

낮은 신호 대 잡음비 환경에서의 퍼지 소속도 천이 C-means 클러스터링을 이용한 음성구간 검출 알고리즘

이기현^{*}, 이윤정^{**}, 조진호^{***}, 김명남^{****}

요 약

음성구간 검출은 음성과 잡음이 섞인 신호에서 음성과 잡음이 섞인 신호에서 음성구간을 찾는 과정으로 잡음제거나 음성 향상을 위한 신호처리에 매우 중요한 과정이다. 지금까지 음성구간 검출에 관한 많은 연구가 있었지만, 낮은 신호 대 잡음비 환경에서 문장형태의 음성신호에 대해서는 좋은 성능을 보이지 못하였다. 본 논문에서는 신호의 엔트로피를 이용한 초기 VAD과정을 거친 후, 퍼지 소속도 천이 c-means 클러스터링 방법을 이용해 주 VAD과정을 거치는 새로운 VAD알고리즘을 제안한다. 제안한 알고리즘의 성능을 비교 평가하기 위하여 백색잡음의 다양한 신호 대 잡음비 환경에서 실험을 수행하였으며 실험결과, 제안한 방법의 우수한 성능을 확인할 수 있었다.

Voice Activity Detection Algorithm using Fuzzy Membership Shifted C-means Clustering in Low SNR Environment

G.H. Lee[†], Y.J. Lee^{**}, J.H. Cho^{***}, M.N. Kim^{****}

ABSTRACT

Voice activity detection is very important process that find voice activity from noisy speech signal for noise cancelling and speech enhancement. Over the past few years, many studies have been made on voice activity detection, it has poor performance for speech signal of sentence form in a low SNR environment. In this paper, it proposed new voice activity detection algorithm that has beginning VAD process using entropy and main VAD process using fuzzy membership shifted c-means clustering. We conduct an experiment in various SNR environment of white noise to evaluate performance of the proposed algorithm and confirmed good performance of the proposed algorithm.

Key words: Speech Detection(음성 검출), Fuzzy Clustering(퍼지 클러스터링), Fuzzy Membership Shift(퍼지 소속도 천이)

※ 교신저자(Corresponding Author): 김명남, 주소: 대구광역시 중구 국제보상로 680, 경북대학교 의학전문대학원 의공학교실(700-842), 전화: 053) 200-5266, FAX: 053) 200-5264, E-mail: kimmn@knu.ac.kr
접수일: 2013년 12월 4일, 수정일: 2014년 1월 15일
완료일: 2014년 2월 11일

^{*} 경북대학교 대학원 의용생체공학과
(E-mail: gihyoulee@gmail.com)

^{**} 경북대학교 대학원 의용생체공학과
(E-mail: whitegleam@nate.com)

^{***} 경북대학교 IT대학 전자공학부
(E-mail: jhcho@ee.knu.ac.kr)

^{****} 경북대학교 의학전문대학원 의공학교실

※ 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (No. 2013 R1A2A1A09015677) 또한 본 연구는 보건복지가족부 보건 의료기술진흥사업의 지원에 의하여 이루어진 것임 (과제고유번호: A092106)

1. 서 론

음성구간 검출(VAD, voice activity detection) 알고리즘은 음성인식, 잡음제거기 등의 다양한 음성처리 시스템에 적용되고 있으며 시스템 성능에 주요한 영향을 미치는 핵심 부분으로 인식되고 있다[1]. 최근 휴대용 IT기기 등에 음성인식 기술을 이용한 여러 가지 기능이 발표되면서 정확한 음성인식에 대한 연구의 필요성이 대두되고 있으며 이러한 환경에서 음성구간 검출은 음성인식에 앞서 반드시 선행되어야 할 연구로써 그 중요성이 더욱 강조되고 있다.

현재 사용되고 있는 음성 인식기는 잡음 제거 기법의 위너필터(Wiener filter)와 칼만 필터(Kalman filter)를 이용한다. 이러한 방식은 정상성(stationary)잡음에서 높은 잡음제거 성능과 함께 좋은 음성인식 성능을 보인다. 하지만 음성인식 전처리 단계에서 잡음제거 단계를 필수적으로 수반하여야 하며 비정상성(nonstationary)잡음에서는 좋은 성능을 보이지 못하는 단점이 있다[2]. 비정상성잡음에서 음성인식을 하기 위해서는 특징추출 방식을 사용한다. 기존의 특징추출방법 중 가장 널리 알려진 방법인 신호의 에너지와 영교차율(zero crossing rate, ZCR)을 이용한 방법들은 작은 연산량과 높은 음성인식 성능을 보이지만 신호의 에너지를 이용한 방법은 낮은 신호 대 잡음비(signal to noise ratio, SNR) 환경에서 성능이 급격히 저하되는 단점이 있으며 영교차율을 이용한 방법은 잡음의 종류에 따라 무성음과 구분되지 않는 단점을 보였다[3]. 잡음에 더 강인한 특징추출을 위한 우도비(likelihood ratio, LR), 엔트로피(entropy) 등과 같은 통계적 특징들을 이용하는 방법이 있다[4]. 이러한 통계적인 특징을 이용하는 방법들의 경우, 잡음의 특징 추출에서 좋은 성능을 보여주고 있지만 연산량이 많거나 음성과 통계적 특징이 비슷한 잡음에서는 좋은 성능을 보여주지 못하였다[4-8]. 또한 Asgari[9]는 주파수 영역에서의 엔트로피를 이용한 음성구간 검출 방법을 제안하였으나 단음절의 음성구간 검출에는 좋은 성능을 보이지만 낮은 SNR환경에서 문장단위의 이어지는 음성에서는 좋은 성능을 보이지 못했다. 그리고 표준 음성 코덱으로 상용 통신제품에 많이 쓰이고 있는 ITU-T의 VAD알고리즘인 G729B는 잡음이 거의 없는 환경에서는 좋은 성능을 보이지만 SNR이 낮은 환경에서는

매우 좋지 않은 성능을 보이는 단점이 있다.

본 논문에서는 신호의 엔트로피를 이용한 초기 VAD과정을 거친 후 퍼지 소속도 천이 c-means 클러스터링 방법을 이용해 주 VAD과정을 거치는 새로운 VAD알고리즘을 제안한다. 제안하는 퍼지 소속도 천이 c-means 클러스터링 방법은 초기 VAD과정에서 시간에 따른 엔트로피를 이용해 음성구간과 잡음구간을 구분한다. 여기서 잡음으로 판단된 구간을 퍼지 소속도 천이과정을 거친 신호의 엔트로피와 에너지, 두 종류의 특징들을 퍼지 소속도로 활용해 음성구간과 잡음구간의 두 개의 클러스터로 나누어 최종적으로 음성구간과 비 음성구간을 결정한다. 제안한 알고리즘의 유효성을 평가하기 위하여 강한 잡음환경에서 실험하였으며 기존의 알고리즘들이 좋은 성능을 보여주지 못한 낮은 SNR환경과 문장형태의 음성신호로 실험을 수행하였다. 실험 결과를 음성구간 검출 그래프와 객관적인 수치로 나타내어 기존의 알고리즘과 비교하였다. 그 결과, 제안한 방법이 기존의 방법에 비하여 성능이 우수한 것을 확인할 수 있었다.

2. 제안하는 음성구간 검출 알고리즘의 개요

2.1 제안한 알고리즘의 개요

본 논문에서는 잡음이 섞인 음성신호에서 음성구간을 검출하기 위해 fuzzy c-means 클러스터링에 기반한 새로운 음성구간 검출 알고리즘을 제안하였다. 그림 1은 제안하는 음성구간 검출 알고리즘의 전체적인 흐름도이다.

먼저, 잡음이 섞인 음성 신호가 입력되면 입력신호의 엔트로피를 계산한다. 그리고 계산된 엔트로피를 이용하여 초기 음성구간 검출을 수행한다. 초기 음성구간 검출을 통해 잡음으로 판단된 신호를 퍼지 소속도 천이 기법을 거친 후, 제안하는 퍼지 소속도 천이 c-means 클러스터링으로 주 음성구간 검출을 한다. 이러한 과정을 거쳐 음성과 잡음이 분리된 음성구간 검출 신호를 얻게 된다.

2.2 엔트로피 음성구간 검출

엔트로피는 역열학적 통계으로 ‘통계적 무질서도’를 의미하며 정보통신이나 신호처리 분야에서는 ‘데

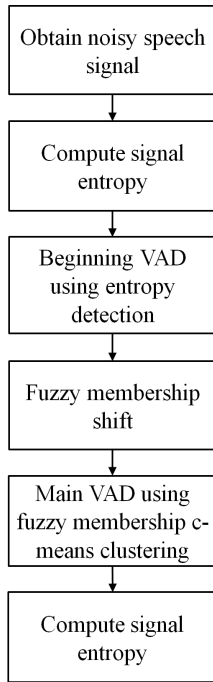


그림 1. 제안하는 음성구간 검출 알고리즘의 블록선도

이터에 내재되어 있는 정보의 양'을 나타내는 척도로 사용된다[10]. 음성신호에서 데이터에 내재되어 있는 통계적 특징을 시간에 따른 엔트로피 변화로 나타내어 추출할 수 있다. 시간 영역에서 특정지점 i 에서의 신호 $x(i)$ 의 확률 $p[x(i)]$ 은 음성신호 전체에서 신호 $x(i)$ 의 상대적 크기로 나타내어 다음과 같이 정의할 수 있다.

$$p[x(i)] = \frac{|x(i)|}{\max(x)} \quad (1)$$

여기서 $\max(x)$ 는 프레임 내 $x(i)$ 의 최대치이며, 주어진 음성신호의 특정 프레임 k 에서 엔트로피 $H(k)$ 는 식 (2)와 같이 정의된다[10].

$$H(k) = - \sum_{i=k-l/2}^{k+l/2} p[x(i)] \log(p[x(i)]) \quad (2)$$

여기서 l 은 엔트로피 계산을 위한 프레임의 길이이다. 또한 식 (2)를 이용해 연속된 데이터에 대해 신호의 엔트로피를 나타내고 프레임 내의 평균엔트로피와 엔트로피의 표준편차에 대한 엔트로피의 시간에 따른 변화를 계산하기 위해 식 (2)에서 계산한 시간에 따른 엔트로피 $E(i)$ 를 다음과 같이 변형하여 전개하였다[11].

$$E(t) = \frac{H(k) - M(H)}{S(H)} \quad (3)$$

여기서 $M(H)$ 는 H 의 평균, $S(H)$ 는 표준 편차이다. 식 (3)을 통해 얻은 음성신호의 엔트로피와 실험에 사용된 음성신호를 그림 2에 나타내었다.

그림 2 (a)와 (b)는 각각 백색잡음이 SNR 5 dB로 섞인 음성신호와 그 음성신호의 엔트로피이다. (a)와 (b)를 비교하여 볼 때 신호의 엔트로피로 나타낸 (b)에서 음성구간은 상대적으로 강조되고 잡음구간은 감쇄된 것을 볼 수 있다. 본 논문에서는 이러한 특징을 이용하기 위해 엔트로피를 이용한 초기 음성구간 검출을 수행한다. 이러한 과정은 식 (4)에 따라 수행된다.

$$VAD_1(t) = \begin{cases} E(t) > 0, & 1 \\ E(t) \leq 0, & 0 \end{cases} \quad (4)$$

여기서 음성구간 검출의 문턱치는 0으로 설정하여 초기 음성구간 검출을 수행하였다.

2.3 FCMC(Fuzzy C-Means Clustering)

일반적인 FCMC알고리즘은 퍼지 집합 이론과 최소자승 에러 평가에 기반을 두고 있으며 소속도와 명기된 정보에 따라 각 데이터 점을 소속 정도에 따라 클러스터로 구성하는 데이터 분류 알고리즘이다 [12,13]. n 개의 벡터 $x_i, i=1, \dots, n$ 의 집합을 c 개의 퍼지 집합으로 나누고 각 집합의 클러스터들의 중심을 찾는다[12]. FCMC에서 소속도 u_{ij} 는 0과 1사이의 값을 나타내며 각 클러스터에 속하는 소속정도의 합은 1이 된다[11]. n 개의 x_i 집합의 \bar{x}_k 벡터와 클러스터의 중심인 c_i 사이의 차이를 측정하는 유클리드 거리와 그 목적함수는 식 (5)와 같이 일반화 된다[12,13].

$$\mathcal{J}(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j^n u_{ij}^m d_{ij}^2 \quad (5)$$

여기서 $d_{ij} = \|c_i - x_j\|$ 는 j 번째 데이터와 i 번째 클러스터 중심값 사이의 유클리드 거리로써 식 (6)과 같이 정의되는 정규화된 유클리드 거리를 사용한다. 또한 m 은 $m \in [1, \infty]$ 인 가중치이다.

$$d_{ij} = \|c_i - x_j\| = \sqrt{\sum_{l=1}^r \frac{c_{il} - x_{jl}}{\sigma_l^2}} \quad (6)$$

여기서 r 은 입력공간의 차원이고 σ_l^2 은 l 번째 변수의 분산이다. 정규화된 유클리드 거리를 사용하는 이

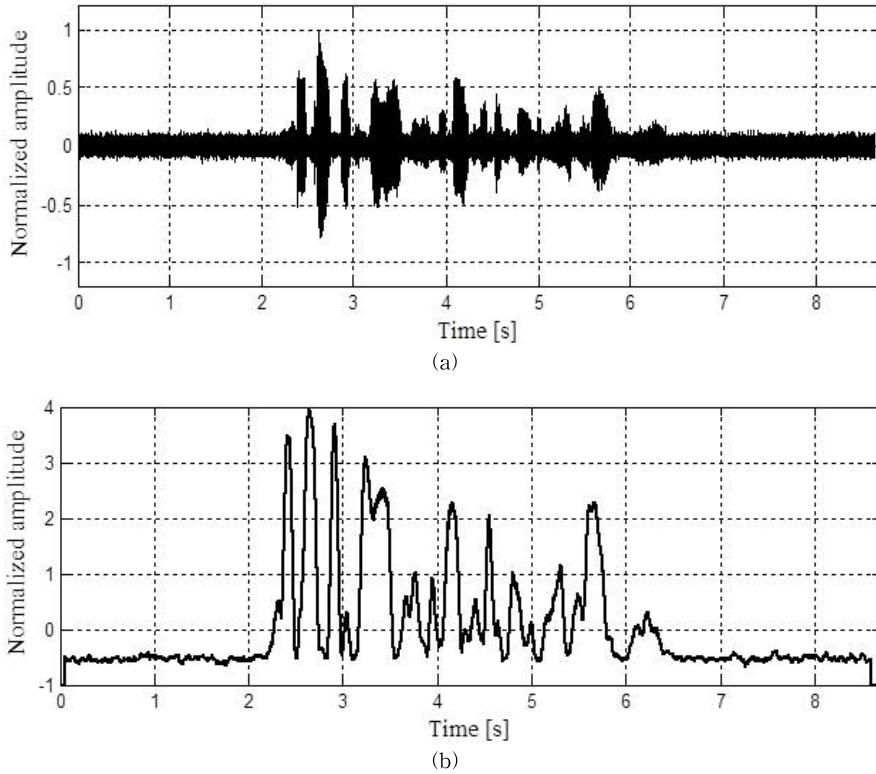


그림 2. 음성신호의 엔트로피 (a) 잡음이 섞인 음성신호, (b) 잡음이 섞인 음성신호의 엔트로피

유는 큰 값을 갖는 입력변수가 작은 값을 가지는 입력변수보다 클러스터의 중심을 결정하는데 더 많은 영향을 미치는 것을 방지하기 위함이다[13]. 그리고 식 (6)이 최소가 되기 위한 필요조건은 식 (7)과 같다 [12].

$$\begin{aligned} \bar{J}(U, c_1, \dots, c_c, \lambda_1, \dots, \lambda_n) &= J(U, c_1, \dots, c_c) + \sum_{j=1}^n \lambda_j (\sum_{i=1}^c u_{ij} - 1) \\ &= \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 + \sum_{j=1}^n \lambda_j (\sum_{i=1}^c u_{ij} - 1) \end{aligned} \quad (7)$$

여기서 λ_j 는 라그랑지(lagrange) 곱셈자이며, 식 (7)이 최소가 되기 위한 클러스터의 중심과 퍼지 소속도의 조건은 다음과 같다[12].

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \quad (8)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \quad (9)$$

FCMC은 반복 처리로 식 (8)과 식 (9)을 갱신시킨다. 더 이상 향상시키지 않을 때까지 반복처리를 하여 식 (8)과 식 (9)의 조건이 만족하도록 하여 소속행렬 U 와 클러스터 중심 c_i 를 결정한다. 그리고 식 (9)는 거리에 기반한 소속도를 계산하며 그 합이 1이 되도록 정규화한다[12,13].

2.3 퍼지 소속도 천이 c-means 클러스터링

앞서 설명과 같이 기존의 FCMC 알고리즘은 입력 벡터들과 각 군집의 중심과의 거리를 이용하여 측정된 유사도에 기초한 목적 함수의 최적화 방식을 사용한다[12]. 그러나 음성구간 검출에 사용될 경우, SNR이 낮은 환경에서는 군집공간상의 거리만을 이용한 FCM 알고리즘은 음성과 크기가 큰 잡음을 구별하지 못하는 결과를 보일 수 있다. 이러한 단점을 극복하기 위해 본 논문에서는 신호의 에너지와 엔트로피 두 가지 특성을 퍼지 소속도로 사용하였으며, 음성과 잡음으로 구분되는 2개의 클러스터의 중심을 업데이트하는 과정에서 데이터의 소속도를 가중시키는 새

로운 퍼지클러스터링 방법을 제안한다. 제안하는 알고리즘의 순서도를 그림 3에 나타내었다.

먼저, 입력된 음성신호 $x(t)$ 를 앞서 설명한 엔트로피를 이용하여 초기 VAD과정을 거친다. 초기 VAD 결과에서 음성으로 판단된 구간은 음성으로 결정하고 잡음으로 판단된 구간에 대하여 제안하는 퍼지 소속도 천이 c -means 클러스터링 방법이 적용되는 주 VAD과정을 거친다. 먼저 초기 VAD결과에서 음성으로 판단된 구간의 퍼지 소속도를 천이시키고 퍼지 c -means 클러스터링 방법으로 두 개의 클러스터로 분류하여 음성구간과 비 음성구간을 검출한다. 검출된 음성구간 결과를 평활화과정을 거쳐 최종 음성구간 검출을 한다.

퍼지 소속도 천이 c -means 클러스터링에는 신호의 엔트로피와 에너지에 대한 두 가지 특징을 사용하여 퍼지 소속도를 결정한다. 주어진 신호의 특정 시간(t)에서 두 가지 퍼지 소속도 함수는 다음 식(10)과 식(11)을 통해 계산된다.

$$UX(t) = x(t)^2 \tag{10}$$

$$UY(t) = E(t) \tag{11}$$

여기서 $UX(t)$ 는 입력신호의 에너지이며 이것을 첫 번째 퍼지 소속도로 사용한다. 그리고 $UY(t)$ 는 식(3)에서 계산된 입력신호의 엔트로피이다. 이것을 두 번째 퍼지 소속도로 활용한다. 그리고 식 (10)과 식(11)을 이용하면 앞서 설명한 식 (8)과 (9)의 일반적인 FCMC알고리즘의 조건은 다음과 같이 변형하여 나타낼 수 있다.

$$CX_i = \frac{\sum_{j=1}^n UX_{ij}^m x_j}{\sum_{k=1}^c \left(\frac{d_{kj}}{d_{kj}}\right)^{2/(m-1)}} \tag{12}$$

$$CY_i = \frac{\sum_{j=1}^n UY_{ij}^m x_j}{\sum_{k=1}^c \left(\frac{d_{kj}}{d_{kj}}\right)^{2/(m-1)}} \tag{13}$$

$$UX_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{kj}}{d_{kj}}\right)^{2/(m-1)}} \tag{14}$$

$$UY_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{kj}}{d_{kj}}\right)^{2/(m-1)}} \tag{15}$$

그리고 초기 음성구간 검출법에서 검출된 음성구간과 잡음구간의 클러스터 중심을 각각 c_s 와 c_n 으로 나누어 클러스터 중심을 계산한다. 다음의 과정을 거쳐 제안하는 퍼지 소속도 천이 c -means 클러스터링의 소속도를 천이시키고 천이된 소속도로 음성구간과 잡음구간의 클러스터 중심을 계산하여 음성구간 검출을 수행한다.

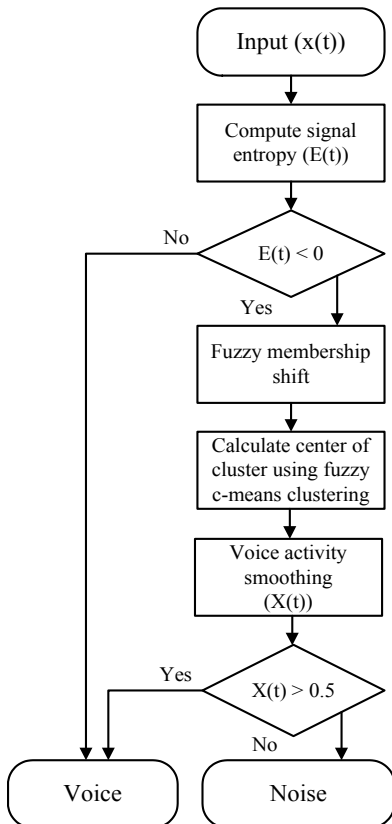


그림 3 음성구간 검출 알고리즘 계산과정의 흐름도

Seed weighted fuzzy c-means clustering

Given the cluster number $c = 2$
 Choose initial estimate of cluster centers c_s, c_n
 REPEAT
 IF $UY(t) > 0$
 $UY(t) = \text{mean}(UY(1:t));$
 $UY(t) = \text{mean}(UX(1:t));$
 END
 Determine $UX(t), UY(t)$, according to Eq (14), (15)
 Update the cluster centers c_s, c_n , according to Eq (12), (13)
 UNTIL no change in the cluster centers

클러스터 개수와 음성구간의과 잡음구간의 클러스터 중심을 초기화 시킨 후 입력신호의 엔트로피가 0보다 큰 구간에서 소속도 UX 와 UY 를 현재시간까지의 평균값으로 바꾸어 음성구간의 소속도를 천이 시킨다. 천이된 소속도 값으로 인해 음성구간의 소속도가 평균값에 가까운 곳으로 밀집되어 소속도 간의 유클리드 거리가 줄어들어 음성구간의 클러스터를 결정할 때 좋은 성능을 가지게 된다. 그리고 식 (14), (15)를 통해 소속도를 결정하고 음성구간의 클러스터 중심과 잡음구간의 클러스터 중심을 식 (12), (13)으로 클러스터의 중심이 변화가 없을때까지 갱신시켜 나누어진 두 개의 클러스터로 음성구간과 잡음구간을 검출한다. 그리고 일반적으로 문장을 말할 때 음성은 음성구간끼리 모여 있는 특징이 있다. 이러한 특징을 이용하면 음성구간 검출 성능을 높일 수 있다. 특정 시점(t)에서 $200ms$ 전후로 음성 클러스터와 잡음 클러스터의 개수를 확인하여 클러스터의 개수가 더 많은 쪽으로 결정하는 음성구간 평활화 과정을 거친 후 최종적으로 음성구간 검출을 한다.

3. 실험 결과 및 고찰

제안한 알고리즘의 성능분석을 위한 실험 데이터로써 TIMIT 데이터베이스의 음성신호 샘플과 NOISEX-92의 잡음신호 샘플을 사용하였다. 데이터 샘플은 16kHz 샘플링레이트(sampling rate)와 32비트(bit)를 가지며 다양한 SNR에 대한 알고리즘의 성능을 평가하기 위해 여러 가지 잡음에 대해 SNR을 0 dB, 5 dB, 10 dB로 구분하여 실험하였다. 제안한 알고리즘의 성능을 확인하기 위해 대부분의 음성구간 검출 알고리즘의 성능비교에 많이 쓰이고 있는 ITU-T의 G729B 음성구간 검출 알고리즘과 Asgari [9]가 제안한 EVAD(entropy voice activity detection) 알고리즘과 비교하였다. 백색 잡음을 섞어 제안한 알고리즘과 기존의 알고리즘들을 이용해 음성구간을 검출한 결과를 그림 4에 나타내었다.

그림 4는 잡음이 비교적 약한 SNR 10 dB에서의 음성구간 검출 결과이다. (a)는 G729B의 음성구간 검출 결과로 약한 잡음에도 많은 잡음구간을 음성구간으로 검출하는 좋지 못한 결과를 보였다. (b)는 EVAD의 결과로 대부분의 음성구간을 잘 찾아내고 있지만 이어지는 음성구간을 모두 분리하여 검출하

여 문장형태의 음성구간을 인식하지 못하는 단점을 보여주고 있다. 그리고 (c)는 제안하는 알고리즘의 결과로 모든 음성구간을 잘 검출했으며 (b)에서 보여주는 문장형태의 음성구간을 인식하지 못하는 부분을 앞서 설명한 음성구간 평활화 과정으로 보정하였기 때문에 이어지는 음성구간을 넓은 음성구간으로 인식하는 결과를 보였다.

그림 5는 비교적 잡음이 강한 SNR 5 dB에서 음성구간 검출을 한 결과이며 그림 6은 잡음이 상당히 강한 SNR 0 dB에서의 음성구간 검출 결과이다. G729B의 경우 잡음이 강한 환경에서는 대부분의 잡음구간을 음성구간으로 잘못 인식하는 결과를 보였으며 EVAD의 경우는 잡음구간을 음성구간으로 잘못 인식하는 경우는 없었으나 SNR 10 dB에서의 결과와 마찬가지로 이어지는 음성구간을 제대로 인식하지 못하였다. 그리고 잡음이 강한 환경에서는 크기가 큰 음성구간만을 음성으로 인식하여 많은 음성구간을 잡음으로 인식하는 오류를 보인다.

세 가지 알고리즘에 대해 객관적인 성능평가를 위해 음성구간 적중률(pause hit rate, PHR)과 비음성구간 오류율(false alarm rate, FHR)을 통해 수치적으로 나타내었다. PHR은 음성구간 내에서 음성을 음성으로 정확하게 찾아낸 비율을 나타내며 FAR은 잡음구간 내에서 잡음을 음성으로 잘 못 찾아낸 오류율을 나타낸다. PHR은 높을수록 FAR은 낮을수록 음성구간 인식에 좋은 성능을 가진 알고리즘이다. 세 가지 알고리즘의 다양한 SNR환경에 따른 PHR, FAR결과를 표 1에 나타내었다.

표 1에서 G729B의 경우 PHR에서 매우 높은 검출율을 보인다. 하지만 FAR 역시 매우 높은 오류율을 보여 잡음구간에서 거의 대부분의 잡음을 음성으로 잘못 판단하는 오류를 보이는 것을 볼 수 있다. EVAD의 경우 SNR환경에 따라 55~63%정도의 PHR을 보였으며 FAR은 모두 0을 나타내었다. 이것은 잡음구간에서 모든 잡음을 잡음으로 판단하였고 음성구간에서 음성을 검출하는 성능이 다소 부족한 것으로 판단된다. 제안하는 알고리즘은 매우 높은 PHR을 보이며 12~14%정도의 FAR을 보여 음성구간에서 대부분의 음성을 잘 찾아내고 잡음구간에서는 약간의 오류는 있지만 비교적 정확하게 잡음으로 구별하는 것을 볼 수 있다. 하지만 PHR과 FAR로는 음성구간과 잡음구간 모두의 객관적인 성능을 알 수

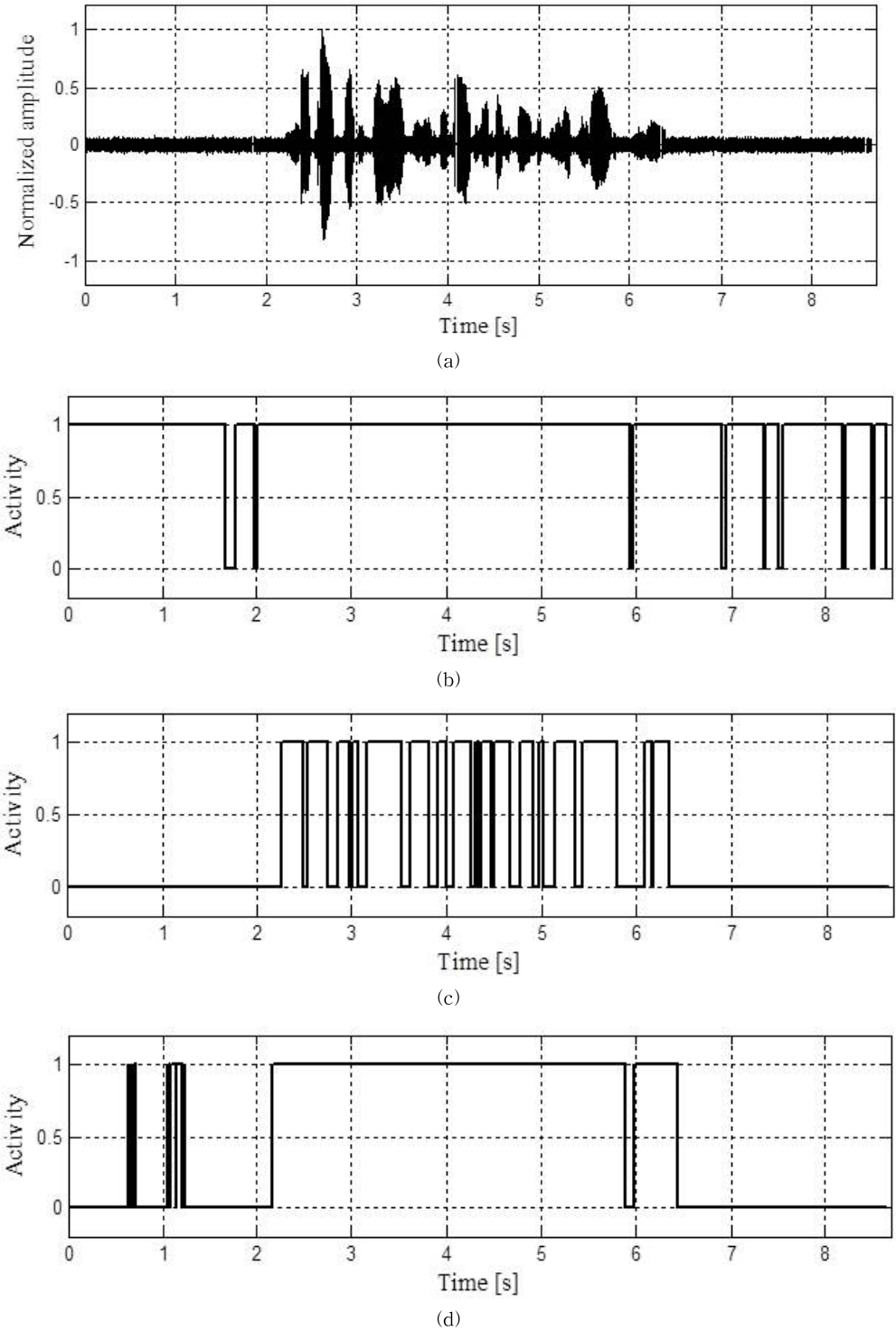
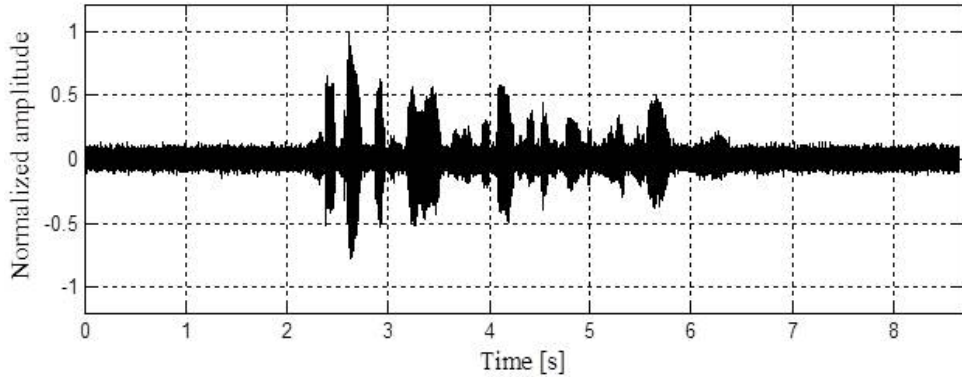
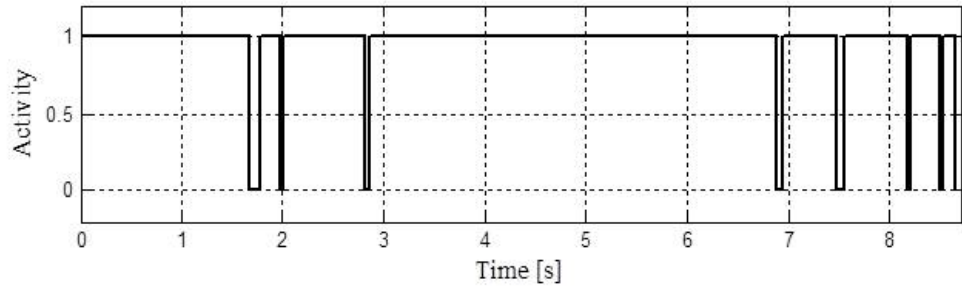


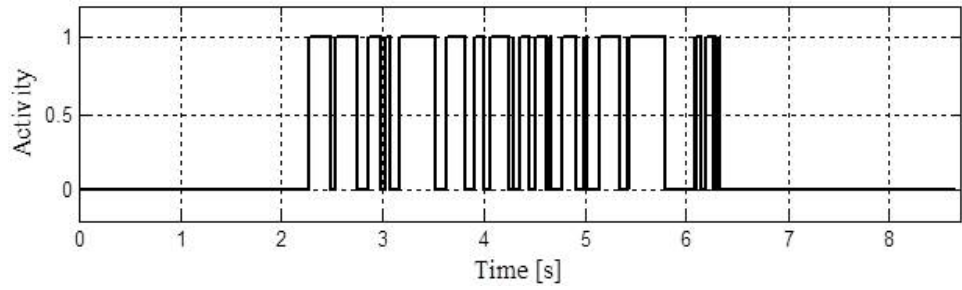
그림 4. SNR 10 dB에서의 음성구간 검출 결과 (a) 입력신호 (b) G729B, (c) EVAD, (d) 제안한 알고리즘



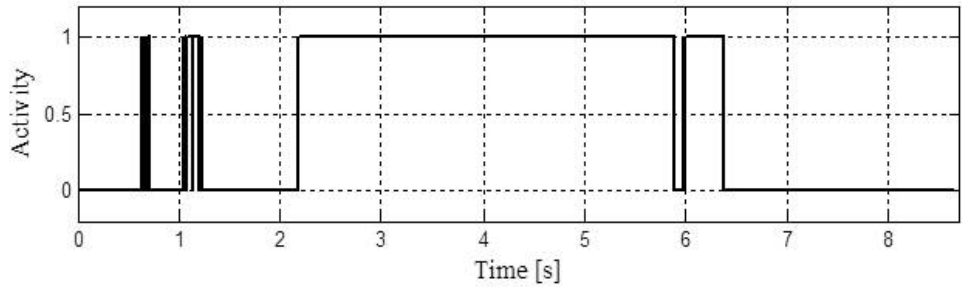
(a)



(b)

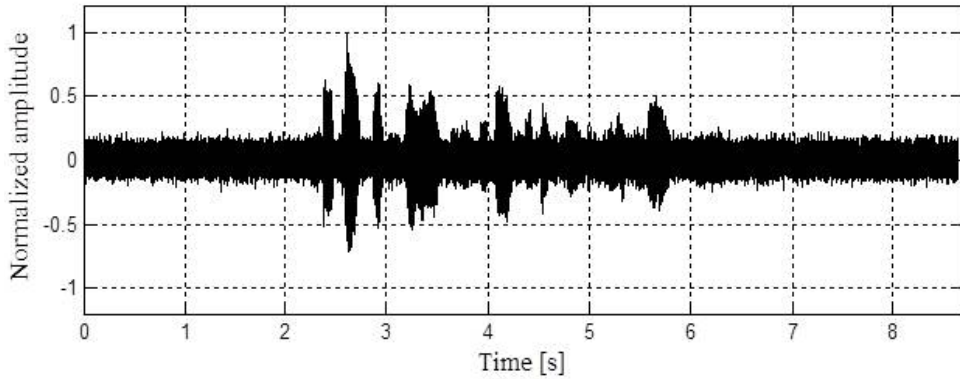


(c)

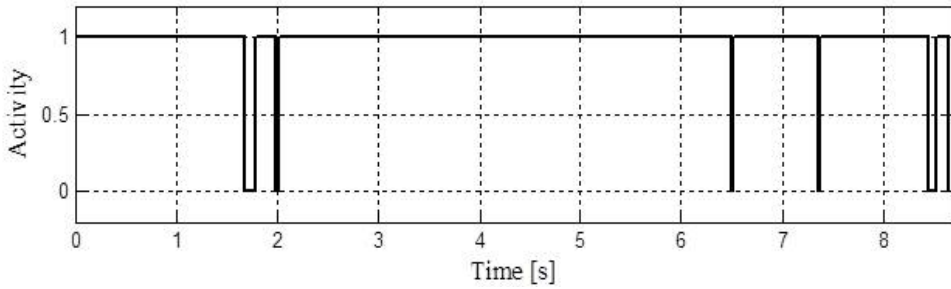


(d)

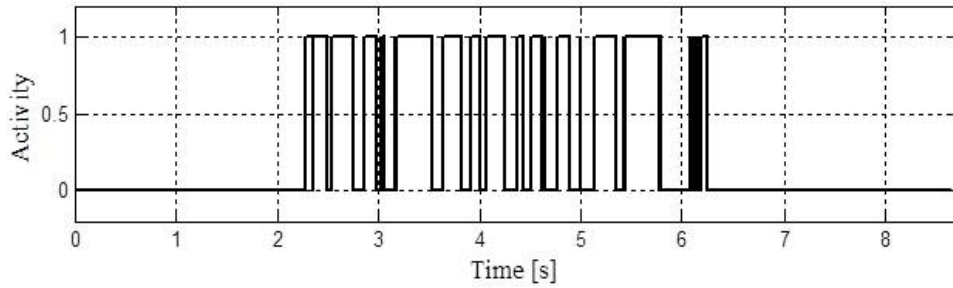
그림 5. SNR 5 dB에서의 음성구간 검출 결과 (a) 입력신호 (b) G729B, (c) EVAD, (d) 제안한 알고리즘



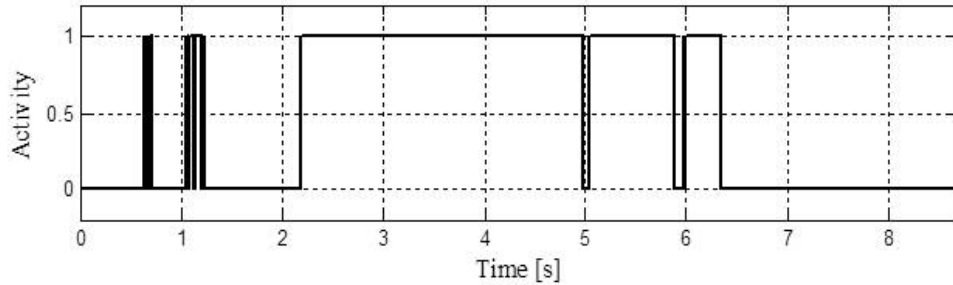
(a)



(b)



(c)



(d)

그림 6. SNR 0 dB에서의 음성구간 검출 결과 (a) 입력신호 (b) G729B, (c) EVAD, (d) 제안한 알고리즘

표 1. 음성구간 검출률 [%]

SNR (dB)	Algorithm					
	G729B		EVAD		Proposed	
	PHR	FAR	PHR	FAR	PHR	FAR
0	99.53	94.9	53.93	0	92.48	5
5	98.83	91.92	58.8	0	94.83	5.05
10	99.3	92.83	64.51	0	96.34	4.49
15	100	96.04	69.75	0	96.94	5.63

없기 때문에 신호 전체에서 음성구간 검출 성능의 객관적인 지표가 될 수 있는 상대적 오류율 (relative error, RE)을 이용하여 세 가지 알고리즘의 성능을 평가하려고 한다. RE는 잡음이 섞이지 않은 깨끗한 음성신호에서 수동으로 음성구간을 찾아 표시한 데이터와 일치하지 않는 비율을 나타낸다. RE는 다음의 식 (16)과 식 (17)을 통해 계산된다.

$$v(t) = \begin{cases} 1, & \text{if } VAD_m(t) = VAD_a(t) \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

$$RE = \frac{\sum_{t=1}^L (v(t))}{L} \times 100 \quad (17)$$

여기서 VAD_m 은 수동으로 찾은 음성구간이며 VAD_a 은 알고리즘을 통해 찾은 음성구간이다. 그리고 L 은 음성신호의 전체 샘플 개수이다. RE는 백분율로 수치가 나타나며 낮을수록 음성구간 검출 성능이 좋음을 의미한다. 앞서 언급한 세 가지 알고리즘의 음성구간 검출 상대적 오류율을 표 2에 나타내었다.

표 2에서 G729B의 경우에는 대체적으로 40% 후반대의 높은 오류율을 보였으며 EVAD는 높은 SNR 환경에서는 비교적 좋은 성능을 보이지만 낮은 SNR 환경에서는 20%이상의 높은 오류율을 보였다. 제안하는 알고리즘은 SNR이 낮아질수록 약간의 오류율 상승이 있으나 전체적으로 7% 미만의 매우 우수한

표 2. 상대적 오류율 [%]

SNR (dB)	Algorithm		
	G729B	EVAD	Proposed
0	48.08	22.84	6.25
5	34.24	20.43	5.11
10	47.15	17.59	4.49
15	48.42	15.00	4.36

음성구간 검출 성능을 보여주었다.

4. 결론

본 논문에서는 퍼지 클러스터링에 기반한 새로운 음성구간 검출 알고리즘을 제안하였다. FCMC 알고리즘을 개선한 퍼지 소속도 천이 c-means 클러스터링 알고리즘을 이용하여 음성구간 검출을 하였다. 기존의 음성구간 검출 알고리즘은 잡음이 심한 환경에서 좋은 성능을 보이지 못하였으며 기본적인 퍼지 c-means 클러스터링 알고리즘 역시 낮은 SNR 환경에서 많은 오류를 보였다. 제안하는 음성구간 검출 알고리즘은 이러한 단점을 극복하기 위해 초기 음성구간 검출에서 음성으로 검출된 구간의 퍼지 소속도를 천이시키는 방법을 이용하여 음성구간 검출 성능을 높이고 퍼지 c-means 클러스터링 방법을 통해 두 개의 클러스터로 나누어 음성구간과 비음성구간을 검출하였다. 그리고 낮은 SNR 환경에서 더 좋은 성능을 나타내기 위해 음성구간 검출결과를 평활화하는 과정을 거쳐 더욱 높은 음성구간 검출 성능을 가지도록 하였다. 실험을 통해 제안한 음성구간 검출 알고리즘이 기존의 알고리즘보다 좋은 성능을 보이는 것을 확인하였으며 특히 낮은 SNR 환경에서 좋은 음성구간 검출 성능을 보이는 것을 확인하였다. 제안된 음성구간 검출 알고리즘은 잡음에 매우 강하며 음성통신을 위한 음성코딩이나 음성인식, 잡음제거, 음성향상 등의 분야에 적용할 수 있다. 특히 잡음이 심한 환경에서 스마트기기를 사용하거나 잡음이 강한 환경에서 사용되는 기기에서 음성명령을 내릴 때 주변 소음에 영향을 받지 않고 사용할 수 있을 것으로 보인다. 향후 다양한 분야에 활용될 수 있을 것으로 기대된다.

참고 문헌

[1] L. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] ETSI, *ETSI standard doc ETSI ES 202 050 v1.1.1*, 2002.

[3] G.K. Choi and S.H. Kim, "Voice Activity Detection Method using Psycho-Acoustic

- Model based on Speech Energy Maximization in Noisy Environments," *Journal of the Acoustical Society of Korea*, Vol. 28, No. 5, pp. 447-453, 2009.
- [4] D.G. Ha, S.J. Cho, G.G. Jin, and O.K. Shin, "Voice Activity Detection based on Signal Energy and Entropy-difference in Noisy Environments," *Journal of the Korean Society of Marine Engineering*, Vol. 32, No. 5, pp. 768-774, 2008.
- [5] J. Ramírez, J.C. Segura, C. Benítez, A. de la Torre, and A. Rubio, "An Effective Subband OSF-based VAD with Noise Reduction for Robust Speech Recognition," *IEEE Trans. on Speech and Audio Processing*, Vol. 13, No. 6, pp. 1119-1129, 2005.
- [6] R. Gemello, F. Mana, and R.D. Mori, "A Modified Ephraim-Malah Noise Suppression Rule for Automatic Speech Recognition," *Proc. the ICASSP 2004*, Vol. 1, pp. 957-960, 2004.
- [7] P. Teng and Y. Jia "Voice Activity Detection Via Noise Reducing using Non-Negative Sparse Coding," *IEEE Signal Processing Letters*, Vol. 20, Issue 5, pp. 475-478, 2013.
- [8] S.W. Deng and J.Q. Han, "Statistical Voice Activity Detection based on Sparse Representation over Learned Dictionary," *Digital Signal Processing*, Vol. 23, Issue 4, pp. 1228-1232, 2013.
- [9] M. Asgari, A. Sayadian, M. Farhadloo, and E.A. Mehrizi, "Voice Activity Detection using Entropy in Spectrum Domain," *Telecommunication Networks and Applications Conference*, pp. 407-410, 2008.
- [10] C.E. Shannon, "A Mathematical Theory of Communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol. 5, No. 1, pp. 3-55, 2001.
- [11] G.H. Lee, P.U. Kim, Y.J. Lee, and M.N. Kim, "Detection of The First And Second Heart Sound using Three-order Shannon Energy Difference," *Journal of the Korea Multimedia Society*, Vol. 14, No. 7, pp. 884-894, 2011.
- [12] J.W. Jung and J.T. Lee, "A Study on the Fault Current Discrimination using Enhanced Fuzzy C-Means Clustering," *Transactions of KIEE*, Vol. 57, No. 11, pp. 2102-2107, 2008.
- [13] J.N. Chio, S.K. Oh, and H.K. Kim, "Genetic Optimization of Fuzzy C-Means Clustering-Based Fuzzy Neural Network," *Transactions of KIEE*, Vol. 57, No. 3, pp. 466-472, 2008.



이 기 현

2009년 8월 경북대학교 천문대기
과학과 (이학사)
2012년 2월 경북대학교 대학원 의
용생체공학과 (공학석사)
2012년 3월~현재 경북대학교 대
학원 의용생체공학과 박
사과정

관심분야: 생체신호처리, 의용전자기기



이 윤 정

2003년 2월 경북대학교 전자전기
공학부 (공학사)
2005년 2월 경북대학교 대학원 의
용생체공학과 (공학석사)
2014년 2월 경북대학교 대학원
의용생체공학과 (공학박
사)

관심분야: 생체신호처리, 의용전자기기



조 진 호

1977년 2월 경북대학교 공업교육
학과 전기공학전공 (공학
사)
1979년 2월 경북대학교 대학원
전자공학과 (공학석사)
1988년 2월 경북대학교 대학원
전자공학과 (공학박사)

1984년~현재 경북대학교 IT대학 전자공학부 교수
2002년 12월 ~현재 경북대학교 의공학연구소 연구소장
관심분야: 생체전자, 생체신호처리, 의용 센서시스템, 완
전이식형 인공지능



김 명 남

1988년 2월 경북대학교 전자공학
과 (공학사)
1990년 2월 경북대학교 대학원
전자공학과 (공학석사)
1995년 2월 경북대학교 대학원
전자공학과 (공학박사)

1996년~현재 경북대학교 의학전문대학원 의공학교실
주임교수
관심분야: 생체신호처리시스템, 의학영상처리