잔향제거를 이용한 음성통신 시스템 성능 향상

김세영* · 강석엽* · 김기만*

Performance Enhancement of Speech Communication System using Reverberation Rejection

Se-young Kim* · Suk-youb Kang* · Ki-Man Kim*

본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다. (계약번호 UD070054AD)
본 연구는 중소기업청의 중소기업기술혁신사업의 지원으로 수행되었습니다. (S1036300)

요 약

본 논문에서는 잔향이 존재하는 환경에서 단일 마이크로폰을 사용한 음성 개선 방법을 제시한다. 스펙트럼 차감법(Spectral Subtraction)은 스펙트럼 상에서 잔향성분 및 잡음을 제거 할 수 있는 효과적인 방법이다. 스펙트럼 차감법은 음성과 비음성 구간의 정확한 구분을 필요로 하며 성능을 향상시키기 위해 본 논문에서는 엔트로피(Entropy) 기반의 음성 구간 검출법을 적용하였다. 제시된 방법을 기존의 에너지 검출 기반의 음성검출법을 적용한 스펙트럼 차감법과 비교하여 성능 평가를 수행하였다. SNR 및 잔향시간에 따른 잔향 제거비율을 평가지표로 사용하였으며, 시뮬레이션 결과 기존의 스펙트럼 차감법과 비교하여 제시된 방법이 우수한 성능을 보였다.

ABSTRACT

In this paper, we propose the speech enhancement algorithm using an one-microphone in a reverberant room environments. Spectral subtraction is the effective method which can reduce the reverberation element and the noise in a spectrum domain. Spectral subtraction needs correct separation of voice section and silent section therefore to improve the performance, voice activity detection(VAD) based on entropy has been applied to the proposed method. We test a performance of the proposed method by comparing with conventional method which used VAD based on energy detection. Reverberation reduction ratio with variable of SNR and a reverberation time is used as a test index. From the simulation result, proposed method shows performance better than conventional method.

키워드

Speech enhancement, Spectral subtraction, VAD, Spectral entropy

* 한국해양대학교 공과대학 전파공학과

접수일자 : 2009. 04. 23
심사완료일자 : 2009. 05. 11

Ⅰ. 서 론

일반적으로 마이크로폰에 수신된 화자의 음성 신호는 잔향 및 주변 잡음등의 영향으로 인해 왜곡된다. 특히 잔향 성분은 음성신호의 품질을 떨어뜨려 원격 화상회의, 자동 음성인식, 화자 위치추적과 같은 실내 음향시스템의 성능을 저하시키는 주된 요인이 된다. 따라서 고품질의 음성신호 획득을 위한 효과적인 잔향제거 기법이 요구되며 현재까지 많은 연구들이 진행되고 있다.

잔향제거 기법은 크게 한 개의 마이크로폰을 이용하는 방법과 다수개의 마이크로폰을 이용하는 방법으로 나눌 수 있으며, 다수개의 마이크로폰을 이용한 대표적인 잔향제거 기법에는 일반적인 Beamforming 기법, Blind deconvolution 기법 등이 있다[1]. 한 개의 마이크로폰을 이용한 잔향 제거 기법에는 켑스트럼 기반의 방법과 변조 전달함수 (MTF)를 이용한 기법, 음성신호의 선형예측 잔여를 이용한 방법등이 있으며 다수의 센서를 이용한 잔향 제거 기법에 비해 구현이 어렵고 많은 양의 잔향신호를 필요로 하는 등의 요구조건으로 인해 효과적인 잔향제거를 수행하는데 많은 어려움이 있다[2].

또한 음성 신호의 특성 및 음향 채널에 대한 사전정보 여부에 따라 잔향제거 기법을 분류할 수 있는데 본 논문에서는 두 가지 정보의 영향이 적은 스펙트럼 개선에 기반한 잔향 제거 알고리즘을 연구하였다.

한 개의 마이크로폰을 이용한 스펙트럼 차감법은 잔향 및 잡음신호만이 존재하는 비음성 구간에서 잔향과 잡음의 스펙트럼을 추정한 뒤 원신호의 스펙트럼에서 차감하여 음성 신호의 품질을 향상시키는 방법이다[3]. 따라서 스펙트럼 차감법의 성능을 높이기 위해서는 음성 및 비음성 구간의 정확한 구분이 요구된다. 기존의 스펙트럼 차감법의 경우 신호의 에너지 임계값 기반의 음성구간 검출법(VAD : Voice Activity Detection)을 적용 하였다. 에너지 임계값 기반의 VAD는 계산량이 많고 잔향시간 및 실내잡음이 증가하는 환경에서 정확한 음성/비음성 구간의 분류가 어렵다. 이는 스펙트럼 차감을 이용한 잔향신호 제거에 성능 열화를 가져오게 된다.

따라서 본 논문에서는 엔트로피(entropy) 기반의 음성 구간 검출법을 스펙트럼 차감법에 적용하여 음성 및 비음성 구간의 분리에 대한 정확도를 높여 효과적으로 잔향을 제거할 수 있는 방법에 대하여 연구 하였다.

본 논문의 구성은 다음과 같다. Ⅱ 장에서는 스펙트럼 차감법에 대하여 설명한다. Ⅲ 장에서는 엔트로피 기반의 VAD에 대하여 설명한다. Ⅳ장에서는 모의실험을 통하여 제안된 방법의 성능을 분석하였다.

Ⅱ. 스펙트럼 차감법

하나의 마이크로폰을 이용한 스펙트럼 차감법은 스펙트럼 향상에 기반한 음성개선 기법으로 주로 신호내에 존재하는 잡음의 제거에 많이 적용되었다. 잡음신호만 존재하는 비음성 구간에서 잡음의 스펙트럼을 추정하고 잡음이 포함된 음성신호의 스펙트럼에서 차감하여 잡음이 제거된 음성신호만 추출하는 방법으로 최근에는 잔향제거에 적용한 연구들도 수행되고 있다.

그림 1의 실내 음향 전달 함수는 일반적으로 직접파와 초기 반사성분 그리고 늦은 잔향성분으로 구성된다. 초기 반사파의 경우 음성신호의 착색현상을 야기하며 늦은 잔향성분은 음성신호의 명료도와 품질을 떨어뜨리는 주된 요인이다.

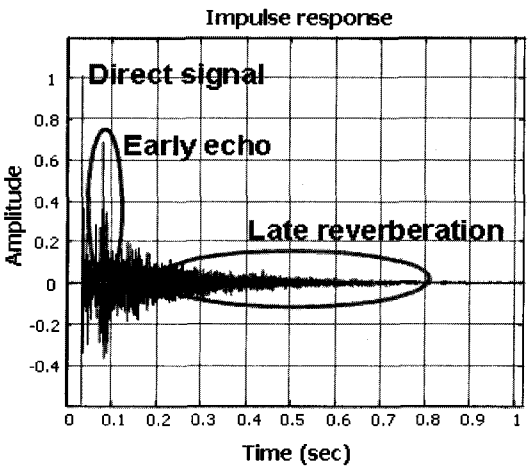

그림 1. 실내 음향 전달 함수
Fig. 1. Acoustic transfer function in a room

일반적으로 늦은 잔향성분의 스펙트럼 특성은 직접파와 초기 반사파 성분을 포함하는 초기 신호의 스펙트럼 특성과 상호 통계적 독립인 것으로 가정한다. 따라서 수신된 음성신호의 스펙트럼 영역에서 잔향성분의 스펙트럼을 제거함으로써 음성신호의 품질을 개선하고자 한다.

그림 2는 스펙트럼 차감법을 이용한 잔향제거 기법의 기본 개념도를 나타낸다. 먼저 잔향이 포함된 수신신호를 주파수 영역에서 분석하고 잔향 성분의 스펙트럼을 추정한다. 그리고 잔향 성분을 억제하는 이득 함수를 구성하여 수신신호의 스펙트럼과 곱한 뒤 시간영역에서 신호를 재구성하면 원하는 음성신호를 추출 할 수 있다.

잔향에 의한 번짐효과(smearing effect)는 입력신호 스펙트럼의 스무딩 현상을 야기시킨다. 따라서 잔향 성분의 전력 스펙트럼은 입력 음성신호 $s(t)$의 전력 스펙트럼이 천이되고 스무딩된 형태로 가정하며 다음과 같이 표현한다[4].

$$|X_r(k;i)|^2 = \gamma\omega(i-\rho) * |S(k;i)|^2 \tag{1}$$

여기서 $|S(k;i)|^2$와 $|X_r(k;i)|^2$은 각각 입력 음성신호와 잔향성분의 단기 전력 스펙트럼을 나타내며, k와 i는 각각 주파수 bin과 시간 프레임 index를 의미한다. $w(i)$는 스무딩 함수를 나타낸다. 지연 인자 ρ는 초기 신호에 대한 잔향성분의 상대적인 지연 시간을 의미하며 일반적으로 실내 음향 전달 함수의 초기 신호와 잔향 간의 상대적인 지연차는 50ms로 나타낸다.

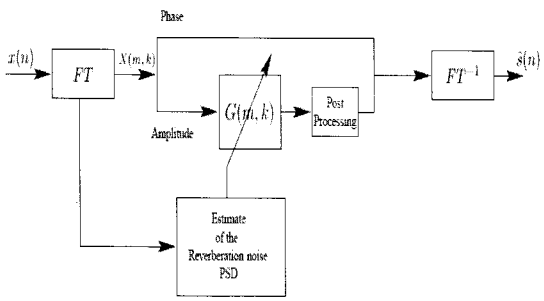

그림 2. 스펙트럼 차감법의 기본 개념도
Fig. 2. Basic diagram of spectral subtraction

스켈링 인자 γ는 잔향성분의 초기신호에 대한 상대적인 음압 강도를 나타내며 본 논문에서는 -10dB를 적용하였다.

마이크로폰에 수신된 신호 $x(t)$는 원래의 음성신호 $s(t)$와 실내 전달함수 $h(t)$의 컨볼루션으로 다음과 같이 표현된다.

$$x(t) = \int_0^{\infty} s(t-\tau)h(\tau)d\tau \tag{2}$$

식 (2)의 전달함수 $h(t)$는 초기 신호성분과 잔향성분으로 구분할 수 있으며 다음과 같이 나타낼 수 있다.

$$x(t) = \int_0^{T_l} s(t-\tau_1)h(\tau_1) + \int_{T_l}^{\infty} s(t-\tau_2)h(\tau_2)d\tau_2 \tag{3}$$

여기서 T_l은 초기 신호성분과 잔향성분을 구분하는 시간 인덱스를 나타낸다. 원 신호 $s(t)$와 전달함수 $h(t)$는 상호 통계적 독립인 관계에 있다고 가정하면 식 (3)은 다음과 같이 전개할 수 있다.

$$E\left[\int_0^{T_l} s(t-\tau_1)h(\tau_1)d\tau_1 \times \int_{T_l}^{\infty} s(t-\tau_2)h(\tau_2)d\tau_2\right] \tag{4}$$
$$= \int_0^{T_l}\int_{T_l}^{\infty} E[s(t-\tau_1)s(t-\tau_2)]E[h(\tau_1)h(\tau_2)]d\tau_2 d\tau_1$$

시간상으로 간격을 가지는 초기신호와 잔향신호는 비상관 특성을 가지므로, $\tau_1 - \tau_2$이 상대적으로 클 경우 $E[s(t-\tau_1)s(t-\tau_2)] \approx 0$을 만족한다.

결국 전달함수 $h(t)$의 초기 신호성분과 잔향성분은 서로 비상관 특성을 가지므로 잔향성분이 제거된 초기 신호성분의 전력 스펙트럼은 수신 신호의 전력 스펙트럼에서 잔향성분의 전력 스펙트럼을 차감함으로써 추정할 수 있다. 잔향 성분이 제거된 원 신호 전력 스펙트럼의 추정치는 수신신호의 전력 스펙트럼에 잔향 스펙트럼의 억압 이득 함수(Gain function)를 곱하여 다음과 같이 구한다.

$$|\hat{S}(k;i)|^2 = |X(k;i)|^2 G(k;i) \tag{5}$$

여기서 $|\hat{S}(k;i)|^2$ 는 원하는 음성신호의 파워 스펙트럼에 대한 추정치를 나타내고 $G(k;i)$는 잔향제거 이득 함수를 의미하며 다음과 같이 표현한다.

$$G(k;i) = \max\left[\frac{|X_r(k;i)|^2 - \gamma w(i-\rho) * |X_r(k;i)|^2}{|X_r(k;i)|^2}, \epsilon\right] \quad (6)$$

ϵ는 최대감쇠 인자를 나타내며, 본 논문에서는 30dB로 적용하였다.

스펙트럼 차감법을 이용한 음성개선의 성능을 높이기 위해서는 비음성 구간에 존재하는 잔향의 제거가 필수적이다. 따라서 음성과 비음성 구간의 정확한 분류가 필요하며 이를 위한 entropy 기반의 음성 구간 검출법을 다음 장에서 설명한다.

Ⅲ. Entropy 기반의 음성 구간 검출

음성구간 검출(VAD)은 수신된 음성신호를 음성 구간과 비음성 구간으로 구분하는 과정으로 음성신호 처리의 여러 분야에 적용되는 기법이다. 음성 구간 검출 기법은 크게 신호의 에너지(energy), 영교차율 (zero crossing rate), LPC 파라메터와 같은 시간영역의 특징을 이용하는 방법과, Likelihood Ratio 등과 같은 통계적인 특성에 기반한 방법들로 나눌 수 있다.

특히 기존의 스펙트럼 차감법에 적용되던 에너지 임계값 기반의 VAD는 계산량이 많고 잔향시간 및 실내잡음이 증가하는 환경에서 정확한 음성/비음성 구간의 분류가 어렵다. 따라서 본 논문에서는 entropy에 기반한 VAD를 적용하여 음성/비음성 구간 검출의 정확성을 높이고자 하였다.

Spectral entropy는 entropy 이론을 음성 신호 처리에 적용한 것으로 잡음 환경에서 신호의 에너지 임계값이나 영교차율 등을 이용한 음성 검출법의 문제점을 개선하기 위한 방법으로 사용되고 있다[5][6].

Spectral entropy를 구하기 위해서는 먼저 고속 푸리에 변환(FFT)를 이용하여 시간 영역의 음성신호를 주파수 영역으로 변환하며 다음과 같이 나타낸다.

$$X(k;i) = \sum_{n=0}^{N-1} s(n;i)\exp(-j\frac{2\pi kn}{N}) \quad (7)$$

여기서 $X(k;i)$는 i번째 프레임의 k번째 주파수 bin을 나타내고 N은 FFT포인트의 개수를 의미한다. Noise의 간섭을 최소화 하고자 대역 통과 필터를 적용하여 spectral entropy를 추정하는 범위를 음성대역이 존재하는 중간 주파수 대역인 350Hz~3000Hz로 제한하였다.

스펙트럼의 확률 질량함수(PMF)는 주파수 성분의 정규화 방법을 사용하여 다음과 같이 구할 수 있다.

$$p(k;i) = \frac{|X(k;i)|}{\sum_{m=0}^{N-1}|X(m;i)|} \quad (8)$$

여기서 k=1,...,N-1 이다. 최종적으로 구해진 spectral entropy는 다음과 같이 나타낸다.

$$H(i) = -\sum_{k=0}^{N-1} p(k;i)\log(p(k;i)) \quad (9)$$

식 (8)의 추정된 spectral entropy의 급격한 변화를 막기 위하 $H(i)$에 대하여 5차 median filter를 적용하였고 이를 초기 임계값과 비교하여 음성 비음성 구간을 구별하였다. Spectral entropy 기반의 음성 구간 검출법은 특히 잡음이 존재하는 환경에서 기존의 에너지 기반의 음성 구간 검출법보다 효율적이다.

Ⅳ. 모의실험 및 결과

본 논문에서 제안하는 spectral entropy 기반의 VAD를 적용한 스펙트럼 차감법의 성능을 분석하기 위하여 모의실험을 수행하였다.

원래의 음성신호와 실내에서 실험을 통해 획득한 실내 음향 전달 함수 사이의 컨볼루션을 통하여 잔향을 가지는 음성신호를 모의하였다. 음향 전달 함수의 잔향시간은 각각 0.3초, 0.78초, 1.3초 이다. 실험에 사용된 음성 데이터는 여성의 음성으로 샘플링 주파수는 16KHz로

설정하였으며 12초의 길이를 가진다.

그림 3은 spectral entropy 기반의 VAD에 의한 음성 구간 검출 결과를 보여준다. 그림 3(a)는 원래의 음성신호에 대한 신호 파형을 나타내며 그림 3(b)는 spectral entropy의 추정치를 나타낸다. 그림 3(c)는 음성/비음성 구간의 판정 결과를 나타내며 긴 비음성 구간 뿐만 아니라 음성과 음성 사이의 짧은 비음성 구간에 대해서도 정확한 분류가 이루어짐을 확인 할 수 있다.

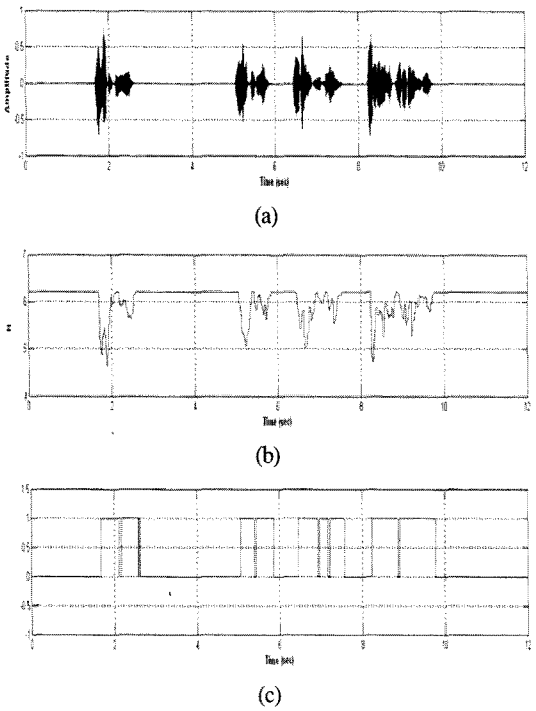

그림 3. Spectral entropy 기반의 음성구간 검출 결과
(a) 원래의 음성신호 (b) Spectral entropy 추정치
(c) 음성/비음성 구간 판정결과
Fig. 3. Results of VAD based on spectral entropy
(a) Original speech signal (b) Estimate of spectral entropy (c) Detection result of voice and silence

잡음이 존재하는 환경에서 제안된 방법의 성능을 분석하기 위해 수신 음성신호에 부가 잡음을 첨가하였다. SNR이 20dB이고 0.3초의 잔향시간을 가지는 음성신호에 대하여 스펙트럼 차감법을 적용한 결과를 그림 4에 나타내었다.

그림 4(a)는 잔향 및 잡음이 부가된 수신 음성신호를 나타낸다. 그림 4(b)에 나와있는 기존의 에너지 임계치 기반의 VAD를 적용한 스펙트럼 차감결과의 경우 음성/비음성 구간의 정확한 검출이 이루어지지 않아 비음성 구간에 여전히 잡음 및 여분의 잔향성분이 포함되어 있음을 확인 할 수 있다. 또한, 비음성 구간의 오검출로 인하여 원래 신호성분이 존재하는 구간에 대하여 신호를 상쇄시키는 결과가 나타났다. 그림 4(c)에 나와있는 제안된 기법을 적용한 결과의 경우 비음성 구간에서의 잡음 및 잔향 성분의 제거가 효과적임을 알 수 있고, 기존의 방법에 비해 원 신호성분을 감쇠시키는 효과도 적게 나타났다.

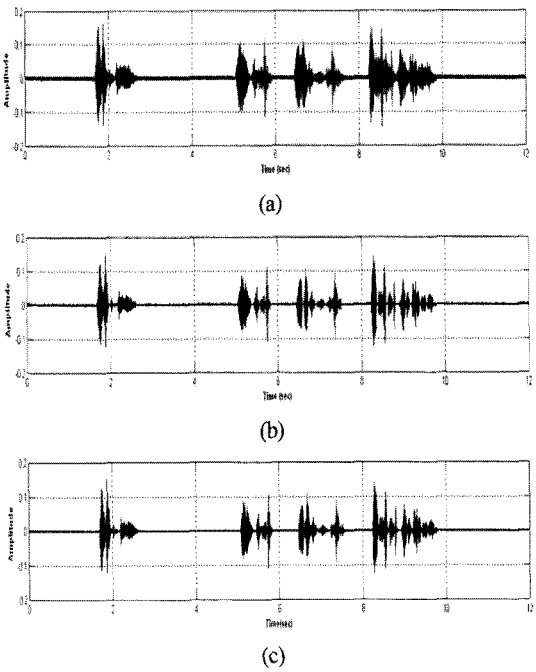

그림 4. 스펙트럼 차감 결과
(a) 잔향 음성신호 (b) 기존의 스펙트럼 차감 결과 (c) 제안된 방법을 적용한 결과
Fig. 4. Results of spectral subtraction
(a) Reverberant speech signal (b) Result of conventional method (c) Result of proposed method

실제 귀로 들었을 경우 기존의 방법을 적용한 결과는 원 신호 성분의 감쇠로 인해 음성신호의 프레임간 연결이 부자연스러웠지만 제안된 방법의 경우는 기존의 방법에 비해 연결이 자연스럽고 명료도 또한 개선되었음

을 확인할 수 있었다.

제안된 방법의 객관적인 평가 지표로써 다음과 같이 비음성 구간에 대한 잔향제거 비율 (Reverberation Reduction Ratio)을 사용하여 성능을 분석하였다.

$$RR = 10\log_{10}\left(\frac{\sum_{n\in\Omega Silence} x(n)^2}{\sum_{n\in\Omega Silence} \hat{s}(n)^2}\right) \quad (10)$$

표 1은 SNR 및 잔향시간에 따른 두 가지 방법의 잔향 제거비를 나타내고 있다. 잡음이 존재하는 환경에서 잔향시간의 증가에 대해서도 제안된 기법을 잔향신호에 적용한 경우 기존의 방법보다 나은 음성 개선 성능을 보여주고 있다.

표 1. SNR 및 잔향시간에 따른 잔향제거 비율
Table. 1. Reverberation reduction ratio according to SNR and reverberation time.

SNR(dB)	Method	Reverberation time(sec)		
		0.3s	0.78s	1.3s
10dB	Conventional	9.4dB	10.6dB	9.6dB
	Proposed	23.5dB	15.1dB	13.7dB
20dB	Conventional	9.4dB	10dB	9.4dB
	Proposed	19.2dB	25.6dB	25dB
30dB	Conventional	9.3dB	10.2dB	9.5dB
	Proposed	11dB	17.7dB	22dB

V. 결론

본 논문에서는 실내에서 잔향이 포함된 음성신호의 품질을 개선하기 위해 spectral entropy 기반의 VAD를 적용한 스펙트럼 차감법에 대하여 연구하였다. 기존의 에너지 임계치 기반의 VAD를 적용한 스펙트럼 차감법에 비해 보다 정확하게 음성 구간을 검출하여 잔향제거 성능을 향상시켰다. 모의실험 결과 잡음이 존재하는 실내 환경에서도 잔향시간의 변화에 따라 잔향제거 비율이 평균 19dB인 우수한 성능을 나타내었다.

향후 다양한 실내환경 및 기타 다른 음향신호에 대한 적용성 검토가 필요하며 잔향 스펙트럼을 효과적으로 추정할 수 있는 방법에 대한 연구가 필요하다.

참고문헌

[1] J. B. Allen, D. A. Berkley and J. Blauert, "Multi microphone signal processing technique to remove room reverberation from speech signals, " *J. Acoust. Soc. Amer.*, vol. 62, no. 4, pp. 912-915.

[2] D. Bees, M. Blostein and P. Kabal, "Reverberant speech enhancement using cepstral processing, " *in proc. IEEE Int. Conf. Acoustics, Speech Signal Processing,* pp. 977-980, 1991.

[3] K. Lebart and J.M. Boucher, "A New Method Based on Spectral Subtraction for Speech Dereverberation," *Acta Acoustica,* vol. 87, pp. 359-366, 2001.

[4] Mingyang Wu and DeLiang Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. Speech Audio Process.*, Vol. 14, no. 3, pp. 774 - 784, May 2006.

[5] R. V. Prasad, R. Muralishankar and S. Vijay, "Voice Activity Detection for VoIP-An Information Theoretic Approach," *in proc. IEEE Int. Conf. Telecommunications,* pp. 1-6, 2006.

[6] 노용완, 이규범, 이우석, 홍광석, "차량 잡음 환경에서 엔트로피 기반의 음성 구간 검출," 한국 신호처리 시스템 학회 논문지, 제19권, 제2호, pp.121-128, 2008.

저자소개

김세영(Se Young Kim)

2005년 2월 한국해양대학교 전파공학과 (공학사)
2007년 2월 한국해양대학교 전파공학과 (공학석사)
2007년 3월 ~ 현재 한국해양대학교 전파공학과 박사 과정
※관심분야 : DSP, 어레이 신호처리

강석엽(Suk-Youb Kang)

1997년 2월 인천대학교 전자공학과 (공학사)
1999년 2월 인하대학교 전자공학과 (공학석사)
2005년 8월 인하대학교 전자공학과 (공학박사)
2001년 7월 ~ 2006년 8월 (주)아이엔텍 대표이사
2006년 8월 ~ 현재 한국해양대학교 BK21 연구교수

김기만(Ki-man Kim)

1988년 2월 연세대학교 전자공학과 (공학사)
1990년 8월 연세대학교 전자공학과 (공학석사)
1995년 2월 연세대학교 전자공학과 (공학박사)
1995년 3월~ 1996년 8월 연세대학교 의과대학 의용공학교실 (Fellow)
1996년 9월~현재 한국해양대학교 전파공학과 교수
※관심분야 : 수중통신, 소나 신호처리, DSP