
주파수 영역 자기 공분산 기울기를 이용한 음성과 자동차 소음 신호의 구분

김선일*

Classification of Speech and Car Noise Signals using the Slope of Autocovariances
in Frequency Domain

Seonil Kim*

요 약

음성 신호와 자동차 엔진 배기음 등의 소음이 섞인 신호에서 통계적 방법을 이용하여 음성 신호와 자동차 소음 신호를 분리하였다. 분리된 신호에서 음성신호를 구분해 내기 위해 128개의 원소를 갖는 신호 조각의 연속으로 신호를 재구성하고 각 신호 조각에 대해 FFT를 구하였다. 각 신호 조각의 FFT 계수 중에서 저주파 영역의 일부 계수 중 계수 각각에 대해 각 신호 조각 사이의 자기 공분산을 구하고 이들을 평균하였다. 그리고 linear regression을 이용하여 평균 자기 공분산 값들을 연결하는 직선의 방정식을 구한 후 이 직선의 기울기를 비교하여 음성 신호와 자동차 소음 신호를 구분하는 방법을 제안하고 유용성을 확인하였다.

ABSTRACT

Speech signal and car noise signal such as muffler noise are segregated from the one which has both signals mixed using statistical method. To classify speech signal from the other in segregated signals, FFT coefficients were obtained for all segments of a signal where each segment consists of 128 elements of a signal. For several coefficients of FFT corresponding to the low frequencies of a signal, autocovariances are calculated between coefficients of same order of all segments of a signal. Then they were averaged over autocovariances. Linear equation was established for the those autocovariances using the linear regression method for each signal. The coefficient of the slope of the line gives reference to compare and decide what the speech signal is. It is what this paper proposes. The results show it is very useful.

키워드

음성 신호, ICA, 자기 공분산 기울기, 신호 분리

Key word

Speech Signal, ICA, Slope of Autocovariances, Segregation of Signals

* 정회원 : 거제대학교(교신저자, seonil@koje.ac.kr)

접수일자 : 2011. 08. 12

심사완료일자 : 2011. 09. 02

I. 서 론

음성 인식 기술이 그 동안 많은 발전을 거듭하여 왔지만 실생활에 그다지 많이 활용되지 않다가 요즘에서야 대중적으로 많이 이용되고 있다. 음성 인식 기술이 그동안 많이 활용되지 못한 것은 연구실에서는 높은 인식율을 보이지만 주변 잡음이 많은 실생활에서는 인식율이 현저히 떨어지기 때문이다.

배경 잡음을 제거하는 방법에는 필터를 이용하여 잡음이 존재하는 영역의 데이터를 제거하는 방법이 있으나 잡음과 함께 해당 영역의 데이터도 손실을 입는다. 하지만 통계적인 방법을 이용하면 섞여있는 신호를 서로 독립적으로 분리해 낼 수 있다.

대표적인 통계적인 방법으로는 BSS(Blind Source Separation)[1]를 들 수 있는데 BSS를 실현하는 방법으로 섞이기 이전에 non-Gaussianity가 강했던 신호들이 서로 섞이게 되면 Gaussianity가 강해지는 경향이 있다는 특성을 이용하여 non-Gaussianity가 최대가 되는 방향으로 신호를 분리해 낸다. 이렇게 서로 독립적인 성분을 최대화하는 방향으로 신호를 분리해 내는 방법을 ICA(Independent Component Analysis)[2] 라고 하며 non-Gaussianity를 측정하는 방법에는 Negentropy[3], Mutual Information[4], Kurtosis[5], Maximum Likely Estimation[6] 등이 있다.

자동차와 같이 원천적으로 소음이 존재할 수 밖에 없는 곳에서 ICA를 이용해 음성 인식율을 높이려는 여러 가지 시도를 발견할 수 있다.[7-9]

ICA에서는 신호원 보다 많거나 같은 수의 마이크를 사용했을 경우에 신호를 분리해 낼 수는 있지만 그 신호 중의 어느 신호가 우리가 원하는 신호인지는 알 수 없다. 자동차에서 나오는 소음과 음성 신호, 2개의 신호를 가정하면 두 개의 마이크를 사용하여 신호를 취득하여 ICA를 이용하여 신호를 분리해 낼 수 있다. 이 때 어느 신호가 음성 신호인지 알아 내기 위해서는 각 신호의 특성을 이용할 필요가 있다.

음성 신호 중에 모음의 일부를 이용하여 이 모음과 분리된 두 신호간의 상관 관계를 알아내어 두 신호를 구분하는 방법이 의미있는 분류 결과를 보여주었지만[10] 어떤 모음을 사용하느냐에 따라 인식율의 차이가 나타나고 적절한 모음을 선정하는 기준이 없다.

자동차 소음과 음성 신호는 신호 간격에 따른 공분산을 살펴보았을 때 시간 영역과 Quefrency 영역에서는 크게 다른 특성을 발견할 수 없지만 주파수 영역의 공분산에서는 공분산의 기울기에서 현저한 격차를 발견할 수 있었다[11]. 따라서 총 81개 문장의 뉴스 음성 신호에 가속과 정속 주행에서의 자동차 소음을 인위적으로 섞인 마이크와 다른 마이크 사이에 2dB 정도 차이가 나게끔 섞은 후 ICA를 이용하여 이들을 분리해 내어 주파수 영역 자기 공분산 기울기를 이용하여 음성 신호와 자동차 소음으로 분류해 내었다.

II. 주파수 영역 자기 공분산

공분산은 식(1)과 같이 나타낼 수 있다. 신호 $x(t)$ 에 대한 자기 공분산 $c_x(t, \tau)$ 는

$$c_x(t - \tau) = E[x(t) - m_x(t)][x(t - \tau) - m_x(t - \tau)] \quad (1)$$

$m_x(t)$ 는 x_t 의 평균값 이고 $x(t - \tau)$ 는 x_t 에서 τ 만큼 이동한 값이다.

자기 공분산을 주파수 영역에서 적용하기 위해서 자동차 소음이나 음성 신호를 128개의 원소를 갖는 신호 조각의 연속으로 이루어지도록 나눈다. 따라서 1280개의 데이터로 이루어진 음성 신호는 10개의 신호 조각을 갖게 된다. 음성 신호 조각의 원소 개수를 128로 한 것은 256개나 64개 보다 자기 공분산 특성이 더 잘 나타났기 때문이다. 각 신호 조각을 FFT 처리하여 주파수 성분을 계산한다. 이는 음성 신호의 SFFT(Short time FFT)와 유사한데 신호군끼리 겹치지 않도록 하는 점만 다르다. 신호군끼리 일부 겹치게 되면 신호군끼리 유사성이 많아 자기 공분산을 구하는 의미가 퇴색하게 된다. 특히 자동차 소음에서 심하다. 각 신호 조각의 FFT 값을 열벡터로 나타내면 한 신호 전체는 식(2)와 같은 행렬로 표현될 수 있다. 여기서 m 은 128이고 n 은 신호의 크기에 따라 달라진다. 1280개의 데이터로 이루어진 신호는 $n = 100$ 이 된다.

여기서 첫 번째 행벡터는 직류 성분을 표시하므로 자기 공분산 계산에서 제외하고 나머지 k 개의 행벡터

에 대해 각 벡터 사이의 자기 공분산을 구한다. 음성에 대한 정보는 저주파에 많이 포함되어 있으므로 저주파 영역의 k 개 주파수에 대해 자기 공분산을 계산하였다. k 는 10으로 설정하였다. 물론 10개 이상도 할 수 있겠지만 계산량만 늘어나고 성능의 개선은 기대할 수 없었다. 자기 공분산을 특정 주파수에 둔감하게 하기 위해 10개 주파수에 대해 계산한 자기 공분산의 평균을 사용하였다.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \dots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{bmatrix} = [A_1 \ A_2 \ \dots \ A_m] \quad (2)$$

식 (2)에서 i 번째 행벡터는

$$A_i = [a_{1,i} \ a_{2,i} \ \dots \ a_{n,i}]^T$$

가 되고 A 에 대한 자기 공분산을 식(3)과 같이 구한다.

$$C = [AutoCov(A_1) \ AutoCov(A_2) \ \dots \ AutoCov(A_m)] = [C_1 \ C_2 \ \dots \ C_m] \quad (3)$$

여기서 $AutoCov$ 는 자기 공분산을, C_i 는

$$C_i = \begin{bmatrix} c_i(-\tau) \\ c_i(-\tau+1) \\ \vdots \\ c_i(-1) \\ c_i(0) \\ c_i(1) \\ \vdots \\ c_i(\tau-1) \\ c_i(\tau) \end{bmatrix}$$

$c_i(0)$ 는 항상 1이 되고 이를 중심으로 양쪽이 서로 대칭이다. $\tau=0$ 부터 $\tau=\pm 10$ 까지 구한다. τ 가 너무 작으면 자기 공분산의 기울기를 구하는 데이터가 모자라고 10보다 커도 추가 정보가 없어서 적절히 10개로 정하였다.

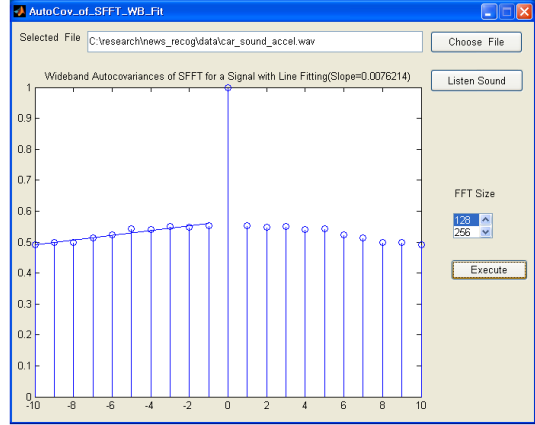


그림 1. 자동차 소음 신호의 주파수 영역 자기 공분산
Fig. 1. Autocovariances of a Car Noise Signal in Frequency Domain

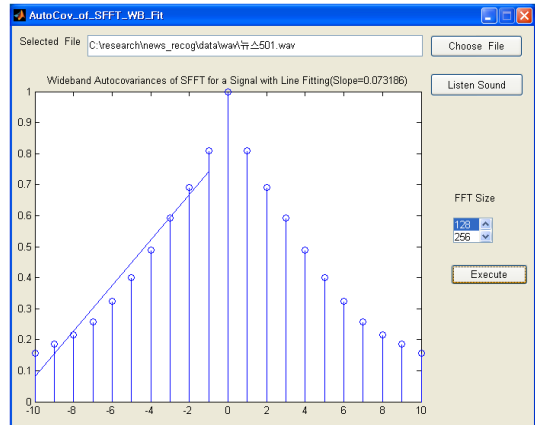


그림 2. 음성 신호의 주파수 영역 자기 공분산
Fig. 2. Autocovariances of a Speech Signal in Frequency Domain

k 개 주파수 구간의 평균 C_{avr} 을 구하면

$$C_{avr} = Avr(C_i) = \frac{1}{k} \sum_{i=1}^k C_i = \begin{bmatrix} c_{avr}(-\tau) \\ c_{avr}(-\tau+1) \\ \vdots \\ c_{avr}(0) \\ \vdots \\ c_{avr}(\tau-1) \\ c_{avr}(\tau) \end{bmatrix} \quad (4)$$

특정 신호에 대한 C_{avr} 값, 즉 $c_{avr}(-10)$ 부터 $c_{avr}(10)$ 까지의 값이 그림 1과 그림 2에 나타나 있다. 그림 1은 자동차를 가속할 때 나타나는 소음 신호에 대한 C_{avr} 값이고 그림 2는 음성 신호에 대한 C_{avr} 값이다.

III. Linear Regression

그림 1이나 그림 2를 보면 $c_{avr}(0)$ 를 제외한 나머지 값의 기울기가 완전히 차이가 난다. 따라서 최소 자승 오류를 갖는 직선의 방정식을 구하면 이 값들의 기울기를 알 수 있고 이 기울기의 차이를 이용하여 음성 신호와 자동차 소음을 구분할 수 있다.

그림 1과 그림 2에는 직선이 그어져 있다. 이는 $c_{avr}(-10)$ 에서 $c_{avr}(-1)$ 까지의 공분산 값을 이용해 이 값들을 지나는 직선의 방정식을 구하되 오차의 제곱이 최소가 되는 직선의 방정식을 구하여 그은 것이다. 이 직선의 방정식은 식 (5)와 같은 Linear Regression[13]을 이용하여 구할 수 있다.

$$y = \beta_0 + \beta_1 x + \epsilon \tag{5}$$

이 식을 식 (6)과 같이 벡터로 표현할 수 있다.

$$Y = X\beta + E \tag{6}$$

여기서

$$Y = [y_1 \ y_2 \ \dots \ y_n]^T$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$\beta = [\beta_0 \ \beta_1]^T$$

$$E = [\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_n]^T$$

식(6)에서 $[x_1 \ x_2 \ \dots \ x_n]^T$ 로는 τ 의 변화 즉 -10부터 -1까지 10개의 값을 사용하고 해당되는 τ 에서의 자기 공분산 평균값이 Y 에 해당되며 이로부터 구해진 $\beta_0 \ \beta_1$ 이 각각 해당 값들을 최소 자승 오류로 연결하는 직선의 절

편 및 기울기이다.

IV. 실험 결과

사용된 음성은 남자 아나운서의 뉴스 음성 81개 문장이고 이 신호들에 자동차가 가속할 때 나오는 소음과 정속 주행할 때 나오는 소음을 각각 섞어서 실험 대상 신호로 이용하였다.

두 신호를 분리하기 위해 Negentropy를 이용한 FastICA를 사용하였다[9]. 실험에서는 자동차 소음의 외 다른 소음이 없다고 가정하였으며 실제 상황처럼 2개의 마이크를 사용하여 녹음하지 않고 두 신호를 인위적으로 두 마이크 사이에 2dB 차이가 나도록 섞었다.

ICA로 자동차 소음과 음성을 분리해내고 분리된 각 신호를 128개의 조각으로 나누고 이 조각으로 FFT를 한 후 저주파에 해당되는 10개의 주파수 영역에 대해 각 주파수 영역별로 각 조각들 사이의 자기 공분산을 구하고 각 주파수 영역별로 구한 자기 공분산을 평균한 후 이 값에서 τ 의 변화에 따른 자기 공분산 값들을 연결하는 직선의 기울기를 Linear Regression으로 구하여 두 신호를 분류하는데 사용하였다.

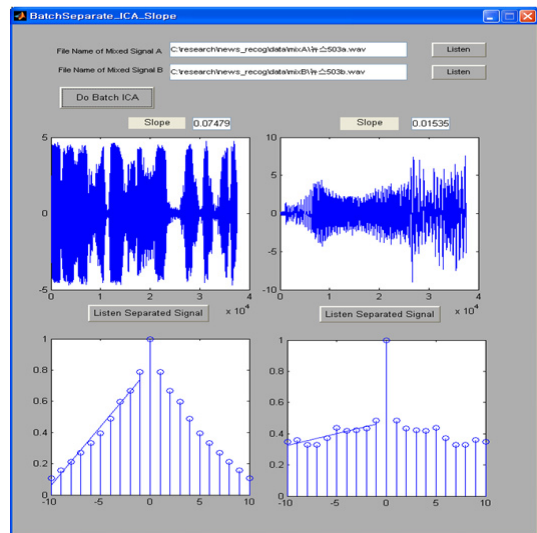


그림 3. 분리된 신호와 주파수 영역 자기 공분산
Fig. 3. Segregated Signals and their Autocovariances in Frequency Domain

81개 신호 중 하나에 대해서 주파수 영역 자기 공분산을 구하고 이 공분산 값들을 연결하는 직선의 방정식을 구하여 공분산 값들에 중첩시켜 그림 3에 그래프로 나타내었다.

그림 3의 위 좌측 그래프는 음성 신호이고 오른쪽은 자동차 소음이다. 아래 좌측 그래프는 음성 신호의 주파수 영역 자기 공분산 값들과 그것들을 연결한 직선이고 오른쪽 그래프는 자동차 소음에 대해 구한 주파수 영역 자기 공분산 값들과 그것들을 연결하는 직선이다.

그림 1과 그림 2는 섞이기 전의 두 신호에 대한 주파수 영역 자기 공분산이고 그림 3은 섞인 신호를 분리해 낸 후의 신호들과 주파수 영역 자기 공분산 그래프이다. 음성 신호에 대한 주파수 영역 자기 공분산 값들과 자동차 소음에 대한 주파수 영역 자기 공분산 값들의 직선의 기울기가 현저한 차이를 보이고 있다. 자동차 소음이 섞인 81개 음성 신호에 대해 ICA로 신호들을 서로 분리해 내고 분리된 신호들에 대한 주파수 영역 자기 공분산 값들을 구하여 이를 연결하는 직선의 기울기 값을 그림 4에 나타내었다.

그림 4의 (a)와 (b)는 각각 음성 신호와 가속중인 자동차 소음 신호의 평균 에너지가 동일하도록(0dB), 음성 신호의 평균 에너지가 자동차 소음 신호에 대해 훨씬 작도록(-40dB) 섞은 다음에 이를 ICA로 분리해 내고 두 신호의 주파수 영역 자기 공분산 기울기를 이용하여 분류해 낸 결과이다. 그림 4의 (c)와 (d)는 일정한 속도로 달리고 있을 때의 결과이며 자동차 소음에 대한 자기 공분산 기울기는 0에 가까워서 그래프 바닥에 그려져서 그리지 않은 것처럼 보인다. 어느 경우든지 81개 신호에 대해 음성 신호와 자동차 소음에 대한 자기 공분산 기울기의 차이가 현격하므로 성공적으로 분리해 낼 수 있음을 그래프를 통해 확인할 수 있다.

각 경우에 대한 주파수 영역 자기 공분산의 81개 신호의 기울기의 평균은 표 1과 같다.

표 1. 81개의 분리된 신호에 대한 자기 공분산 기울기의 평균

Table 1. The averages of values of slopes for the separated 81 signals

상태	음원	0dB	-40dB
가속	음성	0.0788	0.0789
	소음	0.0116	0.0114
정속	음성	0.0786	0.0787
	소음	0.0000	0.0001

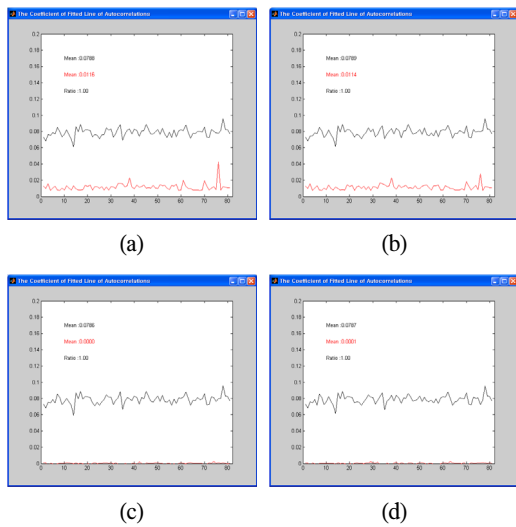


그림 4. 분리된 81개 신호 쌍에 대한 분류 결과
(a) 0dB, 가속 (b) -40dB, 가속
(c) 0dB, 정속 (d) -40dB, 정속

Fig. 4. Results of classification for the 81 segregated signal pairs
(a) 0dB, acceleration (b) -40dB, acceleration
(c) 0dB, constant-speed (d) -40dB, constant-speed

V. 결 론

ICA를 이용해 음성과 자동차 소음을 분리하고 분리된 신호로부터 주파수 영역 자기 공분산 값들의 기울기를 이용하여 자동으로 음성 신호와 자동차 소음을 분류하는 방법을 제안하였다.

음성 신호와 자동차 소음 신호를 분리한 후 들어보면 -40dB로 섞인 신호를 분리하였을 때 음성 신호에는 자동차 소음이, 자동차 소음에는 음성 신호가 약간씩 들렸다. 하지만 주파수 영역 자기 공분산 기울기는 섞기 이전이나 섞은 신호를 분리한 후에도 거의 변화가 없었다.

음성 신호는 SFFT 스펙트로그램을 보면 주파수가 급격히 변화하지 않고 인접 신호와 상당한 연관성을 갖고

변하는 것을 관찰할 수 있다. 따라서 특정 주파수 대역에서 자기 공분산을 구하면 공분산 간격이 증가할 수록 값이 서서히 감소하는 현상이 나타난다[그림1]. 이러한 현상은 음성의 특정 영역에서 특히 강하게 나타난다. 유성음 신호의 에너지 대부분이 집중되어 있는 저주파 영역은 주로 성대가 울릴 때 나는 신호로 구성되므로 이런 현상이 유독 강하게 나타난다.

반면 자동차 소음에서는 동일한 기계음의 반복이므로 신호의 어느 부분에서나 주파수 성분이 일정하다고 볼 수 있다. 따라서 주파수 영역에서 인접하는 신호 조각끼리의 자기 공분산을 구해보면 어느 지점에서나 일정한 값을 나타낸다[그림 2]. 이러한 현상으로 인해 자기 공분산의 기울기 차이가 발생되고 이 기울기를 구하면 두 신호를 구별해낼 수 있다. 현재 실험 결과는 주어진 문장 81개에 대해 모두 구별해내는 성능을 발휘하고 있다.

앞으로는 도로 주행중에 실제로 발생하는 신호를 수집하여 이 방법을 적용해 보고 그 유용성을 확인하고자 한다. 실제 도로 주행 중에는 자동차 소음 뿐만 아니라 주변 잡음도 들어오게 되므로 이로 인해 두 신호를 분리하는데 어려움을 겪을 수도 있을 것이다. 또한 세 개 이상의 신호를 분리했을 때 주파수 영역에서의 자기 공분산 기울기를 적용하면 어떤 문제가 발생하고 이를 어떻게 해결하여야 할지도 도전해야 할 과제이다.

참고문헌

- [1] J. F. Cardoso, "Blind signal separation: statistical principles," Proc. IEEE, vol. 9, no. 10, pp. 2009-2-25, Oct., 1988.
- [2] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, no. 4/5, pp. 411-430, 2000.
- [3] A. Hyvarinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," IEEE Trans. On Neural Networks, vol. 10, no. 2, pp. 626-634, May, 1999.
- [4] P. Conon, "Independent component analysis, A new concept?," Signal Processing, vol. 36, pp. 287-314, 1994.
- [5] J. LeBlanc and P. Leon, "Speech Sепation by Kurtosis Maximization," Proc. ICASSP, vol. 2, pp. 1029-1032, 1998.
- [6] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," Neural Computation, vol. 7, no. 6, pp. 1129-1159, 1995.
- [7] H. Saruwatari, K. Sawai, T. Nishikawa, A. Lee, K. Shikano, A. Kaminuma, M. Sakata and D. Saitoh. "Speech Enhancement Based on Blind Source Separation in Car Environments," Proc. 21st International Conference on Data Engineering." pp. 1205, 05-08 April, 2005.
- [8] J. Lee, H. Jung, T. Lee and S. Lee, "SPEECH CODING AND NOISE REDUCTION USING ICA-BASED SPEECH FEATURES," International Workshop on independent component analysis and blind signal separation, pp. 417-422, 19-22 June, 2000, Helsinki, Finland.
- [9] 김선일, "잡음 섞인 한국어 인식을 위한 ICA 비교 연구," 대한전자공학회 논문지, 제45권, IE-3호, 9월, 2008.
- [10] 김선일, "ICA로 분리한 신호의 분류," 대한전자공학회 논문지, 제47권, IE-4호, 12월, 2010.
- [11] 김선일, "음성 및 음성 관련 신호의 주파수 및 Quefrency 영역에서의 자기공분산 변화," 한국해양정보통신종합학술대회논문집, 춘계15권 1호, pp. 340-343, 대구 EXCO, 2011.
- [12] Sophocles J. Orfanidis, *Introduction To Signal Processing*, Prentice Hall Inc., 1996.
- [13] R. Johnson, K. Tsui, *Statistical Reasoning and Methods*, John Wiley & Sons, Inc. 1998.

저자소개



김선일(Seonil Kim)

1983년 아주대학교 전자공학과
공학사

1985년 아주대학교 전자공학과
공학석사

1996년 아주대학교 전자공학과 공학박사

1985년~1990년 한국기계연구원 선임연구원

1990년~현재 거제대학교 교수

1997년 Visiting Professor, CAIP Center, Rutgers Univ.,
Piscataway, N.J., USA

2007년 Visiting Professor, Dept. of ECE, Georgia
Institute of Technology, Atlanta, G.A., USA

※ 관심분야: 신호 처리, 음성 인식