JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 20, No. 9 : 1673~1678 Sep. 2016

개량된 음성매개변수를 사용한 지속시간이 짧은 잡음음성 중의 배경잡음 분류

최재승*

Background Noise Classification in Noisy Speech of Short Time Duration Using Improved Speech Parameter

Jae-Seung Choi*

Department of Electronic Engineering, Silla University, Busan 46958, Korea

요 약

음성인식처리 분야에서 배경잡음으로 인하여 음성입력이 배경잡음으로 잘못 판단되는 원인이 되어 음성인식율의 저하를 초래한다. 이러한 종류의 잡음대책은 단순하지 않으므로 보다 고도한 잡음처리기술이 필요하게 된다. 따라서 본 논문에서는 잡음환경 중에서 정상적인 배경잡음 혹은 비정상적인 배경잡음과 지속 시간이 짧은 음성을 구별하는 알고리즘에 대하여 기술한다. 본 알고리즘은 다른 종류의 잡음과 음성을 구별하는 중요한 수단으로서 개량된 음성의 특징파리미터를 사용한다. 다음으로 다층퍼셉트론 네트워크에 의하여 잡음의 종류를 추정하는 알고리즘에 대해서 기술한다. 본 실험에서는 잡음과 음성이 구별이 가능하도록 실험적으로 확인하였다.

ABSTRACT

In the area of the speech recognition processing, background noises are caused the incorrect response to the speech input, therefore the speech recognition rates are decreased by the background noises. Accordingly, a more high level noise processing techniques are required since these kinds of noise countermeasures are not simple. Therefore, this paper proposes an algorithm to distinguish between the stationary background noises or non-stationary background noises and the speech signal having short time duration in the noisy environments. The proposed algorithm uses the characteristic parameter of the improved speech signal as an important measure in order to distinguish different types of the background noises and the speech signals. Next, this algorithm estimates various kinds of the background noises using a multi-layer perceptron neural network. In this experiment, it was experimentally clear the estimation of the background noises and the speech signals.

키워드 : 음성인식, 정상잡음, 비정상잡음, 잡음환경, 개량된 음성파라미터

Key word : Speech Recognition, Stationary Noise, Non-Stationary Noise, Noisy Environment, Improved Speech Parameter

Received 03 May 2016, **Revised** 12 May 2016, **Accepted** 27 May 2016

* **Corresponding Author** Jae-Seung Choi (E-mail:jschoi@silla.ac.kr, Tel:+82-51-999-5608)
Department of Electronic Engineering, Silla University, Busan 46958, Korea

Open Access http://dx.doi.org/10.6109/jkiice.2016.20.9.1673 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. 서 론

근래 고도의 정보화 사회의 발전과 함께 개인의 신원 확인 및 인간과 기계와의 인터페이스의 실현 등의 이유로부터 다양한 분야에서 음성 및 화자인식에 대한 기술 실현의 요구가 많아지고 있다. 또한 음성인식 분야의 실용화를 위하여 잡음 환경에서의 음성인식률 향상 등에 많은 문제점들이 남아있다. 이러한 잡음에는 정상적인 배경잡음 외에 비정상적인 잡음이 일반적으로 존재한다. 이러한 비정상적인 잡음 중에는 피치 주파수 및 스펙트럼의 형태가 음성의 주파수 및 스펙트럼과 상당히 유사한 잡음들이 존재하여, 음성인식 장치의 오작동 및 음성인식률 저하의 주요 원인이 되고 있다[1-3].

최근에 음성인식의 연구 분야에 있어서 잡음대책의 중요성이 부각되고 있으며, 음성 및 잡음을 구별하는 연구들이 다수 시도되고 있다[4, 5]. 이러한 음성 및 잡음을 구별하는 수단으로서 스펙트럼의 차이를 이용하는 연구, 켑스트럼 계수의 분포를 이용하는 연구, 스펙트럼의 양자화를 이용하는 연구, 위너필터를 이용하는 연구, 신경회로망을 이용하여 판별하는 연구 들이 제안되고 있다[6-8]. 그러나 이러한 연구들은 주로 정상적인 잡음을 대상으로 하고 있으며, 또한 잡음 종류가 단순하므로 다양한 잡음 하에서 음성과 잡음을 구별하는 실험으로는 만족하지 못하고 있다. 또한 이러한 연구들은 연속한 음성에 잡음이 중첩된 경우를 대상으로 하고 있어서 음성단어 인식장치가 오동작하는 연속시간이 짧은 잡음을 대상으로 하고 있지 않다.

실제 환경에서 음성을 방해하는 배경잡음은 백색잡음과 같은 정상적인 잡음과 도로잡음과 같은 비정상적인 잡음의 합으로 표현하는 경우가 많다. 따라서 본 논문에서는 음성 중에서 정상적인 잡음과 비정상적인 잡음을 구별하기 위하여, 잡음이 섞인 음성 중에서 지속시간이 짧은 음성을 대상으로 하여 배경잡음을 구별하는 알고리즘에 대하여 기술한다.

또한 본 알고리즘은 잡음과 음성을 구별하는 중요한 수단으로서 개량된 음성의 특징파리미터를 사용하여 다층퍼셉트론 네트워크에 의하여 잡음의 종류를 추정하는 알고리즘에 대해서 기술한다. 본 실험에서는 여러 종류의 잡음을 이용하여 음성 중에서 잡음을 구별하는 실험을 실시한다.

Ⅱ. 제안한 알고리즘

음성인식에 유효한 음향 특징량으로서 음성의 스펙트럼 이외에 전력 및 동적 특징량이 있으며, 이러한 파라미터를 조합하는 것에 의하여 음성인식 성능이 향상되는 것이 보고되고 있다[7]. 본 논문에서는 개량된 선형예측계수(Linear Predictive Coefficient, LPC)[9]의 음향 특징량으로 사용하여 다층 퍼셉트론 네트워크에 부가는 방법을 제안한다.

그림 1은 본 논문에서 제안하는 잡음의 스펙트럼을 차감하는 위너필터에 의한 개량된 음성특징 파라미터를 추출하는 알고리즘을 나타낸다. 제안한 알고리즘은 먼저 여러 종류의 잡음으로 중첩된 음성신호를 각 프레임에 대해 해밍창을 통과시킨다. 각 프레임에 대해서 본 논문에서 제안하는 잡음의 스펙트럼을 차감하는 위너필터에 의하여 개량된 LPC 음성특징 파라미터를 추출한다. 그림 2는 본 논문에서 제안하는 개량된 음성특징 파라미터를 사용한 음성인식 알고리즘을 나타내며, 그림 1에서 추출된 개량된 음성특징 파라미터를 사용하여 본 논문에서 제안한 다층 퍼셉트론 신경회로망(Multi-Perceptron Neural Network, MLP)[10]에 의하여 음성인식을 실현하는 알고리즘을 나타낸다.

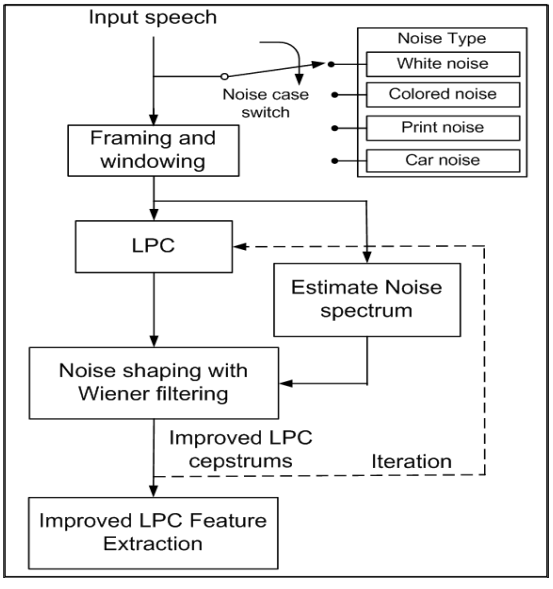

Fig. 1 Proposed improved speech feature extraction method by Wiener filtering

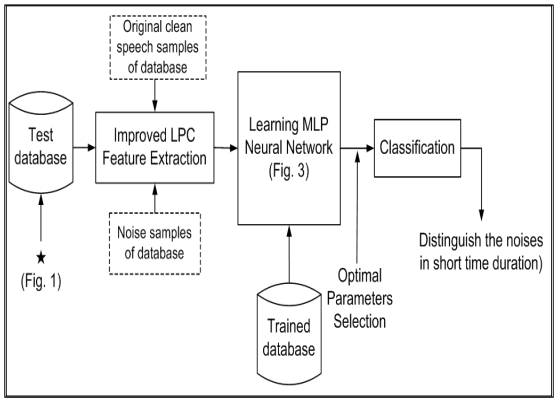

Fig. 2 Proposed speech classification using improved speech feature extraction

그림 3은 본 논문에서 제안하는 잡음의 종류를 분류하기 위한 3층으로 구성된 다층 퍼셉트론 네트워크이다. 역전파 학습 알고리즘(Back-propagation Learning Algorithm)[11]을 대상으로 하는 그림 3의 네트워크는 입력층, 출력층 및 은닉층(hidden-layer)로 불리우는 중간층으로 구성되어 있다. 네트워크의 입력으로는 그림 2에서 구해진 개량된 12차의 LPC 켑스트럼 계수이다. 역전파 학습 알고리즘의 특징은 교사 학습에 있어서 출력층으로부터 입력층으로 오차를 전파하는 특징으로부터 각 유닛에 대하여 최급하강법을 적용하는 것이 가능하게 되었다. 그리고 또 하나의 특징은 각 유닛에 비선형 함수를 도입함으로써 입력부터 출력에의 사상이 보다 일반성을 가져올 수 있게 되었다.

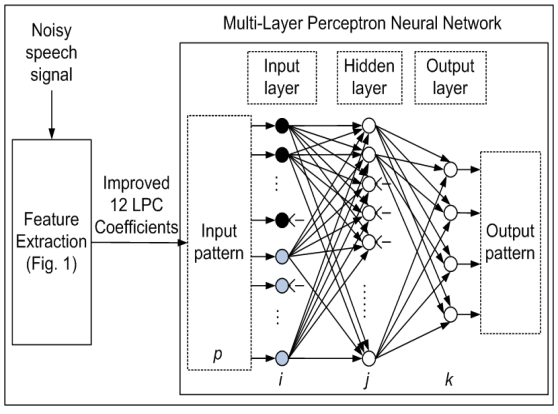

Fig. 3 Proposed multi-layer perceptron network

Ⅲ. 실험 및 학습 조건

본 장에서는 음성 및 잡음 데이터베이스의 실험조건 및 네트워크의 학습조건에 대하여 기술한다.

표 1은 본 실험에서 사용하는 입력 음성데이터의 분석 조건을 나타낸다. 본 실험에서 사용한 음성신호는 8kHz의 표본주파수의 환경에서 녹음된 영어숫자로 구성된 Aurora2 데이터베이스[12] 및 양자화 비트수가 12비트인 일본인 남성화자에 의한 단어의 총 2 종류의 음성 데이터베이스를 사용하였다. 본 실험에서는 음성데이터베이스 중에서 총 10개의 단어를 학습에 사용하였으며, 20개의 단어는 배경잡음 인식에 사용하였다. 여기에서 영어숫자와 일본어는 서로 연관성이 없으며, 일본어 단어는 일본 지명을 나타내는 단어이다.

본 실험에서 사용한 잡음은 컴퓨터에 의해서 작성된 가우스 백색잡음(white noise), 유색잡음(colored noise), 도트 프린터의 동작 시에 녹음한 프린터의 구동잡음(print noise), Aurora2 데이터베이스의 자동차잡음(car noise) 등의 정상잡음 및 비정상잡음을 사용하여 평가하였다. 본 논문에서 사용한 구동잡음 및 자동차잡음은 음성의 스펙트럼과 유사한 모양을 가지고 있다.

표 2에는 본 논문에서 제안하는 다층 퍼셉트론 신경회로망의 분석 조건을 나타낸다. 본 논문에서 제안하는 다층 퍼셉트론에 입력하는 음성특징량으로는 12차의 LPC 켑스트럼 계수를 사용하며, 네트워크에 입력되는 계수는 －1.0부터 +1.0 사이의 값으로 정규화된다.

Table. 1 Analysis conditions of input speech database

Sampling frequency	8 kHz
Method of analysis	12th LPC cepstrum
Analysis window	Hamming window

Table. 2 Learning conditions of multi-layer perceptron network

Number of input layer unit	12
Number of hidden layer unit	30
Number of output layer unit	5
Coefficient of training	0.2
Coefficient of inertia	0.5
Convergence determination error	0.001

Ⅳ. 음성인식 실험

음성신호의 잡음억압 및 분류를 위하여 잡음에 오염된 음성신호의 공간으로부터 잡음이 없는 음성신호의 공간에의 사상을 실현하기 위하여 다층 퍼셉트론을 적용하는 연구가 수행되고 있다. 본 실험에서는 이러한 다층 퍼셉트론을 3장에서 기술한 음성 데이터베이스와 잡음 데이터베이스를 사용하여, 역전파 학습 알고리즘으로 학습시켜 중간층 1층으로 구성된 3층 퍼셉트론이 정확하게 잡음분류를 할 수 있도록 학습을 확인한다.

본 논문에서 사용한 위너필터에 의한 개량 LPC 켑스트럼 계수의 유효성을 입증하기 위하여, 그림 4와 그림 5와 같이 본 실험에서 사용한 남성에 대한 음성신호(제7프레임)의 LPC 켑스트럼의 모양을 나타낸다. 그림 4는 백색잡음에 대하여 LPC 켑스트럼 계수의 한 프레임분의 출력을 나타낸다. 그림 4에서 첫 번째 그림은 원래의 LPC 켑스트럼 계수를 나타내며, 두 번째 그림은 위너필터에 의하여 개량된 LPC 켑스트럼 계수를 나타낸다. 그림 5의 자동차잡음에 대하여 첫 번째 그림은 원래의 LPC 켑스트럼 계수를, 두 번째 그림은 위너필터에 의하여 개량된 LPC 켑스트럼 계수의 한 프레임분의 출력을 각각 나타낸다. 그림 4와 그림 5의 결과로부터 알 수 있듯이, 위너필터에 의하여 개량된 LPC 켑스트럼 계수가 원래의 LPC 켑스트럼 계수보다 분명한 차이점을 가지는 것과 음성의 스펙트럼이 강조된 것을 확인할 수 있었다. 따라서 본 논문에서 제안하는 다층 퍼셉트론 네트워크에 입력되는 개량된 LPC 켑스트럼 계수를 이용하여 네트워크의 학습 및 잡음 분류 테스트에 있어서 분류율이 향상되는 것을 확인할 수 있었다. 그러나 향후의 연구에서는 멜 주파수 켑스트럼계수(Mel Frequency Cepstral Coefficient, MFCC)와 같은 특징추출방법을 사용한 성능비교가 필요하다고 판단된다.

본 실험에서는 3층 구조의 퍼셉트론형의 신경회로망에 선형예측계수 LPC[9]를 입력으로 하여 각 프레임에서 잡음 및 음성을 분류하는 것을 목적으로 하여 분류율을 높이는 실험에 대하여 기술한다.

본 실험에서는 네트워크의 학습데이터로서 음성, 백색잡음, 유색잡음, 프린터 구동잡음을 사용하여 학습을 실시하였다. 표 3은 제안한 알고리즘과 기존의 분류[4]에 의한 방법을 비교한 실험결과이며, 그림 6은 표 3의 분류율을 그래프로 표시한 결과이다. 표 3의 Ref[4]의 알고리즘은 LPC와 고속푸리에변환에 의한 전력스펙트럼의 총 13개의 계수를 사용하여 각 프레임에서 인식율을 구한다. 표 3과 그림 6의결과로부터 학습데이터 및

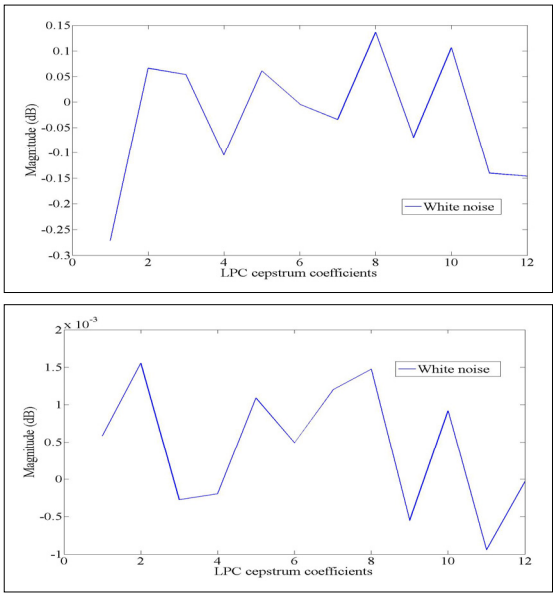

Fig. 4 The outputs for LPC cepstrum coefficients in the case of white noise

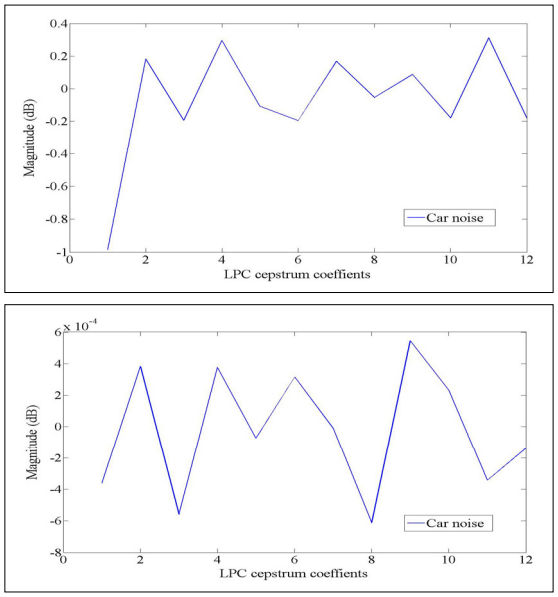

Fig. 5 The outputs for LPC cepstrum coefficients in the case of car noise

평가데이터가 다른 경우의 분류율은 평균 92% 이상인 것을 알 수 있으며, 본 논문에서 제안한 알고리즘의 분류율은 평균 95% 이상인 것을 확인할 수 있었다. 본 알고리즘에 의한 분류율은 기존의 분류법[4]와 비교하여 최대 4.1%(평균 3.04%)이 개선된 것을 알 수 있다. 따라서 8 kHz로 샘플링된 음성 데이터를 사용하여, 중간층 1층으로 구성된 3층 퍼셉트론을 사용하여 back-propagation 학습 알고리즘에 의하여 학습시킨 결과, 양호한 잡음 분류의 실험결과가 학습 이외의 데이터에 대해서도 실험적으로 확인되었다.

Table. 3 The comparison of noise classification rates for the conventional method

Signal and noise type	Classification rates (%)		
	Ref[4]	Prop.	Impr.
Speech signal	94.3 %	98.4 %	4.1 %
White noise	92.7 %	96.3 %	3.6 %
Colored noise	91.6 %	94.5 %	2.9 %
Print noise	90.5 %	93.0 %	2.5 %
Car noise	92.5 %	94.6 %	2.1 %
Average	92.32 %	95.36 %	3.04 %

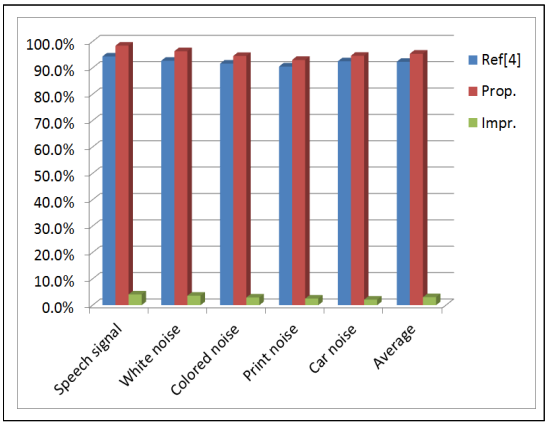

Fig. 6 The figure of noise classification rates

V. 결 론

음성인식처리의 분야에서 배경잡음은 음성인식율의 저하를 초래하고 있으므로 보다 고도한 잡음처리기술이 필요하게 된다. 따라서 본 논문에서는 잡음환경 중의 계속시간이 짧은 정상적인 배경잡음 및 비정상적인 배경잡음을 인간의 음성과 구별하는 알고리즘에 대하여 기술하였다. 본 알고리즘은 잡음과 음성을 구별하는 중요한 수단으로서, 음성의 특징파리미터를 선택하여 다층 퍼셉트론 네트워크에 의하여 잡음의 종류를 구별하는 알고리즘이다.

본 실험에서는 잡음과 음성의 구별이 가능하도록 실험적으로 분명하게 하였다. 따라서 중간층 1층으로 구성된 3층 퍼셉트론을 사용한 역전파 학습 알고리즘에 의하여 학습시킨 결과, 양호한 음성인식 결과가 학습 이외의 데이터에 대해서도 실험적으로 확인할 수 있었다. 특히 Wiener 필터에 의하여 개량된 LPC 켑스트럼 계수를 사용하여 다층 퍼셉트론 네트워크에 사용한 경우에 개량 전의 기존의 인식방법보다 최대 4.1%의 분류율이 구해졌다. 특히 본 논문에서 제안한 알고리즘은 음성의 스펙트럼에 유사한 비정상적인 배경잡음에도 우수하다는 것을 실험적으로 확인할 수 있었다.

REFERENCES

[1] C. Hanilçi, T. Kinnunen, R. Saeidi, J. Pohjalainen, P. Alku, F. Ertaş, J. Sandberg, and M. Hansson-Sandsten, "Comparing spectrum estimators in speaker verification under additive noise degradation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4769-4772, March 2012.

[2] R. Saeidi, J. Pohjalainen, T. Kinnunen, and Paavo Alku, "Temporally Weighted Linear Prediction Features for Tackling Additive Noise in Speaker Verification," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 599–602, June 2010.

[3] R. Su, X. Liu, and L. Wang, "Automatic Complexity Control of Generalized Variable Parameter HMMs for Noise Robust Speech Recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 102-114, Jan. 2015.

[4] J. S. Choi, "Speech and Noise Recognition System by Neural Network," *The Korea Institute of Electronic Communication Sciences*, vol. 5, no. 4, pp. 357-362, Aug. 2010.

[5] N. Moritz, Jörn Anemüller, and B. Kollmeier, "An Auditory Inspired Amplitude Modulation Filter Bank for Robust

Feature Extraction in Automatic Speech Recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1926-1937, Nov. 2015.

[6] J. S. Choi, "Noise Reduction Algorithm in Speech by Wiener Filter," *The Korea Institute of Electronic Communication Sciences*, vol. 8, no. 8, pp. 1293-1298, Aug. 2013.

[7] S. Furui, "Speaker-Independent Isolasted Word Recognition Using Dynamic Features of Speech Spectrum," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 52-59, Feb. 1986.

[8] J. S. Choi, "A Wiener Filter Algorithm of Noise Subtraction Based on Threshold Detection," *Korean Institute of Information Technology*, vol. 13, no. 7, pp. 51-56, July 2015.

[9] X. Zhang, Y. Guo, Xuemei Hou, "A speech Recognition Method of Isolated Words Based on Modified LPC Cepstrum," *IEEE International Conference on Granular Computing*, pp. 481-484, Nov. 2007.

[10] S. K. Pal, S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Transaction on Neural Networks*, vol. 3, no. 5, pp. 683-697, Sep. 1992.

[11] D. Rumelhart, G. Hinton and R. Williams, "Learning representations by back-propagation errors," *Nature,* vol. 323, pp. 533-536, Oct. 1986.

[12] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions," i*n Proc. ISCA ITRW ASR2000 on Automatic Speech Recognition: Challenges for the Next Millennium*, Paris, France, Oct. 2000.

최재승(Jae-Seung Choi)

1989년 조선대학교 전자공학과 공학사
1995년 일본 오사카시립대학 전자정보공학부 공학석사
1999년 일본 오사카시립대학 전자정보공학부 공학박사
2000년~2001년 일본 마쯔시타 전기산업주식회사 (현, 파나소닉 주식회사) AVC사 연구원
2002년~2007 경북대학교 디지털기술연구소 책임연구원
2007년~현재 신라대학교 전자공학과 교수
※관심분야 : 음성신호처리, 신경회로망, 적응필터와 잡음제거 등