교

Table 5. Performance comparison of white by SNR

Input SNR (dB)	Output SNR results (dB)			
	Wiener	MMSE-LSA	PF+MMSE-STSA	Prop.
10	12.13	12.51	13.64	15.57
5	9.00	9.36	10.36	12.54
0	5.80	6.18	7.13	9.13

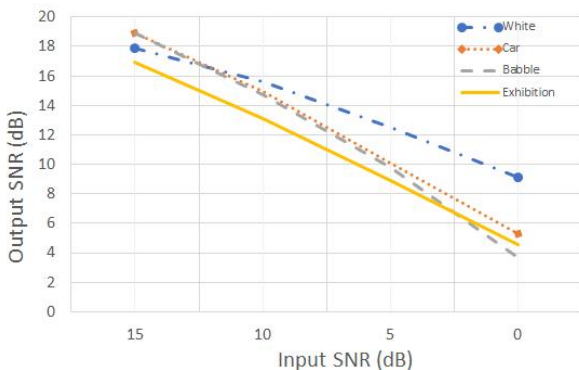


그림 1. 제안한 알고리즘의 성능 비교

Fig. 1. Comparison of proposed algorithm

IV. 결 론

논문에서는 최소 2승오차 단시간 진폭스펙트럼 추정 방법을 사용하여 음성신호 및 잡음신호의 전력을 추정하였으며, 두 종류의 음성 데이터베이스와 다양한 잡음 데이터베이스를 적용하여 높은 잡음 환경 하에서도 신호대잡음비가 상당 부분 향상되는 것을 실험으로 확인할 수 있었다. 특히 입력의 신호대잡음비가 높은 경우에도 출력 결과가 상당히 개선됨으로서 논문에서 제안한 알고리즘의 유효성이 입증되었다. 향후의 연구 과제로서는 다양한 음성신호와 강한 잡음신호에 대하여도 좀 더 강건한 시스템의 구축이 가능한 알고리즘을 테스트하는 것을 목표로 하고 있다.

References

- [1] D. Chazan, G. Cohen, R. Hoory, and M. Zibulski, "Low bit rate speech compression for playback in speech recognition systems", 2000 10th European Signal Processing Conference, Tampere, Finland, pp. 1-4, Sep. 2000.
- [2] B. R. Ravi, S. P. Deepu, M. R. Kini, and D. S. Sumam, "Wavelet based Noise Reduction Techniques for Real Time Speech Enhancement", 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, pp. 846-851, Feb. 2018. <https://doi.org/10.1109/SPIN.2018.8474210>.
- [3] E. Nemer, R. Goubran, and S. Mahmoud, "Robust voice activity detection using higher-order statistics in the LPC residual domain", IEEE Trans. Speech Audio Process., Vol. 9, No. 3, pp. 217-231, Mar. 2001. <https://doi.org/10.1109/89.905996>.
- [4] Y. H. Tu, J. Du, L. R. Dai, and C. H. Lee, "Speech Separation based on signal-noise-dependent deep neural networks for robust speech recognition", 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, pp.

61-65, Apr. 2015. <https://doi.org/10.1109/ICASSP.2015.7177932>.

[5] T. Esch and P. Vary, "Efficient musical noise suppression for speech enhancement system", 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, pp. 4409-4412, Apr. 2009. <https://doi.org/10.1109/ICASSP.2009.4960607>.

[6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator", IEEE Trans. on Speech and Audio Processing, Vol. ASSP-32, No. 6, pp. 1109-1121, Dec. 1984.

[7] H. Tasmaz and E. Ercelesi, "Speech enhancement based on undecimated wavelet packet-perceptual filterbanks and MMSE-STSA estimation in various noise environments", Digital Signal Processing, Vol. 18, No. 5, pp. 797-812, Sep. 2008. <https://doi.org/10.1016/J.DSP.2007.10.003>.

[8] J. J. Lee, H. T. Han, Y. U. Choi, and H. Y. Lee, "Voice Activity Segment Audio Deduction Method using MobileNet", Journal of KIIT, Vol. 20, No. 3, pp. 1-8, Mar. 2022. <http://dx.doi.org/10.14801/jkiit.2022.20.3.1>.

[9] J. S. Choi, "Gender Recognition for Speaker in Colored Noise by Speech/Non-speech Discrimination", Journal of KIIT, Vol. 10, No. 11, pp. 63-68, Nov. 2012. <https://doi.org/10.14801/kiitr.2013.11.4.57>.

[10] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. on Acoustic Speech Signal Processing, Vol. ASSP-27, No. 2, pp. 113-120, Apr. 1979.

[11] A. Yelwande, S. Kansal, and A. Dixit, "Adaptive wiener filter for speech enhancement", 2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC), Indore, India, pp. 1-4, Aug. 2017. <https://doi.org/10.1109/ICOMICON.2017.8279110>.

[12] J. Y. Ahn, S. B. Kim, S. H. Kim, and K. I.

Hur, "A study on Voice Recognition using Model Adaptation HMM for Mobile Environment", The journal of the institute of internet, broadcasting and communication (IIBC), Vol. 11, No. 3, pp. 175-180, Jun. 2011. <https://doi.org/10.7236/JIWIT.2011.11.3.175>.

[13] J. H. Cho, "Efficient Compensation of Spectral Tilt for Speech Recognition in Noisy Environment", The journal of the institute of internet, broadcasting and communication (IIBC), Vol. 17, No. 1, pp. 199-206, Feb. 2017. <https://doi.org/10.7236/JIIBC.2017.17.1.199>.

[14] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "Timit acoustic phonetic continuous speech corpus", Linguistic Data Consortium, 1993. <https://doi.org/10.35111/17gk-bn40>.

저자소개

최재승 (Jae-Seung Choi)



1989년 2월 : 조선대학교
전자공학과(공학사)
1995년 3월 : 일본 오사카시립대학
전자정보공학부(공학석사)
1999년 9월 : 일본 오사카시립대학
전자정보공학부(공학박사)
2000년 5월 ~ 2001년 12월 : 일본

마쯔시타 전기산업주식회사 (현, 파나소닉) AVC사
연구원

2002년 4월 ~ 2007년 2월 : 경북대학교 디지털기술연구소
책임연구원

2007년 3월 ~ 현재 : 신라대학교 전기전자공학과 교수
관심분야 : 음성인식, 음성강조, 잡음제거, 음원분리