

음원신호 추출을 위한 주파수영역 응용모델에 기초한 독립성분분석

최재승*

Independent Component Analysis Based on Frequency Domain Approach Model for Speech Source Signal Extraction

Jae-Seung Choi*

요 약

본 논문은 여러 음원신호가 혼합된 환경에서 목적으로 하는 음원신호만을 분리하기 위하여 마이크로폰을 사용한 블라인드 음원분리 알고리즘을 제안한다. 제안하는 알고리즘은 독립성분분석 방법을 기반으로 한 주파수영역 표현모델이다. 따라서 2 음원에 대한 주파수영역 독립성분분석의 실제 환경에서의 유효성 검증을 목적으로, 음원의 종류를 변경하여 주파수영역 독립성분분석을 실행하여 음원분리를 실시하여 그 향상효과를 검증한다. 파형에 의한 실험결과로부터 원래의 파형과 비교하여 2채널의 음원신호를 깨끗하게 분리할 수 있음을 명확히 하였다. 또한 목표 신호 대 간섭 에너지비율을 사용하여 비교한 실험 결과로부터 본 논문에서 제안한 알고리즘의 음원분리 성능이 기존의 알고리즘에 비하여 성능이 향상되었다는 것을 알 수 있었다.

ABSTRACT

This paper proposes a blind speech source separation algorithm using a microphone to separate only the target speech source signal in an environment in which various speech source signals are mixed. The proposed algorithm is a model of frequency domain representation based on independent component analysis method. Accordingly, for the purpose of verifying the validity of independent component analysis in the frequency domain for two speech sources, the proposed algorithm is executed by changing the type of speech sources to perform speech sources separation to verify the improvement effect. It was clarified from the experimental results by the waveform of this experiment that the two-channel speech source signals can be clearly separated compared to the original waveform. In addition, in this experiments, the proposed algorithm improves the speech source separation performance compared to the existing algorithms, from the experimental results using the target signal to interference energy ratio.

키워드

Blind Speech Source Separation, Independent Component Analysis, Frequency Domain Representation, Speech Source Separation
블라인드 음원 분리, 독립 성분 분석, 주파수영역 표현, 음원 분리

* 교신저자: 신라대학교 스마트전기전자공학부

• 접수일 : 2020. 07. 13
• 수정완료일 : 2020. 08. 29
• 게재확정일 : 2020. 10. 15

• Received : Jul. 13, 2020, Revised : Aug. 29, 2020, Accepted : Oct. 15, 2020

• Corresponding Author : Jae-Seung Choi
Division of Smart Electrical and Electronic Engineering, Silla University
Email : jschoi@silla.ac.kr

1. 서 론

현재의 음성처리시스템에 의한 실제 환경 중에서 여러 음성신호로부터 관측된 혼합음성신호에 대하여 각각의 음원신호에 관한 정보를 추출하는 것은 휴먼 로봇과 인간과의 음성통신신호처리 분야에 중요한 기술이라고 할 수 있다. 따라서 다양한 음원신호가 혼재된 실제 환경에 있어서 목적으로 하는 음원신호만을 분리하기 위해서 마이크로폰을 사용한 블라인드 음원 분리기술의 연구를 진행할 필요가 있다[1-5].

최근, 독립성분분석(Independent Component Analysis, ICA)에 기초한 블라인드 신호분리(Blind Source Separation, BSS)가 특히 주목받고 있다[6]. ICA는 원래 신호성분에 통계적인 독립성을 가정하여 혼합 입력된 음성신호성분으로부터 원래의 신호성분만을 분리하는 기법으로서 이러한 응용기술의 개발이 여러 분야에서 시험되고 있다. ICA를 사용하는 BSS는 음원 및 마이크로폰과 관련된 사전정보가 필요 없으며 관측신호만으로도 음원신호를 분리하는 것이 가능하다. 그러나 ICA는 분리신호의 순서에 임의성이 있기 때문에 주파수 빈마다 추정되는 분리신호의 순서를 적절히 변경해야 하는 후처리 문제가 새롭게 필요하게 되었다. 이 후처리는 퍼뮤테이션 문제로 불리며 지금까지 여러 해결법이 제안되어 있다[7-8]. 이 후에 퍼뮤테이션문제의 해결과 주파수마다 분리행렬의 추정을 동시에 달성하는 수법으로써 ICA를 기초로 한 독립벡터분석(Independent Vector Analysis, IVA)[5, 8-10] 및 Fast ICA[11] 등의 여러 수법이 제안되고 있다. IVA는 ICA를 다변량 확률분포의 모델로 변화시킨 이론으로서 각 음원에 대한 주파수성분을 하나로 정리한 주파수벡터의 생성모델로서 가정한다.

ICA에는 시간영역에서 분리를 실시하는 시간영역 ICA(Time Domain ICA, TDICA)[12]와 주파수영역에서 분리를 실시하는 주파수영역 ICA(Frequency Domain ICA, FDICA)[12-14]의 두 개의 응용 방법이 있다. 각 음원 간의 통계적인 독립성을 이용하는 다채널 음원분리 기법으로서 주파수범마다 선형분리행렬을 추정하는 FDICA에 기초한 기법이 검토되고 있다. FDICA는 주파수범마다 분리신호의 순서가 정해지지 않는 문제를 포함하고 있으며 여러 해결방법이 검토되고 있다. 일반적으로 TDICA는 분리 필터의 차수가

크게 되면 계산부하가 증가하게 되어 알고리즘의 수속성도 감소하게 된다. 그러나 이 독립성분분석을 시간영역에서 사용하게 되면 컨볼루션 혼합모델의 분리 필터를 추정할 필요가 있기 때문에 상당히 어려운 문제가 될 수 있다. 반면에 FDICA는 수속성도 양호하며 잔향이 적은 환경 하에서는 양호하게 동작한다. 따라서 수속성이 양호하고 학습면에서도 비교적 간소한 주파수영역에서의 독립성분분석이 제안되고 있다.

BSS는 고품질인 음성강조에의 응용을 목적으로 하여 광범위하게 연구되고 있으며[15], 이 중에서도 FDICA는 가장 유용한 수법 중의 하나라고 할 수 있다. 따라서 제안한 방법인 주파수영역 ICA를 이용하여 신호를 추출하기 위한 주파수영역 표현을 사용한 ICA 모델의 알고리즘을 제안한다. 제안한 알고리즘에서는 2음원에 대한 주파수영역 표현의 ICA의 환경에서의 유효성 검증을 목적으로, 음원의 종류를 변경하여 주파수영역 표현의 ICA를 실행하여 음원분리를 실시하여 그 향상효과를 검증한다.

II. ICA에 의한 혼합 모델과 분리 모델

BSS의 모델 중의 하나인 ICA는 화자음성 및 음악 등의 비정상신호도 분리 가능하다는 종래 기술에는 없는 특징을 가지고 있으며, 신호원의 추정뿐만 아니라 관측 데이터의 배후에 숨어있는 구조 및 특징의 추출에도 이용되고 있다[6, 11]. ICA를 사용하여 확률적으로 독립한 2개의 음원신호를 분리할 경우에 음원신호 $s(t)$ 는 식 (1)과 같이 표현할 수 있으며, 여기에서 tr 은 전치행렬을 나타낸다.

$$s(t) = [s_1(t), s_2(t)]^{tr} \quad (1)$$

수신된 혼합신호 $x(t)$ 는 식 (2)와 같이 구성할 수 있으며, A 는 원래의 음원신호를 혼합된 신호로 바꾸기 위하여 알려지지 않은 정수인 a_{jk} ($a_{11} \sim a_{22}$)을 구성요소로 하는 2×2 인 경우의 혼합행렬이다.

$$x(t) = A \cdot s(t) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (2)$$

이 때 추정된 행렬 $v(t)$ 을 식 (3)과 같이 기술할 수 있으며, W 는 2×2 에 해당하는 분리행렬이다.

$$v(t) = W \cdot x(t) \quad (3)$$

위에 기술한 내용에 기초하여 혼합신호 $x(t)$ 만을 음원신호로 이용하여 W 를 순서대로 변화하면서 추정된 신호 $v(t)$ 를 구하는 것이 ICA 알고리즘이다.

그림 1은 기존의 BSS 혹은 ICA에 의한 음원분리 모델을 나타내며, 혼합 음원신호를 분리하고 음원을 추정한다. 그림 1에서는 두 개의 마이크로폰으로 수신되는 음원을 이용하여 각 음원은 서로 결합되어 각 마이크로폰에서 혼합된 신호를 생성한다.

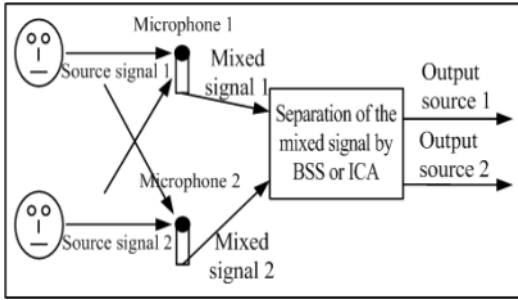


그림 1. 기존의 BSS에 의한 음원분리
Fig. 1 Source separation by conventional BSS

III. 주파수영역 표현 ICA에 의한 음원분리

제안한 알고리즘에서는 2채널 입력음성신호를 대상으로 한 신호의 음원분리를 하며 마이크로폰으로부터의 관측신호를 식 (4)와 같이 컨볼루션으로 정의할 수 있다.

$$x_j(t) = \sum_{k=0}^{M-1} a_{jk}(t) * s_k(t), \quad t = 0, 1, 2, \dots \quad (4)$$

여기에서 $x_j(t)$ 는 수신된 혼합신호이며 $s_k(t)$ 는 음원신호를 나타낸다. 그러므로 추정된 행렬 $v_j(t)$ 를 다음 식과 같이 표현할 수 있으며, 여기에서 $w_{kj}(t)$ 는 분리행렬을 나타낸다.

$$v_j(t) = \sum_{j=0}^{M-1} w_{kj}(t) * x_j(t), \quad t = 0, 1, 2, \dots \quad (5)$$

본 논문에서는 식 (4)를 식 (6)과 같이 주파수영역 표현을 사용한 ICA 모델(Model of Frequency Domain Representation ICA, MFDR-ICA)의 알고리즘을 사용한다.

$$x_j(n, \omega) = \sum_{k=0}^{M-1} a_{jk}(\omega) \cdot s_n(n, \omega) \quad (6)$$

따라서 식 (6)의 주파수영역 혼합신호 $X(n, \omega)$ 를 다음 식과 같이, $a_{jk}(\omega)$ 을 구성요소로 하는 주파수 혼합행렬 $A(\omega)$ 와 음원신호의 스펙트럼 $S(n, \omega)$ 를 사용하여 나타낼 수 있다.

$$X(n, \omega) = A(\omega) \cdot S(n, \omega) \quad (7)$$

여기에서 n 은 시간 프레임의 인덱스를 나타내며, ω 는 주파수를 나타낸다. 또한 식 (7)의 $X(n, \omega)$ 에 대하여, 식 (5)의 $v_j(t)$ 를 단시간 푸리에 변환을 하면 다음 식과 같이 구할 수 있다.

$$V_j(n, \omega) = \sum_{j=0}^{M-1} W_{kj}(\omega) \cdot X_j(n, \omega) \quad (8)$$

즉, 단시간 푸리에변환하여 구해진 $V_j(n, \omega)$ 는 식 (9)의 $V(n, \omega)$ 와 같이 다시 표현할 수 있다.

$$V(n, \omega) = W(\omega) \cdot X(n, \omega) \quad (9)$$

여기에서 $V(n, \omega)$ 는 최종적으로 분리된 주파수영역의 추정 스펙트럼이며, $W(\omega)$ 는 분리된 행렬이다. 따라서 각각의 주파수 ω 에 대하여 혼합된 스펙트럼 $X(n, \omega)$ 로부터 분리행렬 $W(\omega)$ 를 추정하여 최종 분리된 스펙트럼 $V(n, \omega)$ 를 구한 후에 역 단시간 푸리에 변환함으로써 원래의 음원신호에 대응하는 분리신호 $v_j(t)$ 를 추정할 수 있다.

IV. 실험 결과 및 고찰

제안한 알고리즘에서는 입력음원의 혼합신호에 사용되는 음성신호는 표본주파수 16kHz인 남성 3명과 여성 3명의 총 6개의 영어 문장으로 구성된 음성신호를 사용한다[9-10]. 본 실험에서는 3개의 깨끗한 남성 신호와 3개의 깨끗한 여성 신호를 다른 위치의 음원신호로 사용하여 2채널 혼합신호를 얻었다. 본 논문에서 제안한 MFDR-ICA의 실험에서 분석 프레임 길이는 2048 샘플이며 1536 샘플을 중첩시켜 프레임을 이동시키면서 실험을 진행한다. 본 연구의 실험에 사용된 PC는 CPU가 인텔 코어 i7-3.0GHz이며, 매트랩 2019b를 사용하였다.

본 실험에서는 2채널에 입력되는 혼합된 음원신호를 분리하는 실험을 제안하는 MFDR-ICA 알고리즘에 의한 실험결과를 그림 2~그림 7과 같이 각각 구한다. 그림 2와 그림 3은 분리하고자 하는 원래의 음원신호1과 음원신호2를 각각 표시하며, 그림 4와 그림 5는 원래의 음원신호를 각각 혼합한 2채널에 해당하는 혼합된 음원신호를 표시한다. 그림 6과 그림 7은 본 논문에서 제안하고 있는 MFDR-ICA에 의하여 분리된 실험결과를 표시한다. 파형에 의한 실험결과로부터 확인할 수 있듯이 그림 6과 그림 7의 분리결과는 원래의 파형과 비교하여 2채널의 음원신호를 깨끗하게 분리할 수 있음을 명확히 하였다. 특히 묵음 구간에서 더욱더 명확하게 분리할 수 있었다.

제안한 방법에서는 식 (10)과 음성의 신호 대 간섭 신호의 출력 에너지의 비율(Signal to Interference Energy Ratio, SIER)[7]을 사용하여 향상된 음원분리 성능을 분석하며, 음성신호와 간섭신호의 추정된 에너지 값에 대하여 컨볼루션을 사용하여 신호 대 간섭비율이 결정된다. 따라서 혼재된 신호가 성공적으로 추출된 경우 SIER의 값은 크게 되고, 음원추출 성능이 낮아지게 되면 SIER의 값이 작게 됨을 알 수 있다.

$$OutputSIER_j = 10\log_{10} \frac{\sum_t |A_{jj}(t)s_j(t-m)|^2}{\sum_t \left| \sum_{k \neq j} A_{jk}(t)s_k(t-m) \right|^2} \quad (10)$$

여기에서 $s_j(t)$ 는 희망하는 신호이고 $s_k(t)$ 은 원

래 신호와 관계없는 간섭신호에 해당된다.

그림 8 및 그림 9는 각각의 문장에 대해서 본 논문에서 제안한 알고리즘의 음원분리 성능평가를 나타내고 있다. 출력 SIER 값에 의한 성능비교에서 알 수 있듯이 그림 8에서 제안한 알고리즘이 IVA와 비교하여 최대 14.1[dB]이 개선되었으며 Fast IVA와 비교하여 최대 14.3[dB]이 개선되었다. 또한 그림 9에서는 제안한 알고리즘이 IVA와 비교하여 최대 8.1[dB]이 개선되었으며 Fast IVA와 비교하여 최대 6.5[dB]이 개선된 것을 확인할 수 있었다. 따라서 본 논문에서 제안한 알고리즘의 음원분리 성능이 기존의 알고리즘에 비교하여 성능이 개선되었다는 것을 알 수 있었다.

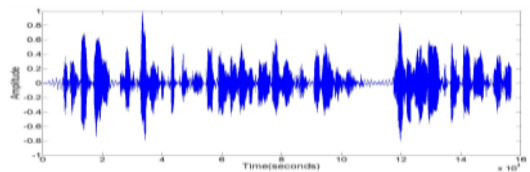


그림 2. 채널1에 대한 입력 음원신호 예
Fig. 2 An example of input speech source for channel 1

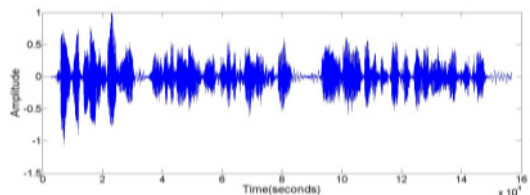


그림 3. 채널2에 대한 입력 음원신호 예
Fig. 3 An example of input speech source for channel 2

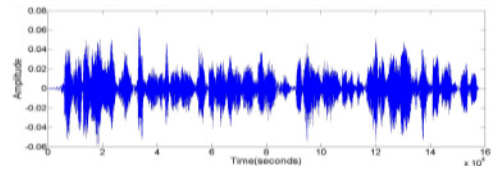


그림 4. 채널1에 대한 혼합 입력음원신호
Fig. 4 Mixed input speech source for channel 1

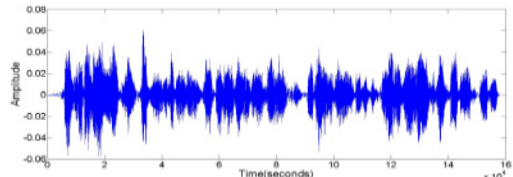


그림 5. 채널2에 대한 혼합 입력음원신호
Fig. 5 Mixed input speech source for channel 2

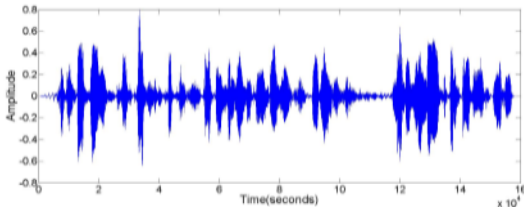


그림 6. 채널1의 분리된 출력음원신호

Fig. 6 Separated output speech source for channel 1

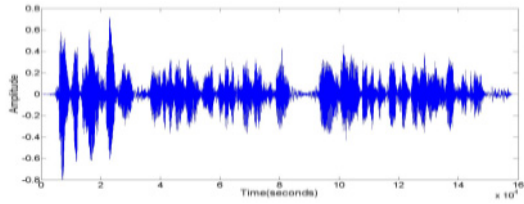


그림 7. 채널2의 분리된 출력음원신호

Fig. 7 Separated output speech source for channel 2

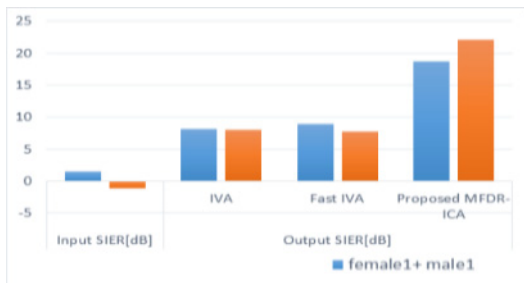


그림 8. "female1+male1"에 대한 음원분리 성능평가

Fig. 8 Performance comparison of speech source separation for "female1+male1"

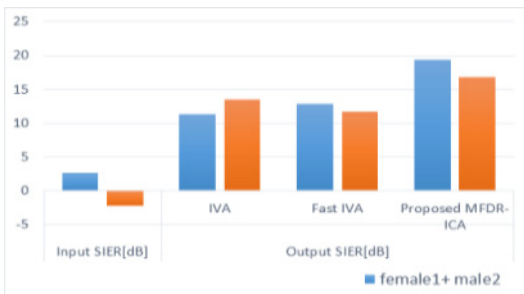


그림 9. "female1+male2"에 대한 음원분리 성능평가

Fig. 9 Performance comparison of speech source separation for "female1+male2"

V. 결 론

본 논문에서는 여러 음원신호가 혼재된 환경에 있어서 목적으로 하는 음원신호만을 분리하기 위해서 마이크로폰을 사용한 BSS 기술에 대하여 그 필요성을 기술하였으며, ICA를 이용하여 신호를 추출하기 위한 독립성분분석의 대표적 수법인 주파수영역 표현 모델에 기초한 MFDR-ICA 알고리즘을 제안하였다.

본 실험에서는 제안하고 있는 알고리즘에 의하여 분리된 실험결과를 나타내었으며, 파형의 실험결과로부터 확인할 수 있듯이 분리 실험결과와는 원래의 파형과 비교하여 2채널의 음원신호를 깨끗하게 분리할 수 있음을 명확히 하였다. 또한 출력 SIER 값을 사용하여 비교한 실험 결과로부터 본 논문에서 제안하고 있는 알고리즘의 음원분리 성능이 기존의 알고리즘에 비교하여 성능이 개선되었다는 것을 알 수 있었다.

References

- [1] K. Nakadai, H. Nakajima, G. Ince, and Y. Hasegawa, "Sound source separation and automatic speech recognition for moving sources," *IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, Taipei, Taiwan, Oct., 2010, pp. 976-981.
- [2] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, "Blind Source Separation Exploiting Higher-Order Frequency Dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, Jan. 2007, pp. 70-79.
- [3] H. T. Kim, "Vocal Separation in Music Using SVM and Selective Frequency Subtraction," *J. of Korea Institute of Electronic Communication Sciences*, vol. 10, no. 1, 2015, pp. 1-6.
- [4] C. B. Lee, "Evaluation of a signal segregation by FDBM," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 8, no. 12, 2013, pp. 1793-1802.
- [5] J. S. Choi, "Mixed Noise Cancellation by Independent Vector Analysis and Frequency Band Beamforming Algorithm in 4-channel Environments," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 5, 2019, pp. 811-816.

[6] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signals," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 3, May 2003, pp. 204-215.

[7] H. Sawada, S. Araki, and S. Makino, "Measuring Dependence of Bin-Wise Separated Signals for Permutation Alignment in Frequency-Domain BSS," *IEEE International Symposium on Circuits and Systems*, New Orleans, LA, USA, May 2007, pp. 3247-3250.

[8] Z. Chu and K. S. Bae, "Post-processing of IVA-based 2-channel blind source separation for solving frequency bin permutation problem," *Phonetics and Speech Sciences*, vol. 5, no. 4, Dec. 2013, pp. 211-216.

[9] X. Wang, X. Quan, and K. S. Bae, "Microphone Array Based Speech Enhancement Using Independent Vector Analysis," *Phonetics and Speech Sciences*, vol. 4, no. 4, Dec. 2012, pp. 87-92.

[10] J. S. Choi, "A Blind Source Separation Method Based on Independent Vector Analysis for Separation of Speech Signal and Noise Signal," *The Journal of Korean Institute of Information Technology*, vol. 16, no. 10, Oct. 2018, pp. 69-74.

[11] E. Bingham and A. Hyvarinen, "A fast fixed-point algorithm for independent component analysis for complex valued signals," *International Journal of Neural Systems*, vol. 10, no. 1, Feb. 2000, pp. 1-8.

[12] T. Nishikawa, H. Saruwatari, and K. Shikano, "Comparison of time-domain ICA, frequency-domain ICA and multistage ICA for blind source separation," *2002 11th European Signal Processing Conference*, vol. II, Sept. 2002, pp. 15-18.

[13] F. Nesta, P. Svaizer, and M. Omologo, "Convolutional BSS of short mixtures by ICA recursively regularized across frequencies," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 3, Mar. 2011, pp. 624-639.

[14] X. Quan and K. S. Bae, "Improvement of convergence speed in FDICA algorithm with

weighted inner product constraint of unmixing matrix," *Phonetics and speech sciences*, vol. 7, no. 4, 2015, pp. 17-25.

[15] H. W. Lee, "Acoustic Echo Cancellation Based on Convolutional Blind Signal Separation Method," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 13, no. 5, Oct. 2018, pp. 979-986.

저자 소개



최재승(Jae-Seung Choi)

1989년 조선대학교 전자공학과 공학사

1995년 일본 오사카시립대학 전자정보공학부 공학석사

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

2002년~2007년 경북대학교 디지털기술연구소 책임연구원

2007년~현재 신라대학교 스마트전기전자공학부 교수
※ 관심분야 : 음성신호처리, 신경회로망, 잡음제거, 음원분리 등