

논문 2010-47SP-5-7

이어폰 주관 평가 기법 : 보정 필터 및 평가 결과

(Subjective Assessment Method for Earphone : Compensation Filter and Assessment Results)

전 상 배*, 성 광 모*

(Sang Bae Chon and Koeng-Mo Sung)

요 약

본 논문에서는 이어폰의 주관 평가 기법에 대해서 소개한다. 음질의 주관 평가는 동시비교를 위한 평가 자극간의 즉각적인 스위칭을 제공하도록 설계되어야 한다. 그럼에도 불구하고 이어폰의 음질 평가의 경우에 있어서는, 물리적으로 평가 자극간의 즉각적인 스위칭이 불가능하여 직접적인 주관 평가에 어려운 점이 있다. 이에, 본 논문에서는 각 이어폰으로부터의 자극을 녹음하여 녹음된 자극간의 동시 비교를 통하여 이어폰을 평가하는 방법을 제안하고, 이에 필요한 선처리들과 평가 결과에 대하여 소개한다.

Abstract

In this paper, a subjective assessment for sound quality evaluation of earphones is introduced. Although instantaneous switching among assessed stimuli is very important in the subjective assessment from the simultaneous comparison's point of view, it is physically impossible in the subjective assessment of earphone. Thus, this paper proposes an indirect assessment method using recorded stimuli and accompanied pre-processing and the assessment results are presented.

Keywords : Quality assessment, earphone

I. 서 론

이어폰은 휴대폰과 노트북 컴퓨터, MP3, PMP 등과 같은 다양한 모바일 기기에서 가장 널리 사용되는 트랜스듀서 기기이다. 그 결과, 이러한 휴대 기기 상에서의 음질은 이어폰의 성능에 의하여 좌우되는 경우가 많아 전체 음질의 중요한 부분을 차지하게 된다.

이어폰의 음질을 평가하는 방법에 있어서, 가장 첫 단계는 주관적으로 그 음질을 평가하는 데에 있다. 주관적인 평가가 이루어진 후, 그 결과에 기반하여 객관적인 인지 모델을 세워서 음질 예측 모델들 만들 수 있

기 때문이다.

본 논문에서는 이어폰 음질 평가의 첫걸음이 될 수 있도록 주관 평가 기법과 이에 필요한 신호처리 기술들에 대하여 소개하고자 한다.

II. 제안하는 주관 평가 기법

공정한 주관평가를 위해서는, 각각의 테스트 자극 (stimulus) 또는 레퍼런스 자극간의 즉각적인 스위칭이 매우 중요하다. Letowski가 [1]에서 정의한 바와 같이, 음질은 청취자의 기억속에 남아있는 레퍼런스의 음상 이미지와 비교하여 평가된다. 청취자의 기억속에 남아 있는 레퍼런스 자극에 대한 기억이 빨리 감소되기 때문에, 동시 비교(simultaneous comparison)은 매우 중요하

* 정회원, 서울대학교 전기컴퓨터공학부
(Seoul National University)

접수일자: 2010년7월5일, 수정완료일: 2010년8월9일

게 된다. 따라서, ITU-R Recommendation BS.1116-1^[2]과 ITU-R Recommendation BS.1534-1^[3]과 같은 표준에서는 주관평가에 있어서 40 ms이하에서 자극간의 즉각적인 스위칭 (instantaneous switching)을 권고하고 있다.^[2]

그러나, 이어폰과 같이 청각기관에 직접적으로 밀착되는 트랜스듀서 시스템에 대한 주관 평가에 있어서는, 이러한 동시 스위칭이 물리적으로 불가능하게 된다. 따라서 각각의 이어폰을 대신하는 물리적 자극을 생성시켜, 자극들에 대한 동시 비교를 제공하여 결과적으로 보다 정확한 주관적 음질 평가를 수행하는 간접적인 방법을 사용하여야 한다.

동시 비교를 제공하며 가장 널리 사용되는 주관 평가 기법인 MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor)^[3] 기법은 intermediate quality audio coding system을 위하여 정의되어 있으나, 이어폰에 대한 음질의 열화를 오디오 코덱에 의한 음질의 열화로 볼 수 있다는 가정 하에서, 본 논문에서는 이어폰의 주관 음질 평가 기법의 기본 골격으로 사용한다. 이 경우, 정확한 이어폰의 음질 평가를 위해서는 이어폰의 특성을 정확히 묘사하는 자극을 어떻게 합성할 것인지에 대한 고찰이 필요하게 된다.

다만, MUSHRA가 일반적으로 3.5 kHz에서 low pass filtering된 신호를 anchor 신호로서 사용하는 반면^[3], 제안하는 이어폰의 음질 주관 평가 기법에서는 사용하지 않았다. 그 이유는, 일반적인 오디오 코딩에서의 비교되는 자극들은 anchor 신호에 비해서 상대적으로 왜곡이 적은 신호들이나, 이어폰에 의한 자극들은

anchor 신호보다 음질의 열화가 더 클 가능성이 높기 때문이다.

결과적으로 제안하는 이어폰의 주관 음질 평가 기법은 그림 1과 같이 세 블록으로 구성된다. 음원 녹음부와 선처리부는 청취평가에 사용할 자극을 인지적인 관점에서 이어폰의 특성을 잘 묘사할 수 있도록 설계되었다. 녹음과 선처리를 거친 자극들에 대하여 ITU-R Recommendation BS.708^[4]의 표준을 따르는 스튜디오 모니터 헤드폰을 사용하여 hidden reference와 함께 다중 자극 비교 (multiple stimuli comparison)의 형태로 청취평가를 수행한다.

가. 녹음부

녹음부에서는 이어폰을 ITU-T Recommendation P.57 Type 3.3^[5]에 정의되어있는 가상 귀 모델 (artificial ear with the pinna simulator)에 착용시킨 후, 이어폰을 통하여 재생되는 소리를 녹음한다. 이는 그림 2에서 나타나는 것과 같이 i 번째 테스트 음원 $x_i(n)$ 이 j

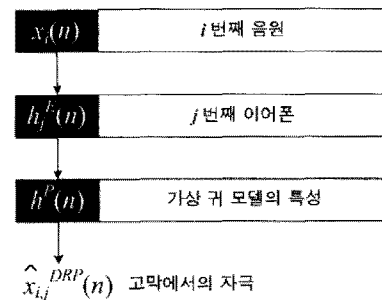


그림 2. 음원 녹음부
Fig. 2. Recording Session.

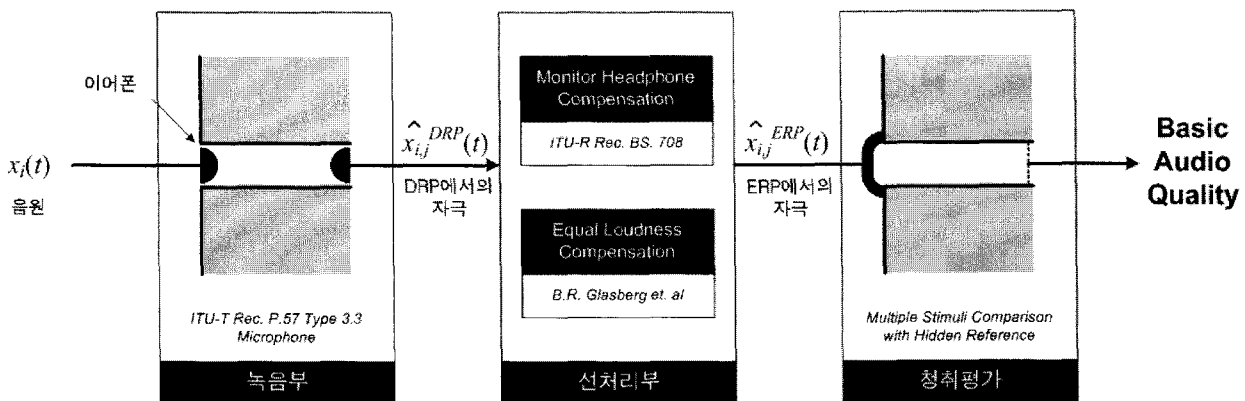


그림 1. 제안하는 주관 평가 기법
Fig. 1. Proposed Subjective Assessment Method.

번째 이어폰 $h_j^E(n)$ 을 통하여 $h^P(n)$ 의 특성의 가상 귀 모델을 통하여 녹음된 $\hat{x}_{i,j}^{DRP}(n)$ 의 자극을 얻는 것을 의미한다. 여기에서 녹음된 자극은 가상 귀 모델의 마이크 위치가 고막의 위치 (DRP : ear-Drum Reference Point)가 되고 이는 다음 식 (1)과 같이 나타낼 수 있다.

$$\hat{x}_{i,j}^{DRP}(n) = h^P(n) * h_j^E(n) * x_i(n) \quad (1)$$

나. 선처리부 (Pre-processing session)

녹음부의 결과로 얻어진 $\hat{x}_{i,j}^{DRP}(n)$ 의 자극이 j번째 이어폰의 특성 $h_j^E(n)$ 을 나타내고는 있으나, 이를 바로 청취평가에 사용하는 것은 다음과 같은 두 가지 이유로 부적절하다. 첫째, 주관 평가에서의 청취 평가는 스튜디오 모니터 헤드폰을 사용하여 진행될 것이나, 이 경우에 헤드폰을 통하여 재생되는 소리는 외이 (outer ear)의 위치에 해당하는 Ear Reference Point (ERP)에서 재생될 것이므로 이러한 DRP와 ERP에 대한 위치 보정이 필요하게 된다. 둘째, 청취평가에서 사용할 스튜디오 모니터 헤드폰의 특성 역시 녹음된 자극에 영향을 끼치므로 이 역시 보정될 필요가 있다. 이러한 두 가지의 문제점들을 해결하기 위하여 주파수 응답을 보정하는 선처리가 필요하게 된다.

(1) 주파수 응답 보정

녹음된 자극인 $\hat{x}_{i,j}^{DRP}(n)$ 를 주파수 보정 없이 직접 청취평가에 사용할 경우, 실제 청취자의 고막에 가해지는 자극은 식(2)와 같이 표현된다.

$$\begin{aligned} x_{i,j}^{DRP}(n) &= h^C(n) * h^H(n) * \hat{x}_{i,j}^{DRP}(n) \\ &= h^C(n) * h^H(n) * h^P(n) * h_j^E(n) * x_i(n) \end{aligned} \quad (2)$$

여기에서, $h^C(n)$ 은 DRP와 ERP의 차이에 해당하는 ear canal의 전달함수이고, $h^H(n)$ 은 사용하는 스튜디오 모니터 헤드폰의 충격 응답이다. 반면, 청취자가 직접 이어폰을 착용하였을 때의 고막에서의 자극 $\hat{x}_{i,j}^{DRP}(n)$ 은 식 (3)과 같이 표현된다.

$$\hat{x}_{i,j}^{DRP}(n) = h^C(n) * h_j^E(n) * x_i(n) \quad (3)$$

이러한 ear canal과 헤드폰의 영향은 그림 3-(a)와 같이 표현될 수 있다. 여기에서, $h^P(n)$ 과 $h^H(n)$ 이 선형이라는 가정 하에 그림 3-(b)와 같이 순서를 변형하

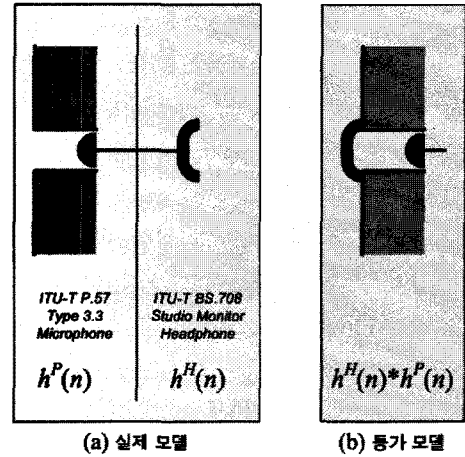


그림 3. Ear Canal과 헤드폰의 영향
Fig. 3. Model for Ear Canal and Headphone.

면, 이 변형된 등가 모델은 사인 스위프^[6]을 통하여 측정할 수 있게 된다. 즉, 스튜디오 모니터 헤드폰을 통하여 재생되는 사인 스위프를 가상 귀 모델을 통하여 녹음하여 등가 모델의 충격 응답 $h^M(n)$ 을 측정한다.

등가 모델의 충격 응답 $h^M(n)$ 이 측정되면, 인버스 필터인 $h^I(n)$ 을 Discrete Fourier Transform을 사용하여 식 (4), (5), (6)를 통하여 구해낸다.

$$H^M(k) = \sum_{n=0}^{N-1} h^M(n) W_N^{kn} \quad (4)$$

$$H^I(k) = \frac{1}{H^M(k)} \quad (5)$$

$$h^I(n) = \frac{1}{N} \sum_{k=0}^{N-1} H^I(k) W_N^{-kn} \quad (6)$$

녹음된 자극 $\hat{x}_{i,j}^{DRP}(n)$ 를 $h^I(n)$ 과 컨볼루션 시킴으로써, ear canal과 헤드폰의 영향이 상쇄된 ERP에서의 자극 $\hat{x}_{i,j}^{ERP}(n)$ 는 식 (7)과 같이 나타낼 수 있다.

$$\begin{aligned} \hat{x}_{i,j}^{ERP}(n) &= h^H(n) * h^I(n) * \hat{x}_{i,j}^{DRP}(n) \\ &= h^H(n) * h^I(n) * h^P(n) * h_j^E(n) * x_i(n) \\ &= h_j^E(n) * x_i(n) \end{aligned} \quad (7)$$

따라서, 인버스 필터링을 통하여 주파수 응답이 보정된 $\hat{x}_{i,j}^{ERP}(n)$ 를 사용할 경우에, 청취자의 DRP에서의 자극 $\hat{x}_{i,j}^{ERP}(n)$ 은 식 (8)과 같이 나타나게 되고, 결과적으로 청취자가 직접 이어폰을 착용하였을 경우인 식

(3)과 동일한 형태를 갖게 된다.

$$\tilde{x}_{i,j}^{DRP}(n) = h^C(n) * h_j^E(n) * x_i(n) \quad (8)$$

(2) 라우드니스 보정

라우드니스가 증가함에 따라 청취자의 선호도가 증가하는 현상을 방지하기 위하여, 청취 평가 이전에 모든 자극들은 그 라우드니스가 보정되어야 한다.

다. 청취평가

청취평가의 자극들을 녹음과 선처리를 통하여 얻은 후, ITU-T Recommendation BS. 708을 따르는 스튜디오 모니터 헤드폰을 통하여 청취평가를 진행한다. 여기에서 사용하는 헤드폰은 인버스 필터를 구하는 과정에서 사용한 것과 동일한 것을 사용한다.

청취평가에서 레퍼런스 자극으로는 녹음과 인버스필터를 거치지 않은 본래의 음원을 그대로 사용한다. 청취평가는 ITU-R Recommendation BS.1534-1의 MUSHRA 방식을 모두 따르나, 예외적으로 anchor 신호는 사용하지 않는 MUSHR (Multiple Stimulus with Hidden Reference)의 형태로 진행된다.

청취평가에 청취자들은 선호도를 제외한 인지되는 모든 왜곡을 바탕으로 한 Basic Audio Quality (BAQ)를 0~100점의 범위 안에서 부여하게 되고, 따라서 이어폰의 미학적, 예술적 측면은 고려대상에서 제외된다.

일반적인 관점에서 음질은 음색적인 측면과 공간감적인 측면에서 평가되나, 본 청취평가에서는 음색적인 측면에 대해서만 평가된다.

III. 주관 평가 실험

가. 실험 개요

제안하는 청취평가 기법을 통하여, 7개의 다른 회사로부터 상용화되어 널리 쓰이는 8개의 다른 이어폰의 음질이 평가되었다. ITU-T Recommendation P.57 Type 3.3의 표준을 따르는 가상 귀 모델로는 B&K 4128C가 자극 녹음에서 사용되었고, ITU-R Recommendation BS.708을 따르는 스튜디오 모니터 헤드폰으로는 Sennheiser HD 650이 청취평가에서 사용되었다.

각각의 이어폰이 이와 대응되는 녹음과 선처리를 거친 자극들과 얼마나 인지적인 관점에서 동일한가를 검

표 1. 주관평가에 사용된 음원

Table 1. Excerpts for Subjective Assessment.

번호	장르	설명 (시간(초))
1	재즈	연주곡(7.3)
2	팝	강한 연주와 여성 보컬 (10.4)
3	재즈	조용한 연주와 여성 보컬 (12.3)
4	R&B	일반적인 연주와 남성 보컬 (13.0)
5	클래식	현악 4중주 (12.3)

증하기 위하여, Sennheiser HD 650에 대한 자극도 녹음, 선처리되어 같이 평가되었다. 따라서, 하나의 음원에 대하여 전체 10개 (8개의 이어폰, 1개의 헤드폰, 하나의 레퍼런스)의 자극들이 비교되었다.

44.1 kHz로 샘플링된 표 1과 같은 다섯 개의 신호가 음원으로 사용되었고, 따라서 레퍼런스를 포함한 전체 10개의 자극 × 5개의 음원 = 50개의 자극이 비교되었다.

나. 선처리부

그림 3-(b)와 같은 등가 모델은 44.1 kHz로 샘플링된 15초 길이의 로그 사인 스위프를 사용하여 측정되었다^[6]. 측정된 15초 구간의 사인 스위프 중, 유효한 구간으로 판단되는 4096 샘플을 측정된 충격 응답 $h^M(n)$ 으로 정의하였다.

식 (4)~(6)에서와 같이 등가 주파수 응답 $H^M(k)$ 로부터 인버스 FIR 필터 $h^I(n)$ 를 구한 후, 100 Hz 이하의 매우 낮은 주파수 응답과 10kHz 이상의 고주파에 대한 보정은 생략하였다. 그 이유로는, 해당 주파수대의 보정 정도가 수십 dB로서, 측정 자체에서 발생하는 오

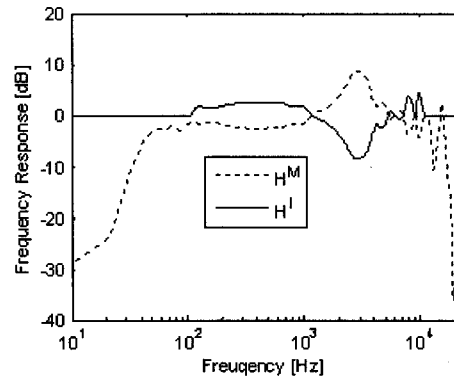


그림 4. 측정된 $H^M(k)$ 와 인버스 필터 $H^I(k)$

Fig. 4. Measured $H^M(k)$ and Inverse Filter $H^I(k)$.

표 2. 평가 자극들의 라우드니스
Table 2. Loudness of Test Excerpts.

번호	라우드니스 (phon)
1	84
2	93
3	87
4	90
5	85

차를 매우 크게 확대할 가능성이 있고, 나아가서 실제 시간 영역 신호에서 clipping과 같은 문제점이 발생할 수 있기 때문이다. 결과적으로 그림 4와 같이 100 Hz ~ 10 kHz 영역만 보정한 필터 $h'(n)$ 을 사용하였다.

주파수 보정이 이루어진 후, 라우드니스 보정을 통하여 청취평가에 사용될 자극들을 인지적인 측면에서 동일한 크기로 보정하였다. 라우드니스 측정으로는 Glasberg의 모델^[7]이 사용되었고, 모든 자극은 ± 2 phon의 오차 내에서 다음 표 2와 같은 라우드니스를 갖는다.

다. 청취 평가

정상적인 청취 능력을 갖는 10명의 남성과 1명의 여성이 청취평가에 참여하였다. 모든 청취자들은 기존에 ITU-R Recommendation BS.1116-1이나 BS1534-1과 같은 청취평가에 경험이 있었다.

라. 청취 평가 결과

그림 5는 청취평가의 결과를 그 평균값과 95퍼센트 신뢰구간으로 나타낸 것이다. 여러 이어폰들이 신뢰구간 상에서 겹침에도 불구하고, 결과적으로 다음과 같은 4개의 그룹을 형성함을 알 수 있다.

- (a) 헤드폰과 레퍼런스
- (b) 이어폰 3, 7, 8번
- (c) 이어폰 1, 4, 5, 6번
- (d) 이어폰 2번

그룹 (a)에서 나타나는 바에서 알 수 있듯, 헤드폰에 대한 자극은 레퍼런스와 큰 차이를 보이지 않았고, 이는 인지적으로 매우 유사함을 나타낸다. 이는 주파수 보정 필터 $h'(n)$ 이 잘 설계되었음을 보여준다.

2번 이어폰에 대한 자극들의 경우, Mean Opinion Score (MOS)가 20~30점 근처의 적은 신뢰구간을 보이고 있는데, 이는 청취자가 anchor 신호를 포함하는 MUSHRA에 익숙해져 있어 자극들 중 하나를 anchor로 간주한 데에서 나온 결과로 보인다. 또한 실제 2번 이어폰의 경우, 그 음질이 저음대역이 고음대역보다 강한 특성을 갖고 있어 더욱 anchor로 간주되었을 가능성이 크다고 판단된다.

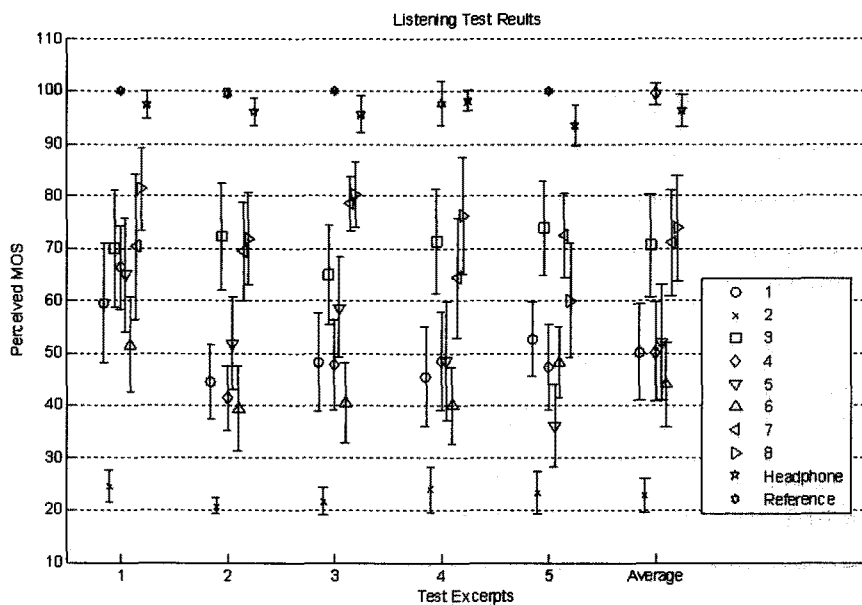


그림 5. 주관 평가 결과

Fig. 5. Subjective Assessment Results.

IV. 결 론

본 논문에서는 이어폰의 음질에 대한 주관 평가 기법이 소개되었다. 보다 정확한 음질의 평가를 위하여 반드시 필요한 동시 비교가, 자극간 즉각적인 스위칭을 제공하는 녹음, 선처리, 청취평가의 간접적인 방법을 통하여 가능하였다. 또한, 선처리에서의 주파수 응답 보정 필터의 성능은 청취평가를 통하여 확인되었다. 본 논문을 바탕으로 청취 평가의 결과를 예측하는 객관 평가 모델에 대한 연구를 진행 중에 있다.

참 고 문 헌

- [1] T. Letowski, "Sound Quality Assessment: Cardinal Concepts," presented at the 87th Convention of the Audio Engineering Society, *Journal of Engineering Society (Abstracts)*, vol. 37, pp.1062, 1989.
- [2] ITU-R Recommendation BS.1116-1, "Method for Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound System," *International Telecommunications Union*, Geneva, Switzerland, 1994.
- [3] ITU-R Recommendation BS.1534-1, "Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems," *International Telecommunication Union*, Geneva, Switzerland, 2001.
- [4] ITU-R Recommendation BS. 708, "Determination of the Electro-acoustical Properties of Studio Monitor Headphones," *International Telecommunication Union*, Geneva, Switzerland, 1990.
- [5] ITU-T Recommendation P.57, "Objective measuring apparatus : Artificial ears," *International Telecommunication Union*, Geneva, Switzerland, 2009.
- [6] S. Muller and P. Massarani, "Transfer-function measurement with sweeps," *Journal of Audio Engineering Society*, vol. 49, pp.443~471, 2001.
- [7] B. R. Glasberg and B. C. J. Moore, "A Model of Loudness Applicable to Time-varying Sounds," *Journal of Audio Engineering Society*, v. 40, pp 331~342, 2002.

저 자 소 개



전 상 배(정회원)

2000년 전북대학교 전기전자 제어공학부 학사졸업
 2005년 서울대학교 전기컴퓨터 공학부 석사졸업
 2009년 서울대학교 전기컴퓨터 공학부 박사졸업

<주관심분야 : 음향학, 음향 및 음성 신호처리>



성 경 모(정회원)

1973년 독일 아헨대학교 전기공학 학사졸업.
 1977년 독일 아헨대학교 전기통신공학 석사졸업.
 1982년 독일 아헨대학교 음향공학 박사졸업.

<주관심분야 : 음향학, 음향 및 음성 신호처리>