
서브밴드에 기반한 스펙트럼 차감 알고리즘

최재승*

Subband Based Spectrum Subtraction Algorithm

Jae-Seung Choi*

요약

본 논문에서는 거리측정, 로그전력, 실효치 방법에 의하여 유성음, 무성음, 묵음 구간을 검출하여, 서브밴드 필터에 의한 잡음제거 알고리즘을 제안한다. 제안한 알고리즘은 각 프레임에서 서브밴드 필터를 사용하여 잡음으로 오염된 음성신호로부터 백색잡음 및 도로잡음의 스펙트럼을 차감하는 방법이다. 본 실험에서는 Aurora-2 데이터베이스에 포함된 음성신호와 잡음신호를 사용하여 스펙트럼 차감 알고리즘의 결과를 나타낸다. 잡음에 의하여 오염된 음성신호에 대하여 신호대잡음비를 사용하여 본 알고리즘이 유효하다는 것을 확인한다. 실험으로부터 백색잡음에 대하여 평균 2.1 dB, 도로잡음에 대하여 평균 1.91 dB의 출력 신호대잡음비가 개선된 것을 확인할 수 있었다.

ABSTRACT

This paper first proposes a classification algorithm which detects a voiced, unvoiced, and silence signal using distance measure, logarithm power and root mean square methods at each frame, then a spectrum subtraction algorithm based on a subband filter. The proposed algorithm subtracts spectrums of white noise and street noise from noisy signal based on the subband filter at each frame. In this experiment, experimental results of the proposed spectrum subtraction algorithm demonstrate using the speech and noise data of Aurora-2 database. Based on measuring the speech-to-noise ratio (SNR), experiments confirm that the proposed algorithm is effective for the speech by contaminated the noise. From the experiments, the improvement in the output SNR values was approximately 2.1 dB and 1.91 dB better for white noise and street noise, respectively.

키워드

Subband, distance measure, classification algorithm, spectrum subtraction.

서브밴드, 거리척도, 분류알고리즘, 스펙트럼차감

1. 서론

최근, 음성인식 시스템에 적용되는 음성인식 알고리즘은 배경잡음 환경 하에서 음성부호화 및 음성인식기와 같은 음성신호처리 시스템에 보편적으로 적용되어 응용되는 추세에 있다. 특히 음성인식 알고리즘에 의

하여 추정된 배경잡음 신호는 잡음제거 혹은 음성인식 파라미터의 추정에 영향을 미치는 요소이기 때문에 음성 인식기에 적용되고 있으며, 따라서 다양한 배경잡음 환경 하에서 효과적인 인식방법이 필요하다[1, 2, 3]. 현재의 음성인식 기술의 수준으로 여러 방면에 응용할 경우에 잡음이 없는 환경에서는 비교적 성공적으

* 신라대학교 전자공학과(jschoi@silla.ac.kr)

접수일자 : 2013. 01. 20

심사완료일자 : 2013. 03. 25

게재확정일자 : 2013. 04. 22

로 적용될 수 있다. 그러나 배경잡음이 존재하는 환경에서는 음성 인식률이 비교적 낮기 때문에 실제로 적용되기 어려운 상황이다. 일반적으로 잡음을 추정하는 알고리즘으로는 음성구간의 검출 알고리즘, 음성의 전력 스펙트럼에 의한 잡음전력 추정 알고리즘 등으로 분류할 수 있으며, 여러 방면의 연구자들에 의하여 잡음 환경 하에서 음성인식기의 성능 향상을 위하여 노력하고 있다[4].

배경잡음 환경 하에서 잡음 성분을 추정하기 위해서 일반적으로 음성/비음성 구간을 판단할 수 있는 Voice Activity Detector(VAD), 전력스펙트럼 방법을 사용하고 있다[5, 6, 7]. VAD를 사용할 경우에 각 프레임에서의 에너지와 영교차율을 이용한 알고리즘이 일반적으로 사용되었으나, 음성/비음성 구간을 추정하는데 오차가 발생하게 되면 예상하지 못한 결과를 가져오고 있다. 따라서 이러한 문제점을 해결하기 위하여 본 논문에서는 각 프레임에서의 거리측정 방법에 의하여 유성음, 무성음, 묵음 구간을 검출하는 알고리즘을 먼저 제안한다. 이 알고리즘은 각 프레임에서 캡스트럼(Cepstrum), 로그전력(Log power), 실효치(Root Mean Square, RMS) 방법을 사용하여 유성음, 무성음, 묵음 구간을 검출하는 방법이다.

일반적인 유성음/무성음/묵음 검출 방법들은 낮은 잡음 환경에서는 비교적 우수한 성능을 보여주고 있지만, 높은 잡음 환경에서는 검출률이 현저히 떨어지는 문제점을 가지고 있다. 따라서 본 논문에서는 이를 해결하기 위하여 캡스트럼, 로그전력, 실효치 방법을 결합하여 유성음/무성음/묵음을 검출하는 방법을 제안한다. 이 알고리즘을 적용하면 높은 잡음 환경 하에서도 유성음/무성음/묵음의 검출의 정확도가 비교적 양호할 수 있었다. 이러한 유성음/무성음/묵음 검출 알고리즘을 이용하여 전체 주파수 영역을 각 서브밴드 영역으로 분류하여, 이 후에 구간 검출된 결과를 사용하여 각 대역별로 잡음을 억제하는 서브밴드필터[8, 9, 10]에 의한 스펙트럼 차감 알고리즘을 제안한다. 본 실험에서는 스펙트럼 차감 알고리즘의 성능을 서브밴드필터를 사용하여 각 프레임에서 잡음을 제거하는 실험결과를 나타낸다. 잡음에 의하여 오염된 음성에 대하여 신호대잡음비(Signal-to-Noise Ratio, SNR)를 사용하여 본 알고리즘이 유효하다는 것을 확인한다. 백색잡음에 대하여 2 dB 이상, 도로잡음에 대하여 1.9 dB 이

상이 개선된 것을 확인할 수 있었다.

II. 검출 알고리즘

배경잡음이 존재하는 실제 환경 하에서의 유성음/무성음/묵음 검출은 잡음의 스펙트럼 성분이 음성의 스펙트럼 성분과 비교적 유사하기 때문에 잡음을 음성으로 잘못 검출하는 경우가 종종 발생한다. 따라서 이러한 문제점을 해결하기 위하여 본 논문에서는 캡스트럼, 로그전력, 실효치 방법을 결합하여 유성음/무성음/묵음을 검출하는 방법을 제안한다. 본 논문에서는 음성인식 시스템에 입력되는 유성음/무성음/묵음 신호를 인식하는 시스템을 구축하는 것을 목적으로 한다. 본 논문에서는 입력신호의 주파수 특성을 이용하여, 캡스트럼의 거리측정 방법, 로그전력에 의한 방법, 실효치의 에너지 측정에 의한 방법 등에 의하여 유성음/무성음/묵음 신호를 식별하는 방법을 제안한다.

먼저 유성음/무성음/묵음 신호를 식별하기 위하여, 유성음 구간과 비유성음 구간으로 검출한다. 따라서 입력되는 음성데이터를 캡스트럼 영역에서의 각 음성데이터의 유클리드(euclid) 거리로 표현되는 식 (1)과 같은 캡스트럼 거리식으로 측정한다[11].

$$Eud_dis_{j,*} = \sqrt{\sum_{i=1}^{coef} [y_{j,i} - Cen_{*,i}]^2} \quad (1)$$

여기에서 $y_{j,i}$ 는 잡음이 중첩된 음성신호의 캡스트럼 계수, i 는 각 프레임에서의 캡스트럼 ($i=1, 2, \dots, coef$)을 나타내며, 본 실험에서의 $coef$ 수는 12개로 결정하였다. 그리고 $Cen_{*,i}$ 는 캡스트럼의 중심을 나타낸다. 따라서 식 (2)와 같이 캡스트럼 거리가 일정한 문턱치 이상일 때에 유성음 신호로 판정하며, 문턱치 이하일 때에는 비유성음 신호로 판정한다. 본 실험에서는 문턱값 δ 가 0.4 이상이면 유성음 신호로 판정한다.

$$Eud_dis_{j,*} > \delta \quad (2)$$

다음으로 식 (1)과 식 (2)에 의해서 유성음 구간 및 비유성음 구간으로 검출한 후에, 비유성음 구간을 다시 무성음 구간과 묵음 구간으로 검출한다. 따라서 무

성음 구간과 묵음 구간의 신호를 식별하기 위하여 각 프레임에서 고속 푸리에변환(Fast Fourier Transform, FFT)한 후에 로그 스펙트럼 및 전체 문장의 실효치를 구하여 문턱값과 비교한다. 따라서 식 (3), 식 (4)와 같이 무성음/묵음 신호를 식별한다. 그러므로 식 (3)을 만족할 때에는 무성음으로 판별되며, 식 (4)를 만족할 때에는 묵음으로 판별된다.

$$\text{Log}|Y(\omega)|_{[r]} > \frac{TH_1}{\alpha} \quad (3)$$

$$\text{Log}|Y(\omega)|_{[r]} < \frac{TH_2}{\alpha} \quad (4)$$

여기서 $\text{Log}|Y(\omega)|_{[r]}$ 은 각 프레임의 구간에서 얻어진 문장의 실효값에 대하여 로그 전력값을 취한 값을 나타내며, TH_1 과 TH_2 는 각각 문턱값을 나타낸다. 본 실험에서의 α 는 문턱값을 결정하기 위한 파라미터이며, 실험적으로 최적의 값으로 결정한다.

이상과 같이, 최종적으로 그림 1과 같이 캡스트럼 거리 방식과 로그전력 에너지를 결합하여 유성음/무성음/묵음 신호를 검출하였다. 즉, 캡스트럼 거리를 문턱값과 비교하여 각 프레임에서 큰 값을 갖는 구간을 유성음으로 분류하고, 작은 값을 비유성음으로 분류하였다. 이 후에 분류된 결과 중에서 비유성음으로 분류된 구간에 대하여, 로그 스펙트럼 및 전체 문장의 실효치(Root Mean Square, RMS) 방법을 이용하여 무성음 및 묵음 구간을 재분류하였다. 먼저 음성이 입력되면 전처리 과정을 거친 음성신호를 가지고 음성의 특징 파라미터인 FFT의 캡스트럼 계수를 구한다[12]. 표준패턴은 특정화자의 파라미터를 입력시켜 놓고 입력된 음성과 유클리드 거리차를 계산하여 비교한다. 입력음성과 표준패턴과의 유사도는 유클리드 거리(Euclidean distance)를 사용하였다. 이 유클리드 거리는 파라미터 개수가 동일한 차원인 표준패턴과 미지패턴의 차에 해당하며, 어떤 음성의 N차 파라미터의 유클리드 거리의 식의 결과가 최소가 되는 표준패턴을 결정하여 입력음성을 판정한다.

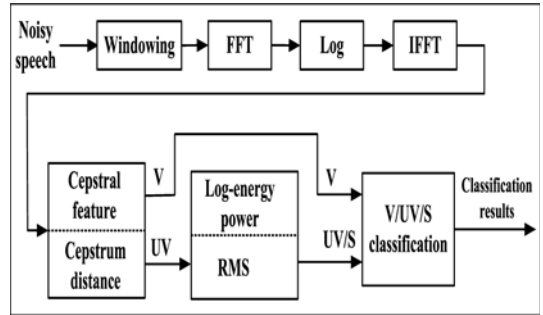


그림 1. 유성음/무성음/묵음 검출 알고리즘
Fig. 1 Voiced/unvoiced/silence classifier algorithm

III. 서브밴드에 기초한 스펙트럼 차감 알고리즘

본 논문에서 제안한 알고리즘에 의해 구현된 시스템의 블록도를 그림 2에 나타낸다. 시스템의 입력으로 들어오는 음성신호는 8 kHz로 표본화된 잡음이 섞인 음성신호이다. 본 논문에서는 먼저 음성신호로부터 FFT를 수행한 후에, 2장에서 구현한 구간 검출 알고리즘에 의하여 유성음/무성음/묵음으로 분류한다. 이러한 인식결과를 사용하여 음성과 잡음의 주파수 분포 특성을 고려하여 각 서브밴드에 따라 가중치를 주어 결합하여, 정류기와 대수 연산을 통하여 잡음을 억제하는 알고리즘이다. 제안한 알고리즘에서는 음성신호의 정보가 대부분이 저주파 대역에 분포하는 음성신호의 주파수 분포 특성을 충분히 고려하여, 추정된 검출 결과를 이용하여 서브밴드에 따라 가중치를 주어 결합하는 방법이다.

본 논문에서는 음성의 입력 크기에 따라 중심주파수를 다르게 적용하여 주파수 밴드별에 따라서 서로 다른 청각필터 형태를 가진 3개의 서브밴드 필터를 구성한다. 본 논문은 4 kHz의 주파수 대역 중에서 저역에 해당하는 0~2 kHz를 2개의 대역으로, 고역에 해당하는 2~4 kHz를 1개의 대역으로 세분화하여 주파수 능력을 극대화하고자 하는 목적의 연구이다. 본 논문에서는 이러한 특성을 이용하여 서브밴드 필터에 의하여 잡음환경에 적응시켜 잡음에 강한 잡음억압 알고리즘을 제안한다. 본 논문에서는 4 kHz의 주파수 대역(8 kHz의 샘플링 주파수)에 대하여 표 1과 같이 전체 주파수 대역을 3개의 다중 대역으로 나눈다.

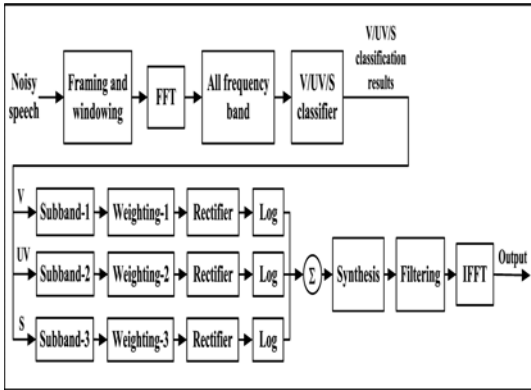


그림 2. 서브밴드에 의한 스펙트럼 차감 알고리즘
Fig. 2 Noise subtraction algorithm by subband

표 1. 각 서브밴드의 주파수 범위
Table 1. Frequency range of each sub-band

서브밴드	주파수 범위[Hz]
Sub-band 1	0.0 Hz~1.0 kHz
Sub-band 2	1.0 kHz~2.0 kHz
Sub-band 3	2.0 kHz~4.0 kHz

IV. 음성 및 잡음 데이터

본 실험에서 사용한 음성신호는 8 kHz의 샘플링 주파수를 가진 환경에서 녹음된 영어숫자로 구성된 Aurora2 데이터베이스(Database, DB)[13]를 사용하였다. 본 실험에서는 Aurora2 DB의 테스트 셋 A, B, C 중에서 임의적으로 남성 화자에 의한 10문장을 선택하여 평가용으로 사용하였다. 본 실험에 사용한 음성신호는 한 프레임의 분석구간의 길이가 16 ms이며, 분석구간의 데이터에 대하여 해밍창(Hamming window)을 통과시켰다.

본 실험에서 사용한 잡음데이터는 Aurora2 DB의 도로잡음(street noise)을 사용하였으며, 컴퓨터에 의해서 작성된 가우스 백색잡음(white noise)을 사용하여 평가하였다. 이러한 배경잡음은 샘플링 주파수 8 kHz이다. Aurora2 DB의 각 테스트 셋에 포함된 배경잡음들이 음성데이터에 부가된 잡음이 중첩된 음성데이터(즉, $SNR_{input} = 20 \text{ dB}, 15 \text{ dB}, 10 \text{ dB}, 5 \text{ dB}, 0 \text{ dB}$)들이 포함되어 있다. 백색잡음에 대해서는 Aurora2 DB의

음성데이터에 별도로 백색잡음을 부가하여 잡음이 중첩된 음성데이터를 구하였다.

잡음부가음성은 위에서 기술한 음성데이터에 백색잡음 및 도로잡음을 중첩하여 작성하였다. 각 음성데이터마다 서로 다른 백색잡음 및 도로잡음을 중첩함으로써 입력 신호대잡음비(SNR_{input})가 15 dB의 잡음이 중첩된 음성을 작성하였다. 또한, SNR_{input} 으로서는 음성 $S(n)$ 과 잡음 $N(n)$ 의 전체에 해당하는 전력의 비율로서 정의되는 전역 SNR_{input} 을 사용하였다.

본 음성인식시스템의 평가방법으로는 각 프레임에서의 구간 인식률을 도입하였다. 이 인식률은 입력문장의 모든 프레임 수에 대하여 각 프레임에서 인식률이 정확하게 검출된 프레임수의 비율로 정의한다. 본 논문에서 제안한 잡음제거의 평가방법으로는 식 (5)와 같은 출력 신호대잡음비(SNR_{output} (Output Signal-to-Noise Ratio))를 사용하였다. 여기에서, $\hat{S}(n)$ 은 강조된 출력 음성신호이다.

$$SNR_{output} = 10 \cdot \log_{10} \left(\frac{\sum_{n=1}^N S(n)^2}{\sum_{n=1}^N [S(n) - \hat{S}(n)]^2} \right) \quad (5)$$

V. 실험 결과

본 논문에서는 유성음/무성음/목음 신호의 구간 인식 및 잡음억제 알고리즘 구현을 위하여, 제안한 알고리즘을 이용하여 실험을 한다. 실험환경은 Intel(R) Core(TM)2 Quad 2.4GHz CPU와 3.25GB RAM이 장착된 IBM 호환 컴퓨터에서 Windows XP OS를 사용하여 Microsoft VC++ 6.0으로 구현하였다.

본 실험에서 평가용으로 사용하는 음성은 Aurora2 데이터베이스의 테스트 셋 A, B, C로부터 잡음이 중첩된 음성 데이터들이 임의적으로 선택되었으며, 잡음 데이터는 학습 시에 사용한 동일한 잡음인 백색잡음 및 Aurora2 데이터베이스의 도로잡음이 선택되었다. 제안한 시스템은 4장에서 기술한 음성데이터 및 잡음 데이터를 사용하여 검출률 및 잡음억제의 실험을 실시하였다.

표 2는 백색잡음 및 도로잡음에 대하여 캡스트럼

거리 측정 방법을 사용하여, 유성음과 비유성음으로 분류하는 검출률의 실험 결과를 나타낸다. 표 3은 로그에너지와 RMS 방법을 사용하여, 백색잡음 및 도로잡음에 대하여 무성음과 묵음으로 분류하는 검출률을 나타내고 있다. 표의 실험결과로부터 SNR_{input} 이 15 dB인 경우에 각 프레임에서 유성음/무성음/묵음 신호의 분류가 양호하게 검출되는 것을 알 수 있다. 표 3과 표 4는 본 논문에서 제안하는 그림 2의 서브밴드에 의한 스펙트럼 차감 알고리즘을 사용하여, SNR_{input} 이 15 dB인 경우에 백색잡음 및 도로잡음에 대하여 출력 SNR의 실험 결과를 나타내고 있다. 표의 실험 결과로부터, 백색잡음에 대해서는 서브밴드를 사용하지 않은 방법보다 약 2.1 dB 향상되었으며, 도로잡음에 대해서는 SS 방법보다 약 1.91 dB 향상된 것을 알 수 있다.

따라서 이상의 실험결과로부터, 유성음/무성음/묵음 구간의 분류 및 잡음억제를 통하여 각 프레임에서 잡음제거가 가능한 것을 보였으며, 입력 SNR이 15 dB인 잡음이 중첩된 음성신호에 대해서 본 방식에 의한 잡음억제 효과가 있는 것을 알 수 있었다.

표 2. 켈스트럼 거리측정에 의한 검출율 (%)
Table 2. Classification rates by cepstrum distance

Speech	검출율 (%)			
	White noise		Street noise	
	유성음	비유성음	유성음	비유성음
S1	96.3%	93.5%	94.2%	92.1%
S2	95.8%	92.9%	93.6%	91.8%
Total	96.1%	93.2%	93.9%	92.0%

표 3. 로그에너지 및 RMS에 의한 검출율 (%)
Table 3. Classification rates by log-energy and RMS

Speech	검출율 (%)			
	White noise		Street noise	
	무성음	묵음	무성음	묵음
S1	95.2%	92.6%	93.5%	91.0%
S2	94.6%	91.7%	92.7%	90.3%
Total	94.9%	92.2%	93.1%	90.7%

표 4. 백색잡음인 경우의 SNR 실험 결과
Table 4. Experimental results of SNR in white noise

Speech	Input SNR[dB]	Output SNR [dB]		
		No subband	Prop.	Impr.
S1	15.00	13.50	15.63	2.13
S2	15.00	13.93	15.89	1.96
Total	-	-	-	2.1

표 5. 도로잡음인 경우의 SNR 실험 결과
Table 5. Experimental results of SNR in street noise

Speech	Input SNR[dB]	Output SNR [dB]		
		No subband	Prop.	Impr.
S1	15.00	10.04	11.98	1.94
S2	15.00	10.15	12.03	1.88
Total	-	-	-	1.91

VI. 결론

본 논문에서는 켈스트럼 유클리디언 거리와 로그에너지를 결합하여 유성음/무성음/묵음을 분류하는 알고리즘을 제안하여, 이 알고리즘을 사용하여 잡음을 억제하는 서브밴드 필터에 의한 잡음억제 알고리즘을 제안하였다. 본 알고리즘은 유클리디언 거리측정을 사용하여 유성음/비유성음의 분류, 로그전력 및 RMS를 이용하여 무성음/묵음 신호의 분류가 가능하였다. 또한 subband를 사용하지 않은 경우의 SNR과 비교하여 본 논문에서 제안한 방법이 우수하다는 것을 실험으로 확인하였다. Aurora 2.0 DB의 음성을 사용하여, SNR_{input} 이 15 dB인 경우에 백색잡음에 대하여 평균 2.1 dB, 도로잡음에 대하여 평균 1.91 dB의 SNR_{output} 이 향상되었다. 향후의 연구과제는 입력 SNR 및 배경잡음의 종류를 늘려서 본 알고리즘을 개선할 필요가 있다.

참고 문헌

- [1] T. Hirahara and H. Iwamida, "Auditory spectrograms in HMM phoneme recognition," Proc. Int. Conf. Spoken Lang. Process.,

- ICSLP-90, pp. 1381-1384, 1990.
- [2] C. Y. Lee and Y. G. Lee, "Study of the noise processing to technique speech recognition system", Journal of the Korea Society of Computer & Information, Vol. 7, No. 2, pp. 73-78, June, 2002.
- [3] 이창영, "FIR 필터링과 스펙트럼 기울이기가 MFCC를 사용하는 음성인식에 미치는 효과", 한국전자통신학회논문지, 5권, 4호, pp. 363-371, 2010.
- [4] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Processing. Vol. 27, No. 2, pp. 113-120, 1979.
- [5] 방동혁, 길세기, 강현덕, 윤광섭, 이상민, "디지털 보청기에서 마이크로폰 어레이를 이용한 잡음제거", 전기학회논문지, 58권, 4호, pp. 857-866, 2009.
- [6] Y. Wu and Y. Li, "Robust speech/non-speech detection in adverse conditions using the fuzzy polarity correlation method," IEEE International Conference on Systems, Man, and Cybernetics, Vol. 4, pp. 2935-2939, 2000.
- [7] 최재승, "퍼셉트론 신경회로망을 사용한 유성음, 무성음, 묵음 구간의 검출 알고리즘", 한국전자통신학회논문지, 6권, 2호, pp. 237-242, 2011.
- [8] J. Chen, K. K. Paliwal and S. Nakamura, "Sub-band based additive noise removal for robust speech recognition," in Proc. Euro-speech, pp. 571-574, 2001.
- [9] L. Lin, W. H. Holmes and E. Ambikairajah, "Subband noise estimation for speech enhancement using a perceptual Wiener filter," IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. I-80-I-80-3, 2003.
- [10] 한영오, "GRNN을 이용한 동영상 움직임 예측 및 대역분할 부호화에 관한 연구", 한국전자통신학회논문지, 5권, 3호, pp. 256-261, 2010.
- [11] 최재승, "멀티밴드 필터에 의한 환경잡음억압 알고리즘", 한국컴퓨터정보학회논문지, 17권, 8호, pp. 92-97, 2012.
- [12] J. He, L. Liu, and G. Palm, "On the use of residual cepstrum in speech recognition," 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 5-8, 1996.
- [13] H. Hirsch, and D. Pearce, "The AURORA

experimental framework for the performance evaluations of speech recognition systems under noisy conditions", in Proc. ISCA ITRW ASR2000 on Automatic Speech Recognition: Challenges for the Next Millennium, Paris, France, pp. 181-188, September 2000.

저자 소개



최재승(Jae-Seung Choi)

1989년 조선대학교 전자공학과 공학사

1995년 일본 오사카시립대학 전자정보공학부 공학석사

1999년 일본 오사카시립대학 전자정보공학부 공학박사
2000년~2001년 일본 마쯔시다 전기산업주식회사(현, 파나소닉 주식회사) AVC사 연구원

2002년~2007 경북대 디지털기술연구소 책임연구원

2007년~현재 신라대학교 전자공학과 교수

※ 관심분야 : 음성신호처리, 신경회로망, 적응필터와 잡음제거, 디지털 TV 및 멀티미디어 등