

장애 음성 판별을 위한 의료/전자 융복합 소프트웨어 개발

문지혜, 이지연
중원대학교 의료공학과

Development of medical/electrical convergence software for classification between normal and pathological voices

Ji-Hye Moon, JiYeoun Lee

Department of Biomedical Engineering, Jungwon University

요 약 장애음성을 판별할 수 있는 소프트웨어가 개발 될 경우, 원격의료와 언어치료 등 여러 융복합 분야에서의 활용도가 매우 높다. 본 논문은 성대 진동에 대한 변화율을 나타내는 의료정보인 음향학적 파라미터와 신호처리 기반 고차 통계량에 기반을 둔 파라미터를 융합하여, CART(Classification And Regression Trees) 분석을 통해서 정상/장애음성 판별 프로그램을 구현하였다. 사용된 음향학적 파라미터는 Jitter(%)와 shimmer(%)이다. 그리고 본 연구에서 제안된 고차통계량 기반 파라미터는 왜도(Skewness)와 첨도(Kurtosis)의 평균과 분산이다. Kay Elemetrics의 데이터베이스에서 무작위로 발췌된 정상음성 53명, 장애 음성 173명의 /아/ 발화를 이용하여 결정트리(Decision tree) 기반 장애음성 판별을 위해 평균적으로 83.15%의 성능을 보이는 알고리즘을 구현하였다. 그 결과를 바탕으로 추후 상용화를 고려하여 사용자 친화적인 프레임 워크에 의해 콘텐츠를 생성하는 융복합형 기능이 포함된 장애음성 판별 프로그램을 개발하였다.

주제어 : 고차통계량, 음향학적 분석, 융복합 음성분석 소프트웨어, 의료전자

Abstract If the software is developed to analyze the speech disorder, the application of various converged areas will be very high. This paper implements the user-friendly program based on CART(Classification and regression trees) analysis to distinguish between normal and pathological voices utilizing combination of the acoustical and HOS(Higher-order statistics) parameters. It means convergence between medical information and signal processing. Then the acoustical parameters are Jitter(%) and Shimmer(%). The proposed HOS parameters are means and variances of skewness(MOS and VOS) and kurtosis(MOK and VOK). Database consist of 53 normal and 173 pathological voices distributed by Kay Elemetrics. When the acoustical and proposed parameters together are used to generate the decision tree, the average accuracy is 83.11%. Finally, we developed a program with more user-friendly interface and frameworks.

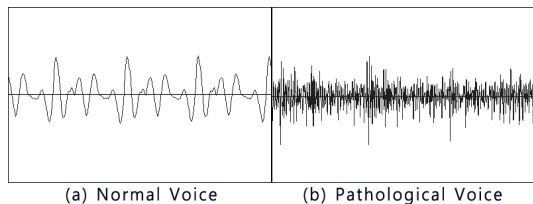
Key Words : Higher-order Statistics, Acoustical analysis, Convergence voice analysis software, Biomedical electricity

* 이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 No. 2014-00540001
Received 30 October 2015, Revised 30 November 2015
Accepted 20 December 2015
Corresponding Author : JiYeoun Lee(Jungwon university)
Email: jylee@jwu.ac.kr

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

인구의 고령화 및 의료비 부담 증가에 따라 질병의 예방 및 일상 관리의 중요성이 대두되어지고 있다[1]. 의료 산업의 연구 방향성이 질병의 치료에서 질병의 진단, 사후 관리, 질병의 예방으로 전환[1]되었으며, 이 과정에서 음성 신호는 별도의 숙련 과정 없이 편리하고 정확히 현재 상태를 기계에 전달할 수 있다는 점에서 큰 장점을 가지고 있다. 따라서 차세대 융복합 인터페이스 기술로써 정상/장애 음성 판별이 재활을 넘어 간편한 음성건강 진단의 지표로써 사용되고 있다[1,2,3,4,5,6]. 국내 음성치료 전문가들은 장애음성 분석을 위해 MDVP (Multidimensional Voice Program), TF32, Dr. Speech 등을 사용하는 것으로 보고되었다[1,3,7]. [Fig. 1]의 장애와 정상음성 파형에서 볼 수 있듯이, 정상음성 파형은 주기적으로 안정적이기 때문에 정확한 음성 특징 정보의 추산이 가능하다. [Fig. 1] (b)의 장애음성은 성대의 떨림 및 폐쇄 상태로 인해 발성에 문제가 발생하기 때문에 기식화 성분이 포함되어 불규칙한 구간이 발생해 부정확한 음성 특징이 추산될 수 있다[3,6,8,9]. 그러나 시중에 출시된 음성 분석 프로그램은 장애음성 특징을 고려하지 않고 정상음성 데이터베이스만을 가지고 개발되었기 때문에 위의 프로그램 중에서 어떤 것이 더 정확한지 또는 신빙성이 있는지 판단할 수 없다.



[Fig. 1] Waveform of normal and pathological voice signal

현재까지 개발된 수많은 음향학적 파라미터 중에서 장애 음성 판별에 중요한 역할을 하는 특징 파라미터로는 Pitch, Jitter(%), Shimmer(%), Harmonics-to-noise ratio (HNR) 등이 있으며, 이 특징 파라미터들은 성대의 기본 진동을 나타내는 기본 주파수(Fundamental frequency)를 기반으로 한다[7,8]. 그러나 위의 파라미터들은 음성 신호에서 잡음이 혼재되어 있으면 음성의 정확한 분석에

효과적이지 않다. 따라서 최근에 3차 이상의 통계적 특성을 이용하여 신호를 처리하는 고차통계량(Higher-order statistics, HOS)을 이용한 파라미터가 개발되었다[3,7,9,10]. 대부분의 장애음성 판별에 관한 연구는 GMM(Gaussian mixture model), SVM(Support vector machine), 그리고 NN(Neural network) 등과 같은 분류 알고리즘에 초점을 맞추어 정상/장애음성 판별 성능을 개선하는 방법에 대해 주로 연구해왔다[11,12,13,14].

환자를 위한 최적의 치료는 적절한 진단[1]에서 출발하며, 적절한 진단은 정확한 프로그램 개발에서부터 출발한다. 어떠한 전문가라 하더라도 항상 객관적인 진단만을 내릴 수 없으며, 환자의 동향을 일일이 확인할 수 없기 때문에 정확한 음성건강 평가 프로그램 개발은 필수적이다. 따라서 의료정보인 음향학적 파라미터와 신호 처리 기반 고차 통계량을 융합하여 사용자 친화적인 프레임 워크에 의해 콘텐츠를 생성하는 융복합 기능이 포함된 장애/정상음성의 분류를 위한 프로그램을 개발하고자한다.

2. 음향학적 분석(Acoustical analysis)

2.1 주파수 변화율(Jitter, %)

Pitch는 수식 (1)의 자기상관함수 (Auto-correlation function)를 이용하여 추출할 수 있다[11]. T 는 주기이고, $x(t)$ 는 특정 시간 t 에서의 신호 값, $x(t+\tau)$ 는 t 에서 τ 만큼 떨어진 시간의 신호 값이다.

$$\overline{x(t) \cdot x(t+\tau)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t+\tau) \quad (1)$$

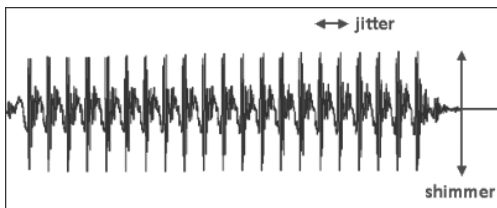
Jitter(%)는 단위시간 동안 Pitch의 변화율이다[11]. 수식(2)과 같이 시간 축에서 주기의 불규칙한 변동을 나타내며 [Fig. 2]에서도 볼 수 있다. 수식 (2)에서 $T(i)$ 는 수식 (1)에서 구한 피치 값, N 은 피치 주기의 전체 개수를 나타낸다.

$$Jitter(\%) = 100 \times \frac{\frac{1}{N-1} \sum_{i=0}^N |T(i) - T(i+1)|}{\frac{1}{N} \sum_{i=1}^N |T(i)|} \quad (2)$$

2.2 진폭 변화율(Shimmer, %)

Shimmer(%)는 수식 (3)과 같이 음성 구간별 진폭 값의 변화율이라 정의되며 진폭 변화율을 %로 표시한다. 또한 [Fig. 2]에서와 같이 특정 주기 당 진폭의 변화에 대한 불규칙성을 나타낸다[11]. 수식 (3)에서 $A(i)$ 는 진폭 (Amplitude), N 은 진폭의 전체 개수이다.

$$Shimmer(\%) = 100 \times \frac{\frac{1}{N-1} \sum_{i=0}^N |A(i) - A(i+1)|}{\frac{1}{N} \sum_{i=1}^N |A(i)|} \quad (3)$$



[Fig. 2] Representation of Jitter(%) and Shimmer(%)

3. 고차 통계량 분석 (Higher-order statistics)

고차통계 방법은 3차 이상의 통계적 특성을 이용하여 신호를 처리하는 방법이다. Gaussian 잡음 환경에서 왜곡이 발생하여 성능이 저하되는 자기상관함수와 같은 2차 통계와 달리 Gaussian 특성을 가지는 잡음과 대칭적인 확률 분포를 가지는 잡음에 강인한 특성을 가진다. Non-Gaussian 분포를 가지는 장애음성 신호의 불안정적이고 불연속적인 요소를 추산하는 효과적이라고 알려져 있다[3,5,7,9].

본 논문에서는 정상과 장애음성을 판별하는 파라미터로써 3차와 4차 통계변수인 왜도(Skewness)와 첨도(Kurtosis) 평균(Mean)과 분산(Variance)을 사용한다. 왜도와 첨도의 정의는 다음 수식 (4)와 (5)와 같이 표현된다[7]. N 는 전체 프레임 갯수, x_i 는 신호 값, \bar{x} 는 평균을 의미한다.

$$Skewness = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^{\frac{3}{2}}} \quad (4)$$

$$Kurtosis = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^2} \quad (5)$$

Skewness를 S_i , Kurtosis를 K_i 라 하였을 때, Mean of skewness(MOS), Variance of skewness(VOS), Mean of kurtosis (MOK), Variance of kurtosis (VOK) 파라미터는 수식 (6), (7), (8), 그리고 (9)와 같이 제안되어진다. \bar{S} 는 분석 구간의 왜도 평균 값, \bar{K} 는 분석 구간의 첨도 평균 값을 의미한다.

$$Mean\ of\ Skewness\ (MOS) = \frac{1}{N} \sum_{i=1}^N S_i \quad (6)$$

$$Variance\ of\ Skewness\ (VOS) = \frac{1}{N-1} \sum_{i=1}^N (S_i - \bar{S})^2 \quad (7)$$

$$Mean\ of\ Kurtosis\ (MOK) = \frac{1}{N} \sum_{i=1}^N K_i \quad (8)$$

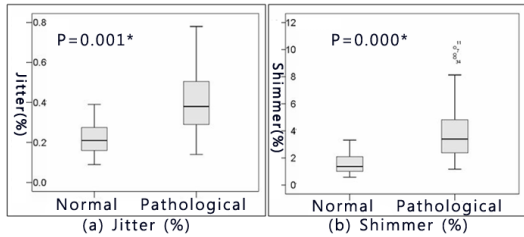
$$Variance\ of\ Kurtosis\ (VOK) = \frac{1}{N-1} \sum_{i=1}^N (K_i - \bar{K})^2 \quad (9)$$

4. 실험 결과

본 논문에서는 데이터베이스로써 Disordered voice database version 1.03 Model 4337 (Kay Elemetrics, Lincoln Park, NJ)에서 발췌된 정상 음성 53명, 장애 음성 173명의 /아/ 발화를 사용하였다[12]. 훈련 데이터용으로 정상음성 37명, 장애음성 121명을 사용하였고, 나머지는 장애음성 판별 소프트웨어 개발 후 테스트용으로 사용하였다. 그리고 공인된 음향학적 분석 프로그램 TF32를 사용하여 Jitter(%)와 Shimmer(%)를 추출하였다.

[Fig. 3]는 정상/장애 음성에서 추출된 음향학적 파라미터들의 분포를 나타낸다. 공통적으로 각 분포는 최대값과 최소값으로 표시되며, 박스 구성 (Box plot)은 중간 값(Median), 1사분위수(First quartile)와 3사분위수(Third quartile)로 표현되어진다. [Fig.3]의 (a)와 (b)는 각각 정상음성과 장애음성에서의 피치 변화율과 진폭의 변화율을 나타낸다. 장애음성의 불규칙한 파형 특징에서 예상되어지듯이, 장애음성에서 변화율이 큼을 알 수 있다. SPSS의 Mann-Whitney 분석방법을 이용하여 유의

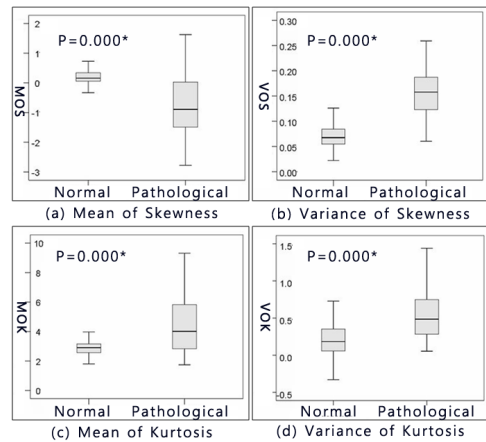
수준 0.01에서 각 파라미터 분포도에서 정상음성과 장애 음성 간의 유의미한 차이가 나타남이 확인하였다.



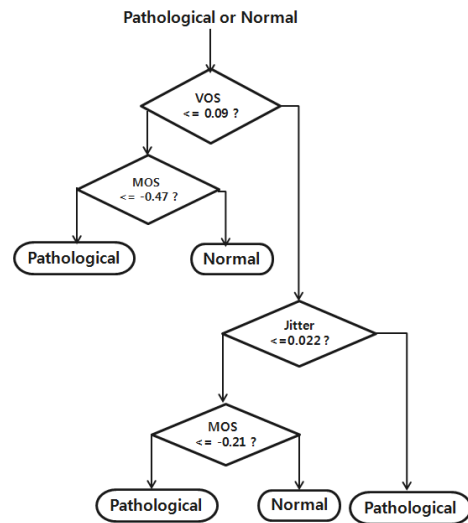
[Fig. 3] Distribution between normal and pathological signals (a) Jitter(%) (b) Shimmer(%)

[Fig. 4]는 제안된 정상/장애 음성에서 추출된 고차통계량 기반 파라미터들의 분포를 나타낸다. [Fig. 4]의 (a), (b), (c), (d)는 각각 MOS, VOS, MOK, 그리고 VOK 파라미터들의 분포도를 나타내며, 정상음성과 장애음성에서의 고차통계량의 변화율을 보여준다. 전체적으로 장애 음성에서의 변화율이 크게 나타나며, 침도 기반 파라미터보다 왜도 기반 파라미터에서 정상음성과 장애음성 사이의 차이가 더 명확하게 나타난다. 모든 파라미터에의 분포도에서 정상음성과 장애음성 간의 유의미한 차이가 나타났다.

본 연구에서는 CART(Classification and regression trees)를 사용하여 정상음성과 장애음성을 분류하였다. CART는 각 독립 변수를 이분화 하는 과정을 반복하여 결정 트리(Decision tree)를 형성하는 방법으로, 서로 다른 파라미터들의 분류 및 예측을 위해 효과적으로 사용된다[13,14,15]. Jitter(%)와 shimmer(%)와 같은 음향학적 파라미터와 고차통계량 기반 파라미터는 범위와 단위가 다르기 때문에 같이 복합적으로 동시에 사용하여 성능을 체크하기가 쉽지 않다. 이런 단점을 해결하기 위한 방법 중의 하나로써 CART 분석을 통해 형성된 Tree를 통해서 성능을 측정할 수 있다. [Fig. 5]은 6개의 파라미터를 사용하여 CART 분석을 통해 형성된 Tree를 나타낸다. 여러 파라미터 중에서 정상음성과 장애 음성 판별에 효과적인 파라미터는 Jitter(%), MOS, 그리고 VOS 파라미터임을 알 수 있다.



[Fig. 4] Distribution between normal and pathological signals (a) MOS, (b) VOS, (c) MOK, (d) VOK



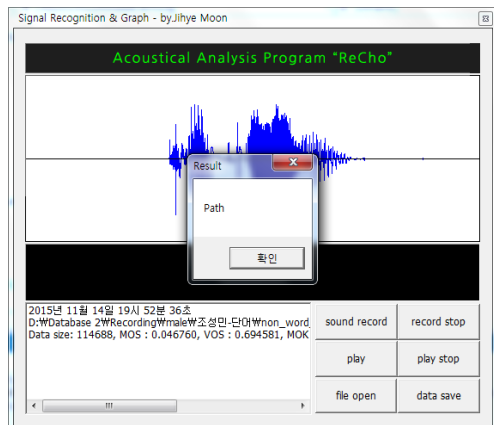
[Fig. 5] Decision tree formed from decision tree

<Table 1>은 테스트용 데이터베이스를 이용하여 장애 음성 판별 성능을 나타낸다. 정상음성의 판별율은 37개의 데이터 중 27개의 성공 비율인 약 73%였고, 장애음성 판별율은 121개의 데이터 중 113개의 성공 비율인 약 93%였다. 장애음성 판별을 위해 평균적으로 83.15%의 분류 성능을 보였다. 이를 토대로 [Fig. 6]의 정상/장애 음성 판별 프로그램인 “ReCho”를 구현하였다. eCho는 소리의 울림 혹은 반복을 의미하는 echo에서 유래되었으며, 음성 신호를 되새긴다는 분석한다는 의미에서 접두

어 Re를 붙임으로 "ReCho"라 명명하였다. "ReCho"는 음성 파일을 열고 저장하는 기능(File Open/Data Save)과 재생기능(Play), 실시간으로 음성을 녹음하여 정상 음성 과 장애음성을 판별하는 기능(Sound Record/Stop)을 가진다. 불필요한 기능을 배제함으로 사용자 만족도에 부합하고 사용자 친화적인 정상/장애 음성 판별에 최적화된 프로그램이다.

<Table 1> Classification performance between normal and pathological voices

Result(%)	Normal voices	Pathological voices
Normal voices	73% (27/37)	27% (10/37)
Pathological voices	6.7% (8/121)	93.3% (113/121)



[Fig. 6] Medical/electrical convergence program for classification between normal and pathological voice

5. 결론

본 논문에서는 의료정보인 음향학적 파라미터와 신호 처리 기반 고차 통계량을 이용한 파라미터를 융합하여 제시된 6가지 파라미터를 이용하여 의료/전자 융복합 장애음성 분류 프로그램을 개발하였다. 그 파라미터들은 jitter(%), shimmer(%), MOS, VOS, MOK, 그리고 VOK이다. CART 분석을 통해 형성된 결정트리에서는 jitter(%), VOS, 그리고 MOS 파라미터가 효과적으로 장애음성 분류에 사용될 수 있음을 보였으며, 결정트리

(Decision tree) 기반 장애음성 판별을 위해 평균적으로 83.15%의 성능을 보이는 알고리즘을 구현하였다. 따라서 위의 알고리즘을 기반으로 정상/장애 음성을 빠르고 정확히 분류하기 위한 "ReCho"이 개발되었으며, 이것은 CART 기반 결정트리를 이용함으로써 실시간으로 빠르게 데이터를 분석할 수 있다는 장점을 가진다. 현재, 사용자 친화적인 프레임 워크에 의해 콘텐츠를 생성하는 융복합형 기능이 포함된 장애음성 판별 프로그램을 개발하였으며, 추후 "ReCho"를 임상적으로 적용하기 위해서는 다양한 환경에서의 실험이 필요하며, 이를 토대로 분류 성능의 정확도를 높인 후 상용화를 할 예정이다.

ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No. 2014-00540001).

REFERENCES

- [1] Jinsu Lee, KHIDI Brief vol.140, pp.1-2, Korea Health Industry Development Institute, 2014.
- [2] Hwa-Young Pyo; Hyun Sub Sim, A Study for the Development of Korean Voice Assessment Model for the Patients with Voice Disorders: A Qualitative Study, The Korean Association of Speech Sciences, vol. 14, no.2, pp. 7-22 (16 pages), 2007.
- [3] Ji-Yeoun Lee; Minsoo Hahn, Automatic Assessment of Pathological Voice Quality Using Higher-Order Statistics in the LPC Residual Domain, EURASIP Journal on Advances in Signal Processing, Volume 2009, Article ID 748207, 8 pages, 2009.
- [4] Soon-Bok Kwon; Soon-Woo Kwon, The Effect of Self Voice Feedback Training Using Praat on the Voice Improvement of Patient with Vocal Nodules, Journal of Special Education & Rehabilitation Science, Vol. 46, No. 1, pp. 191~215, 2007.
- [5] J.B. Alonso et al., "Automatic Detection of

Pathologies in the Voice by HOS Based Parameters,” EURASIP Journal on Applied Signal Processing, vol. 4, pp. 275-284, 2001.

[6] Ki-Chang Nam; Seung-Hoon Lee; Jai-Nam Choi; Hong-Shik Choi; Do-Hyun Nam; Deok-Won Kim, Comparison of vowel pitch results among several commercial voice analysis programs, ICS’05, pp.54-56, 2005.

[7] Ji-Yeoun Lee, Performance Improvement of Automatic Pathological Voice Quality Assessment Based on Higher-Order Statistics, ICU-School of Engineering [Thesis(doctoral)], pp.109, 2008.

[8] Bong-Hyun Kim; Dong-Uk Cho, Pronunciation Influence Analysis of Carbonate Drink and Eucalyptus Fragrance by Applying Speech Signal Processing Techniques, The Journal of Korean Institute of Communications and Information Sciences, Volume 37, Issue 5C, pp.420-428, 2012.

[9] Taeyoung Shin; Giseong Kim; Yeonguk Kwon; Hyeongsun Kim, Speaker Identification based on Higher-Order Statistics in Noisy Environment, The journal of the acoustical society of Korea, v.16, no.6, pp. 25-35, 1997.

[10] Tae Young Shin; Jae Ho Kim; Kyung Sik Son; Hyung Soon Kim, Pitch Determination and Voiced/Unvoiced Decision of Noisy Speech Based on the Higher-Order Statistics, SCAS, Vol. 12, no. 1, 1995.

[11] JiYeoun Lee; Seong Hee Cho., Perturbation analysis using a moving window for disordered voices, International Journal of Engineering, Science and Innovative Technology, Vol. 3, No. 1, pp. 1~10, 2012.

[12] Kay Elemetrics Corp. Multi-dimensional voice program: software instruction manual. Pine Brook: NJ: Kay Elemetrics Corp, 1993.

[13] J.I. Godino-Llorente; N. Sáenz-Lechón; V. Osma-Ruiz; S. Aguilera-Navarro; P. Gómez-Vilda, An integrated tool for the diagnosis of voice disorders, Medical Engineering & Physics, Vol. 28, No. 3, pp. 276~289, 2006.

[14] Xiang Wang; Jianping Zhang; Yonghong Ya,

Discrimination Between Pathological and Normal Voices Using GMM-SVM Approach, Journal of Voice, Vol. 25, No. 1, pp. 38~43, 2011.

[15] R. Das, A comparison of multiple classification methods for diagnosis of Parkinson disease, Expert Systems with Applications, Vol. 37, No.2, pp.1568~1572, 2010.

문 지 혜(Moon, Ji Hye)



- 2013년 2월 ~ 현재 : 중원대학교 의료공학과 재학
- 관심분야 : 생체신호처리, 신경공학, 재료공학
- E-Mail : bmel3moonjihye@gmail.com

이 지 연(Lee, Ji Yeoun)



- 2003년 2월 : KAIST, 전자공학과 (공학석사)
- 2008년 8월 : KAIST, 전자공학과 (공학박사)
- 2008년 9월 ~ 2011년 2월 : UCLA, University of Wisconsin-Madison, 연구원
- 2011년 3월 ~ 현재 : 중원대학교 의료공학과 조교수
- 관심분야 : 생체신호처리, 의료전자, 의공학기술
- E-Mail : jylee@jwu.ac.kr