

머리 전달 함수의 보간에 적합한 왜곡 척도

A Relevant Distortion Criterion for Interpolation of the Head-Related Transfer Functions

이 기 승*, 이 석 필**
(Ki-Seung Lee*, and Seok-Pil Lee**)

*건국대학교 정보통신대학 전자공학부, **전자부품연구원 방송통신융합 연구센터
(접수일자: 2008년 10월 1일; 수정일자: 2008년 12월 3일; 채택일자: 2008년 12월 29일)

양이 재생 환경에서 다양하고 정확한 공간 이미지 형성을 위해서는 다양하고 세분화된 머리 전달 함수가 요구된다. 방대한 양의 머리 전달 함수를 효과적으로 감축하기 위한 방법으로, 모든 방향에 대한 머리 전달 함수를 몇 개의 대표 값들을 이용해 보간을 통해 얻는 방법이 주로 사용되고 있다. 본 논문에서는 머리 전달 함수의 보간 시 중요한 역할을 하는 보간 왜곡의 측정 방법에 대해 연구하였다. 다양한 객관적 왜곡 측정 방법을 이용하여 보간된 머리전달 함수와 본래의 머리전달 함수 간 차이를 표현하였으며 차이 값과 청취 테스트의 결과를 비교, 분석하였다. 분석 결과로부터 음원의 공간 이미지 차이를 가장 잘 반영하는 객관적 왜곡 측정 방법을 선택하였으며 이를 실제 보간 기법에 적용하였다. 3명의 사람으로부터 측정된 머리전달함수와 1개의 마네킹에서 측정된 머리전달함수에 제안된 방법을 적용한 결과, 3명의 사람에 대한 머리전달 함수에 대해서는 멜-주파수 cepstral 왜곡이, 4 종류의 머리전달 함수에 대해서는 시간 영역의 신호 대 왜곡비가 음원의 공간 이미지 차이를 가장 잘 예측하는 왜곡 척도임을 알 수 있었다.

핵심용어: 머리전달함수, 보간, 왜곡

투고분야: 전기음향 분야 (3.1)

In the binaural synthesis environments, wide varieties of the head-related transfer functions (HRTFs) that have measured with a various direction would be desirable to obtain the accurate and various spatial sound images. To reduce the size of HRTFs, interpolation has been often employed, where the HRTF for any direction is obtained by a limited number of the representative HRTFs. In this paper, we study on the distortion measures for interpolation, which has an important role in interpolation. With the various objective distortion metrics, the differences between the interpolated and the measured HRTFs were computed. These were then compared and analyzed with the results from the listening tests. From the results, the objective distortion measures were selected, that reflected the perceptual differences in spatial sound image. This measure was employed in a practical interpolation technique. We applied the proposed method to four kinds of an HRTF set, measured from three human heads and one mannequin. As a result, the Mel-frequency cepstral distortion was shown to be a good predictor for the differences in spatial sound location, when three HRTF measured from human, and the time-domain signal to distortion ratio revealed good prediction results for the entire four HRTF sets.

Keywords: Head related transfer function, Interpolation, Distortion measure

ASK subject classification: Electro-Acoustics (3.1)

I. 서론

다양한 신호처리 기술의 진보에 따라 오디오 신호 처리에 있어서도 단순한 재생이 아닌 재생음의 입체감을 증가

시키려는 노력이 진행되고 있다. 입체 음향을 구현하는 방법으로 서라운드 스피커 등을 이용하는 다채널 재생 방법과 헤드폰을 사용하는 양이 재생 방법 (binaural synthesis) [1] 등을 들 수 있다. 이 중 양이 재생 방법은 복수개의 스피커를 설치하기 어려운 환경이나 이동 가능한 소형의 장치들, 예로서 휴대폰이나 개인 멀티미디어 재생기 (Personal multimedia player, PMP) 등에 있어서 유용한

책임저자: 이 기 승 (kseung@konkuk.ac.kr)
143-701 서울특별시 광진구 화양동 1번지 143-701
건국대학교 정보통신대학 전자공학과 1417호
(전화: 02-450-3489; 팩스: 02-3437-5235)

입체 음향 기법으로 그 중요성이 점차 증대되고 있다 [1].

양이 재생 환경에서 입체 음향을 구현하는 방법으로, 좌, 우 채널에 대한 시간지연 (interaural time delay, ITD), 음량차 (interaural intensity difference, IID), 위상차 (interaural phase difference, IPD)를 음원의 위치에 따라 적절히 제어하는 방법을 들 수 있는데 [1], 보다 정확한 음상의 위치 제어를 위해서 머리전달함수 (head-related transfer function, HRTF)를 사용하는 기법이 사용되고 있다 [2-6]. HRTF는 자유 공간 상에서 음원과 귀간의 경로를 모델링하는 함수로써, 머리의 중심부를 기준으로 음원이 향한 방위각 (azimuth) 및 고도각 (elevation)에 따라 각기 다른 주파수 응답을 갖는 것으로 알려져 있다 [2]. 따라서 특정 방위각, 고도각에 대응하는 좌, 우 채널의 HRTF를 통과시킨 음원은 해당 방위각과 고도각에서 발생하는 소리로 인지된다. HRTF는 청취자의 머리 모양, 귀모양 등에 따라서도 그 특성이 변화하는데 [2-4], 이는 개인별로 맞춤화된 HRTF를 사용해야 정확한 음상이 형성됨을 의미한다. 또한 이동 음상 효과 (moving sound image effect) [6-7]를 얻기 위해서는 3차원 공간상의 모든 방향에서 측정된 HRTF가 필요한데, 이는 입체 음향의 구현 시 방대한 양의 HRTF가 필요함을 의미한다. 양이 재생 기법이 주로 소형의 장치들에 적용된다는 사실을 고려할 때, 매우 방대한 양의 HRTF를 저장하는 것은 다소 비현실적임을 알 수 있다.

이러한 문제에 따라, 모든 방향에 대한 HRTF를 저장하지 않고, 특정한 방향에 대한 일부 HRTF만을 저장하고 나머지 방향에 대해서는 저장된 대표 HRTF를 이용하여 HRTF를 구현하는 방법이 주로 이용되고 있다 [3][5-7][9-11]. 몇 개의 대표 HRTF로부터 모든 방향에 대한 HRTF를 얻기 위한 방법으로 부호화 기법 등에 널리 사용되고 있는 벡터 양자화 (vector quantization) 기법과 유사하게 3차원 공간상에 분포된 HRTF를 구획 분할 (cell partitioning) 하여, 특정 구획에 포함된 모든 HRTF를 해당 구획의 대표 HRTF로 근사화하는 방법이 있으며 [7][10-11], 원하는 위치에서 근접하여 위치하고 있는 대표 HRTF들을 이용하여 보간 (interpolation)을 수행함으로써 HRTF를 얻는 방법 등이 있다 [3][5-6][9]. 이러한 기법들은 모두 근사화된 HRTF와 실제 측정된 HRTF 간의 차이를 나타내는 왜곡 척도 (distortion measure)가 필요하다. HRTF에 대한 왜곡 척도에 대한 연구로서 Nicol 등은 스펙트럼 자승평균오차 (mean square error, MSE)를 포함하는 4종류의 거리 측정 방법을 분석하고 이에 대한 결과를 제시하였다 [7]. 그러나 이 연구에서는 근사화

HRTF의 왜곡이 청취상 음원 위치 변동에 끼치는 영향과 같은 주관적인 관점이 고려되지 않았으며, 단순히 방향의 차이에 따른 왜곡의 변화, 왜곡의 분산 등이 제시되었다. 본 논문에서는 다양한 거리 측정 방법으로 표현된 근사화된 HRTF와 실제 측정된 HRTF 간의 왜곡 척도가 주관적인 공간 이미지 변화에 어떠한 영향을 끼치는가를 정량적으로 분석하였으며, 이로부터 HRTF의 근사화에 가장 적합한 왜곡 척도를 찾는 방법을 제안하였다.

II. HRTF의 근사화를 위한 왜곡 측정 방법

몇 개의 대표 HRTF를 이용하여 모든 방향에 대한 HRTF를 얻는 HRTF의 근사화 기법에서는, 대표 HRTF를 선택하는 방법이 고안되어야 한다. 대표 HRTF의 선택은 대표 HRTF를 이용하여 근사화된 HRTF $\hat{H}(f)$ 와 실제로 측정된 HRTF $H(f)$ 간의 전체 왜곡 (overall distortion)이 최소화되는 관점에서 이루어진다.

두 변수간의 왜곡 측정은 부호화 기법, 음성 인식 등의 분야에서 주로 연구되었으며 선 스펙트럼 쌍 (line spectrum pair, LSP)의 유클리디언 거리, 켈프스트럼 (cepstrum)의 유클리디언 거리 등이 주로 사용되고 있다 [12]. 이러한 왜곡 척도는 사용되는 목적에 부합한 특성을 지니고 있는데, 부호화에서는 주관적으로 인지되는 잡음을 잘 반영하고 있는 왜곡 척도가 이용되고 있으며 인식에서는 인식하고자 하는 단어 사이의 차이를 명확하게 구분할 수 있는 왜곡 척도가 사용되고 있다. 그러나 HRTF의 근사화에서는 HRTF 자체를 청취자가 듣거나, HRTF를 이용하여 인식을 수행하는 것이 아니기 때문에 부호화와 인식에서 사용된 왜곡 척도와는 다른 측정 방법이 고안되어야 한다.

HRTF의 주요 역할이 음상의 위치 결정에 있다면, 이에 대한 왜곡 척도는 재생음의 음상 위치 변동 정도를 잘 반영하는 특성을 지니고 있어야 한다. 본 논문에서는 기존에 다른 분야에 적용되었던 왜곡 척도를 포함하여 다양한 척도에 대해 어떠한 척도가 HRTF의 근사화 목적에 가장 부합한지를 살펴보았다. 본 논문에서 고려된 왜곡 척도들은 다음과 같다.

2.1. 머리 전달 충격파 응답의 신호 대 왜곡비

HRTF는 주파수 도메인에서의 특징 변수로서, 이에 대응하는 시간 영역의 특징 변수를 머리 전달 충격파 응답 (head-related impulse response, HRIR)이라 한다. HRIR

을 이용한 왜곡 척도로서, 아래 식으로 주어지는 신호대 왜곡비 (signal to distortion ratio, SDR)를 사용할 수 있다.

$$C_{SDR} = 10 \log_{10} \left(\frac{\sum_{n=1}^N \{h(n) - \hat{h}(n)\}^2}{\sum_{n=1}^N h^2(n)} \right) \quad (1)$$

여기서 $h(n)$ 과 $\hat{h}(n)$ 은 각각 측정된 HRIR 과 근사화된 HRIR 의 n -번째 샘플을 나타내며 N 은 HRIR 의 길이를 나타낸다. SDR은 HRIR로부터 직접적으로 계산되므로, 다른 왜곡 척도에 비해 간단하게 계산될 수 있다는 장점이 있다. 또한 음상 위치 변화에 비교적 큰 영향을 끼치는 두 HRIR 간의 시간적 차이 (time lag)에 따라 그 변화 정도가 크게 나타나기 때문에 음상의 위치 변화 정도를 잘 반영한 것으로 예상된다. 이를 고려하면 두 HRIR 간의 시간 지연 차이와 시간 지연이 보상된 HRIR 간의 왜곡을 각각 분리하여 나타내는 방법을 고려할 수 있는데, 본 논문에서는 식 (1)과 같은 단순화된 왜곡 척도를 사용하였다. 이와 같은 왜곡 척도는 왜곡 측정이 시간 도메인에서 이루어지기 때문에 주파수 도메인에서 표현되는 귀의 특성 등을 반영할 수 없다는 단점이 있다.

2.2. Avedano 왜곡 척도

두 번째로 고려된 왜곡 척도는 Avedano 등에 의해 소개된 척도로서 [13], 데시벨 스케일로 표현된 주파수 왜곡을 나타낸다.

$$C_A = 10 \log_{10} \left\{ \frac{\sum_{m=1}^M (|H(m)| - |\hat{H}(m)|)^2}{\sum_{m=1}^M |H(m)|^2} + 1 \right\} \quad (2)$$

여기서 $|H(m)|$ 과 $|\hat{H}(m)|$ 은 각각 측정된 HRTF와 근사화된 HRTF의 m 번째 주파수 성분에 대한 크기 (magnitude)를 나타내며 M 은 주파수 성분의 전체 개수를 나타낸다. 일반적으로 $H(m)$ 은 HRIR $h(n)$ 의 M -point DFT (Discrete Fourier Transform)을 통해 얻어진다. C_A 는 앞서 살펴본 C_{SDR} 과 비교하여 각 주파수 성분의 크기만을 고려하고 위상 특성은 고려되지 않았기 때문에 두 HRTF 간의 위상 차이 또는 시간 지연 등이 내포되지 않은 왜곡 척도라 할 수 있다.

2.3. 가중 주파수 왜곡 척도

일반적으로 인간의 귀는 높은 주파수 보다 낮은 주파수의 왜곡에 대해 더 민감하게 반응하는 것으로 알려져 있다 [12]. 이러한 사실에 근거하여 주파수 밴드별로 각각 다른 가중치 (weight)를 도입하여 왜곡을 측정할 수 있다. 가중치를 도입한 주파수 왜곡 척도는 아래와 같다.

$$C_{FW} = \frac{1}{M} \sum_{m=1}^M \{w(m) (|H(m)| - |\hat{H}(m)|)^2\} \quad (3)$$

여기서 $w(m)$ 은 m 번째 주파수 성분에 대한 가중치를 나타낸다. 본 논문에서 가중치는 아래 식으로 주어지는 "Munich" 식 [14]을 사용하였다.

$$w(m) = 25 + 75(1 + 1.4(\frac{mf_s}{2M})^2)^{0.66} \quad (4)$$

f_s 는 샘플링 주파수를 나타낸다. C_{FW} 는 귀의 주파수 대역별 민감도를 내포하고 있으므로 인간의 귀에 실제적으로 인지되는 왜곡과 가까울 것으로 기대된다. 앞서 살펴본 Avedano 척도와 마찬가지로 HRTF 의 왜곡 측정에 중요한 역할을 담당할 것으로 예상되는 위상 차이를 반영할 수 없다는 단점이 있다.

2.4. 위상 왜곡 척도

음상의 위치는 각 주파수의 위상 응답, 즉 시간 지연에 영향을 받을 것이라는 가정에 근거하여, 아래와 같은 위상 왜곡 척도를 고려할 수 있다.

$$C_{PM} = \frac{1}{M} \sum_{m=1}^M \{H_c(m) - \hat{H}_c(m)\}^2 \quad (5)$$

여기서 $H_c(m)$ 과 $\hat{H}_c(m)$ 은 각각 m 번째 주파수 성분에 대한 측정된 HRTF와 근사화된 HRTF의 unwrapping 된 위상값을 나타낸다. 위상 왜곡은 스펙트럼의 크기 차이를 고려하지 않고 단순히 주파수 성분의 위상 차이, 즉 시간 지연만을 고려한 왜곡 척도이다. HRTF의 위상 스펙트럼에 대한 사람 귀의 반응 특성은 Kulkarni의 연구 [8]에서 고찰되었는데, 낮은 주파수에 대한 양쪽 채널간의 ITD가 귀에 민감하게 반응하는 파라미터임을 보고하였다. 이러한 결과를 바탕으로 C_{PM} 의 계산 시에도 낮은 주파수의 성분만이 고려될 수 있으나, 본 논문에서는 식 (5)와 같은 전 주파수 대역에서 계산된 위상 왜곡을 사용하였다.

2.5. 멜 밴드 스펙트럼 왜곡 척도

인간의 귀는 멜 (Mel) 스케일로 소리를 인지한다고 알려져 있다 [12][14]. 주파수 f 에 대한 멜 스케일은 다음과 같다.

$$Mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (6)$$

이와 같은 사실에 근거하여 HRTF의 왜곡도 멜 밴드 단위로 계산될 수 있다. 멜 밴드 스펙트럼 왜곡 척도는 HRTF의 크기 스펙트럼에 대해 각 멜 밴드별 대역 통과 필터를 통과시키고, 여기서 얻어진 대역별 신호의 에너지를 구하고 대역별 에너지에 대한 자승 오차 합을 구하여 얻는다. 즉,

$$C_{MWD} = \frac{1}{N_B} \sum_{k=1}^{N_B} \{ Y_k - \hat{Y}_k \}^2 \quad (7)$$

여기서 N_B 는 멜 밴드의 전체 개수로서, 본 논문에서는 MPEG Audio 압축 알고리즘에 사용된 24밴드가 사용되었다. 각 멜 밴드별 중심 주파수는 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500 Hz이다. Y_k 와 \hat{Y}_k 는 k -번째 멜 밴드에 대한 측정된 HRTF와 근사화된 HRTF의 에너지 값으로 다음과 같이 계산될 수 있다.

$$Y_k = \sum_{m=1}^M |H(m)| X_k(m), \quad \hat{Y}_k = \sum_{m=1}^M |\hat{H}(m)| X_k(m) \quad (8)$$

$X_k(m)$ 은 k -번째 멜 밴드에 대한 대역 통과 필터의 m -번째 주파수 성분에 대한 주파수 응답 (크기 응답)을 나타낸다.

2.6. 멜 캡스트럴 왜곡 척도

캡스트럼 변수는 주파수 도메인에서 표현된 신호의 변화도를 나타내는 변수로서 2.5 절에서 소개되었던 멜 주파수별 에너지 값에 이산 여현 변환 (discrete cosinc transform, DCT)를 수행함으로써 얻어진다 [12]. 즉, 멜 캡스트럴 왜곡 척도는 아래 식으로 주어진다.

$$C_{MCD} = \frac{1}{N_C} \sum_{k=0}^{N_C-1} \{ M_k - \hat{M}_k \}^2 \quad (9)$$

여기서 N_C 는 캡스트럼 계수의 개수를 나타내며, M_k 와 \hat{M}_k 는 각각 측정 HRTF와 근사 HRTF의 k -번째 캡스트럼 계수를 나타낸다. 캡스트럼 계수는 아래 식으로 주어진다.

$$M_k = \sum_{i=1}^{N_B} Y_i \cos \left[\frac{\pi}{N_B} \left(i + \frac{1}{2} \right) k \right], \quad k = 0, \dots, N_C - 1 \quad (10)$$

캡스트럼 변수는 음성 인식, 화자 인식과 같은 분류 (classification) 목적의 응용 분야에 널리 사용되는 변수로서 [12], 멜 주파수별 에너지값 보다 낮은 차원 (dimension)의 변수로 나타낼 수 있다. 본 논문에서 캡스트럼 계수 N_C 는 16이다. 캡스트럼 왜곡 척도는 두 HRTF 간에 스펙트럼의 미세한 차이보다는 포락선의 차이와 같은 전반적인 차이를 고려하는 왜곡 척도라 할 수 있다. 그러나 멜 밴드 스펙트럼 왜곡 척도와 캡스트럴 왜곡 척도는 모두 크기 스펙트럼만을 고려한 변수로서, 위상 왜곡을 내포하지 못하고 있다.

III. 청취 테스트

앞서 살펴본 6 종류의 왜곡 척도에 대해 어떠한 척도가 인지되는 음원의 위치 변화에 유의한 연관성을 갖고 있는지 알아보기 위해 청취 테스트를 수행하였다. 청취 테스트 시 음원으로 HRTF와 관련된 연구에 비교적 널리 사용되고 있는 유색 잡음 (colored noise, -3 dB/oct)을 사용하였다. 음원은 1 초의 길이를 갖고며 신호가 완만하게 증가, 감소하도록 전, 후 100 msec 기간 동안 코사인 함수를 곱하였다.

음원은 HRTF의 샘플링 주파수와 동일하게 44.1 KHz로 샘플링 되었으며, 크기는 청취자가 편안함을 느낄 수 있는 레벨로 적절히 설정하였다. 청취 테스트에는 총 10명의 정상적인 청력을 가진 시험자가 참여 하였으며, 시험자의 연령대는 20대 5명, 30대 3명, 40대가 2명이었다.

사용된 HRTF는 MIT KEMAR 데이터 [16]와 3명의 성인 남자로부터 직접 측정된 데이터 (NJ1, NJ2, NJ3)가 사용되었다. 측정자는 장상의 머리 모양을 갖고 있었으며, 완전 무향 실 (4x3x1.5, WDI)에서 최대 길이 열 (maximum length sequence, MLS) 방법 [15]을 이용하였다. HRTF의 측정에는 양이 마이크로폰 type 4101 (B&K, Denmark)를 피실험자의 귀에 장착시키고 conditioning amplifier 2693-A (B&K, Denmark)로 증폭한 후 Fire face 400

digital audio interface (RME, Germany)를 이용하여 PC로 data를 전송하였다. 음원 발생을 위한 스피커 시스템으로 8020A (Genelec, Finland) 스피커 시스템이 사용되었으며 음원과 머리 중심부까지의 거리는 약 1.63 m였다. HRTF 측정을 위한 소프트웨어는 Purebits 사의 SamplePRO가 사용되었으며 사용된 마이크로폰, 스피커의 특성으로 인한 영향을 감소시키기 위해 취득된 샘플에 대해 마이크로폰, 스피커의 충격과 응답을 역-전분류선하였다. 각 HRTF 데이터는 0도의 고도각을 갖는 수평면에 대해 5도 간격의 방위각에서 측정된 총 72개의 HRTF가 사용되었다.

청취 테스트에 참여한 10명과 HRTF 측정에 참여한 3명은 서로 다른 사람으로서 청취 테스트에 참여한 모든 피실험자는 비개인화된 HRTF (non-individualized HRTF)를 사용하였다.

청취 테스트에는 쌍-비교 법 (paired comparison)이 사용되었으며 다음과 같이 진행되었다. 먼저 임의의 방위각에 대해 측정된 HRTF를 통과한 신호를 들려주고, 1초간의 휴식기를 거쳐 동일한 방위각을 갖는 근사화된 HRTF를 통과한 신호를 들려주었다. 두 신호를 듣고 난 후, 청취자는 두 신호가 동일한 방향에서 발생하는지 (0) 서로 다른 방향에서 발생하는지 (1)를 테스트지에 기입하도록 하였다. 방향 차이를 판단하기 힘든 경우, 무제한의 반복 청취가 가능하도록 하였다. 또한 모든 청취자는 테스트 전에 소리의 방향감에 익숙하기 위한 과정으로 임의로 선택된 10개의 방향에 대해 먼저 방향을 알려주고, 해당 방향에서 측정된 HRTF를 통과한 신호를 들려주었다. 테스트 시 음색 변화로 인한 두 음간의 차이를 방향감의 차이로 판단하는 것을 방지하기 위하여 모든 피실험자는 두 음간의 차이를 오로지 방향감에 의해서만 구분하도록 테스트 전에 주지 시켰다. 근사화된 HRTF를 얻기 위해 쌍선형 보간 기법 (bi-linear interpolation)을 HRIR에 대해 적용하였으며, 다양한 왜곡 값을 갖는 근사 HRTF를 얻기 위해 보간 시의 보간 간격 (interpolation interval)은 5도에서 45도로 가변하였다. 이러한 보간 간격에 따라 얻어지는 근사 HRTF의 총 개수는 576개로서, paired comparison 또한 576 번이 시행되었다. 청취 테스트 결과의 신뢰도를 높이기 위하여 576개의 조합 중에 임의로 100개를 선택하여 이에 대해서는 반복 테스트를 수행하였는데, 첫 번째 테스트 결과와 두 번째 테스트 결과가 95% 이상의 일치율을 나타낸 피실험자의 청취 테스트 결과만을 유효 자료로 사용하였다.

청취 테스트는 비교적 조용한 환경에서 개방형 헤드폰

(Beyerdynamic DT990 PRO, Germany)을 사용하였으며 오디오 신호 발생에는 HRTF의 측정 시에 사용되었던 Fireface 400이 사용되었다. 테스트 실의 크기는 6.7x3.8x2.5 (WDH)로서 평균 산향시간은 220 msec, 초기 감쇄 시간은 200 msec, 사용된 헤드폰의 고조파 왜곡은 0.3% (1 kHz) 신호대 잡음비는 23.1 dB였다. 사용된 헤드폰의 특성으로 인한 청취 테스트의 오류를 방지하기 위하여 사용된 헤드폰의 충격과 응답을 무향실에서 미리 측정하였으며, 이를 역분류선한 값을 실제 테스트에 사용하였다. 피실험자 개인의 헤드폰-귀 전달 특성 (headphone-to-ear canal transfer function)은 고려하지 않았다.

IV. 청취 테스트 결과를 통한 왜곡 척도의 평가

III장에 소개된 각 왜곡 척도가 청취상의 음원 위치 차이를 얼마나 잘 예측하는지 알아보기 위해 본 논문에서는 수용자 특성 곡선 (receiver operator characteristic curve, ROC-curve) [17]를 사용하였다. 예측 가능성을 평가하기 위해 주로 사용되는 상관 계수 (correlation coefficient)나 회귀 분석법 (regression analysis)은 두 변수가 선형적인 상관 관계를 갖는 경우 의미 있는 상관 분석이 가능하지만, 비선형적인 상관 관계에서는 그 역할이 불분명한 것으로 알려져 있다 [17]. HRTF의 왜곡 값과 청취자의 테스트 결과 값이 항상 선형 상관 관계를 갖는다는 보장이 없기 때문에 본 논문에서는 ROC 곡선을 이용하여 예측 가능성을 평가하였다.

평가 방법은 다음과 같다. 청취 테스트 결과와 왜곡 척도를 서로 연관시키기 위해 두 개의 조건 확률밀도 함수 $p(C|0)$ 와 $p(C|1)$ 을 먼저 정의한다. 각각은 음상의 위치가 서로 동일한 경우 (0)와 서로 다른 경우 (1)의 각 왜곡값 C 에 대한 확률을 나타낸다. 그림 1은 $p(C|0)$ 와 $p(C|1)$ 의 예로서, 음상의 위치가 서로 동일한 경우에는 작은 C 값에 큰 확률값이 분포되는 것을 알 수 있으며, 반대로 음상의 위치가 서로 다르다고 판정되는 경우에는 큰 값에 많이 분포함을 알 수 있다.

그림 1에서 x-축에 표시된 문턱치 값 β 에 대해 오 경보 (false alarm) 확률 P_f 와 적중 (hit) 확률 P_D 를 구하면 다음과 같다.

$$P_f(\beta) = \int_{\beta}^{\infty} p(C|0)dC, \quad P_D(\beta) = \int_{\beta}^{\infty} p(C|1)dC \quad (11)$$

즉, 왜곡이 문턱치 보다 큰 경우에 청감상 음원의 위치가 변화할 것이라고 가정하는 경우, 오 정보 확률은 왜곡이 문턱치보다 큰 경우에도 음원의 위치가 변동하지 않을 확률의 총합이며 적중 확률은 문턱치가 큰 경우에 청감상 음원의 위치가 변동할 확률의 총합이다. 문턱치 β 에 따라 P_f 와 P_D 를 각각 x-축과 y-축 값으로 도시함으로써 ROC-곡선을 얻을 수 있다. 그림 2는 ROC-곡선의 예를 나타낸다. 그림에서 점선으로 나타낸 대각선은 왜곡 정도의 주관적인 음원 차이 예측성이 전혀 없는 경우로써, 오 정보 확률과 적중 확률은 문턱치값과 무관하게 항상 0.5를 갖는 경우이다. 가장 예측성이 가장 높은 경우는 문턱치가 어떤 값을 넘게 되면 오 정보 확률은 0이 되고 적중 확률은 1.0이 되는 경우인데, 이는 ROC-곡선이 P_f 와 P_D 좌표 상에서 왼쪽 위로 향하는 경우이다. 그림 2에

나타난 A, B, C 3개의 ROC-곡선에 대해, 가장 왼쪽 위를 향하고 있는 A-곡선이 예측성이 높음을 나타내며 이에 따라 곡선 A에 해당하는 왜곡 척도가 가장 예측성이 높음을 알 수 있다. 반면 곡선 C에 해당하는 왜곡 척도가 가장 예측성이 낮다.

각 왜곡 척도에 대한 ROC-곡선을 얻는 과정은 다음과 같다. 먼저 특정한 왜곡 척도에 대한 왜곡값 C 를 테스트에 사용된 모든 근사 HRTF에 대해 구한다. 다음으로 각 근사 HRTF에 대한 주관적인 청취 결과로부터 조건 확률 밀도 함수 $p(C|0)$ 와 $p(C|1)$ 를 구한다. 조건 확률 밀도 함수를 이용하여 왜곡값의 문턱치에 따라 P_f 와 P_D 를 구한다. 이를 2차원 공간상에서 도시하면 해당 왜곡 척도에 대한 ROC-곡선이 얻어진다. 6종류의 왜곡 척도마다 각각의 ROC-곡선을 도시하였을 때, 왼쪽 위 방향으로 가장 치우쳐 있는 왜곡 척도가 가장 높은 예측성을 지닌 왜곡 척도라 할 수 있다. 다음 장에서 각 왜곡 척도에 대한 ROC-곡선을 제시하고 이로부터 각 왜곡 척도에 대한 예측성을 비교한다.

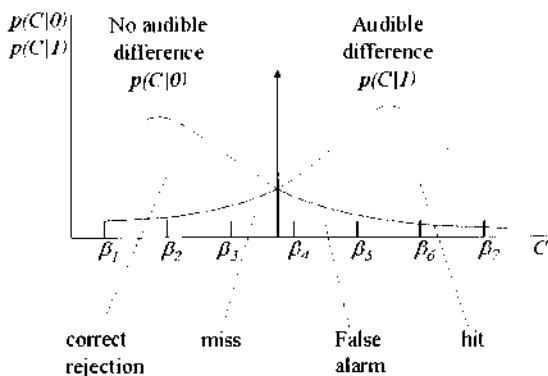


그림 1. 청취자의 음원 방향 차이 여부에 따른 조건 확률 분포함수의 예

Fig. 1. An example of the conditional probability density function according to the differences in subject's perceptual sound location.

V. 실험 결과

그림 3은 본 논문에서 사용된 6개의 왜곡 척도에 대한 ROC-곡선을 도시하고 있다. 각 그림에 대해 4종류의 HRTF에 대한 곡선이 함께 도시되어 있다. HRTF는 좌, 우 2채널로 주어지기 때문에 대표 왜곡을 나타내는 방법이고, 좌, 우 2채널에 대한 왜곡 중 큰 왜곡을 대표 왜곡값으로 선정하는 경우와 두 왜곡값의 평균값을 취하는 방법 2가지를 고려할 수 있다. 로그 단위 왜곡 척도인 C_{SDR} , C_1 는 그 평균값이 로그 영역에서의 평균값으로 나타낸다. 본 논문에서는 왜곡 측정에 앞서 모든 HRTF 데이터에 대해 측정 시 사용된 마이크폰, 스피커의 충격파 응답을 역 콘볼루션하여 측정 환경에 따른 HRTF 데이터의 편향 요인 (bias factor)를 제거하였다.

그림 3은 좌, 우 채널에 대한 왜곡값 중 큰 왜곡값을 대표 왜곡으로 선정하는 경우의 ROC-curve이다. 그림에서 보면 IIRIR의 신호대 왜곡 비 (C_{SDR}), Avedano 왜곡 척도 (C_1), 가중 주파수 왜곡 척도 (C_{FWD}), 멜 밴드 스펙트럼 왜곡 척도 (C_{MB})는 4개의 HRTF 데이터에 대해 대체적으로 일관된 ROC 특성을 보이는 반면, 위상 왜곡 (C_{PD}) 및 캡스트럼 왜곡 (C_{MD})는 그 데이터 별 편차가 크게 나타나고 있다. KFMR HRTF 데이터가 다른 HRTF 데이터에 비해 낮은 예측성을 나타내고 있다. 스펙트럼

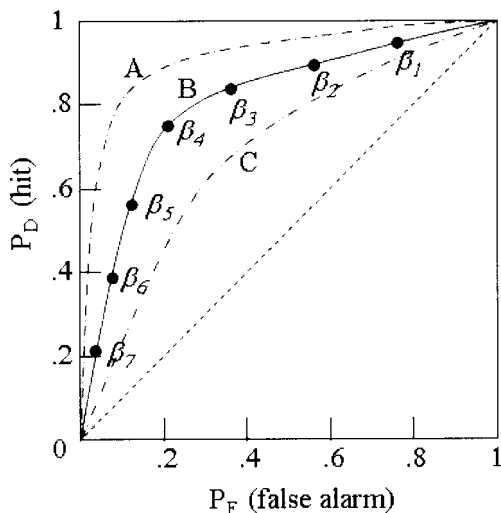


그림 2. 수용자 특성 곡선의 예

Fig. 2 An example of the receiver operator characteristic curve.

의 크기 응답에 대한 왜곡 척도인 C_A 와 위상 응답에 대한 왜곡 척도인 C_{FD} 를 비교하면, C_A 가 월등하게 좌상단에 분포함을 알 수 있는데, 이는 크기 응답의 왜곡이 청취상

의 음원 위치 변동에 더 유의한 영향을 끼침을 나타내는 것이다. 청취상 음원의 위치는 좌, 우 채널의 시간지연에 영향을 많이 받기 때문에 위상 응답에 민감할 것으로 예

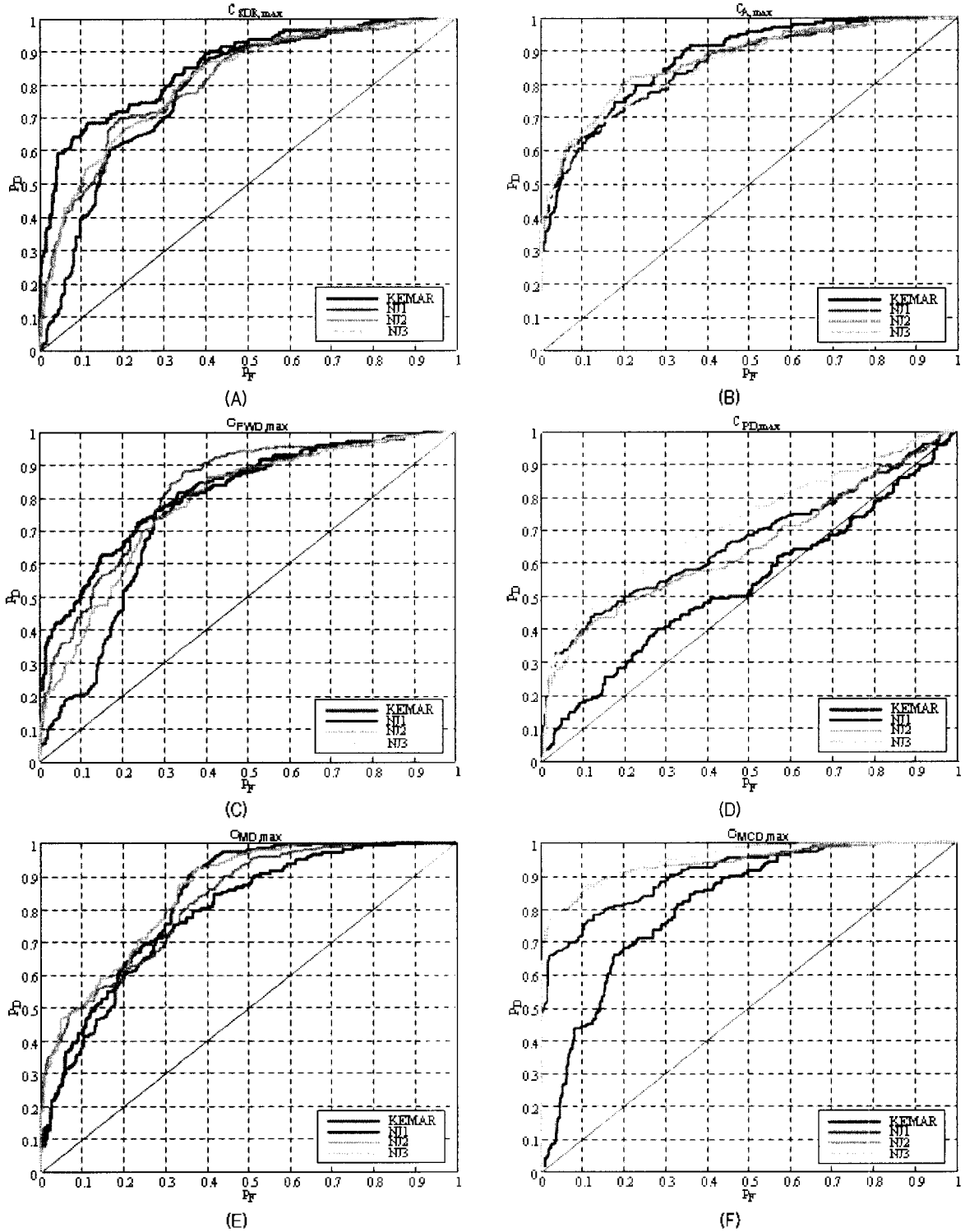


그림 3. 각 왜곡 척도 별 ROC 곡선 (채널 별 최대 왜곡값을 대표 왜곡척도로 간주하는 경우) (A) HRIR의 신호 대 왜곡비, (B) Avedano 왜곡 척도, (C) 가중 주파수 왜곡 척도, (D) 위상 왜곡 척도, (E) 멜 밴드 스펙트럼 왜곡 척도, (F) 멜 케스트럴 왜곡 척도
 Fig. 3. The ROC curves for each distortion criterion (In the case of taking the maximum distortion as the representative distortion) (A) Distortion in HRIR-SDR, (B) Avedano distortion, (C) Weighted distortion in frequency domain, (D) Phase distortion, (E) Mel-band spectrum distortion, (F) Mel-Frequency Cepstral distortion.

측되었으나, 실험 결과에서 보듯이 단순히 주파수 별 위상 차이만으로 왜곡을 표현하는 경우 크기 차이로 왜곡을 표현하는 경우와 비교하여 청취상 위치 변동을 잘 인지하지 못하는 것으로 나타났다. 시간 도메인에서 측정되는 왜곡 척도인 C_{SMM} 가 실제적으로는 위상 응답의 차이도 함께 반영하고 있는데 이에 대한 ROC 곡선과, 주파수 도메인에서 크기 응답만으로 측정되는 왜곡 척도인 C_A 의 ROC 곡선을 비교하면 C_{SMM} 이 약간 낮은 예측성을 나타내는 것을 알 수 있다. 이는 스펙트럼의 크기 차이만으로 표현되는 왜곡 척도가 위상의 차이를 포함한 왜곡 척도에 비해 높은 예측성을 나타낸다고 해석할 수 있다.

주파수별 사람의 귀와 유사한 민감도를 반영한 가중치를 반영하여 왜곡을 측정하는 C_{FUD} 는 가중치를 반영하지 않은 주파수 영역의 왜곡 척도인 C_A 와 비슷하거나 낮은 예측성 (NJ2 데이터의 경우)을 나타내었다. 이는 HRTF의 차이에 따른 방향성 인지가 주파수 별 크기응답에 대한 민감도를 반영함에 따라 예측 성능이 증진되지 않음을 의미한다. 이러한 결과는 인간의 귀가 갖는 또 다른 특성인 멜 밴드 주파수 해상도를 반영한 C_{MB} 가 C_A 와 유사한 ROC 특성을 나타낸 것과 유사한 결과로 볼 수 있다. 이는 본 논문에서 적용된 귀의 특성이 대부분 주파수의 크기 응답 (magnitude response)과 관련된 특성이며 [12] 위상 응답과는 상대적으로 덜 관련된 것에 그 원인이 있는 듯하다. 캡스트럼 계수를 사용한 C_{MD} 척도는 3개의 사람으로부터 얻은 HRTF 데이터에 대해서는 매우 우수한 예측 성능을 보이는 반면 KEMAR 데이터에 대해서는 C_A 보다 약간 낮은 예측성을 나타내고 있다. 이는 그림 3에서, NJ2, NJ3에 해당하는 HRTF의 ROC 곡선이 좌, 상단에 매우 근접해 있으며 NJ1 HRTF도 다른 왜곡 척도에 비해 좌 상단에 근접하여 분포하는 것으로 알 수 있다. 반면 KEMAR HRTF는 멜 캡스트럼 왜곡 척도가 C_{MD} 를 제외한 다른 왜곡 척도보다 대각선 방향에 더 근접하여 분포하며 예측성이 떨어짐을 나타낸다. 따라서 KEMAR HRTF 데이터를 논의의 대상에서 제외한다면 캡스트럼 왜곡 척도가 청취상의 위치 변동을 가장 잘 예측할 수 있는 왜곡 척도라 말할 수 있지만, KEMAR 데이터를 고려하면 항상 그러한 특성을 나타낸다고 보장할 수는 없다. 이는 HRTF가 시간 영역, 주파수 영역에서 표현되는 경우에 마네킨 (KEMAR)과 사람의 머리 (NJ1, NJ2, NJ3)에 따라 예측 성능 차이가 크게 나타나지 않는 반면 캡스트럼이 표현되는 큐퍼런스 도메인 (queferency domain) [12]에서는 그 차이가 크게 나타난 것으로 해석할 수 있다. 이러한 차이는 KEMAR HRTF의 측정에 사용된 스피커, 마이

크로폰이 NJ HRTF의 측정에 사용된 것과는 다른 특성을 갖는 것에 원인이 있을 수 있다. 그러나 청취 테스트 시스템, 마이크로폰의 충격파 응답을 역콘볼루션한 데이터를 사용하였기 때문에 측정 기기로 인한 영향은 미미하다고 볼 수 있다. 두 번째 원인으로서는 머리 모양의 차이 및 실험 환경의 물리적 차이를 생각할 수 있다. 실제로 KEMAR HRTF가 측정된 마네킨은 머리의 반경이 약 8 cm, 사람의 경우는 12.2 cm~9.1 cm였으며 머리의 중심부에서 음원까지의 거리는 KEMAR에서 1.45 m 사람의 경우는 1.63 m였다. 이러한 물리적인 차이는 음원에서 귀까지의 음파 도달 시간에 영향을 끼치는데, 두 데이터 간의 위상 응답에도 유의한 영향을 가져올 수 있다. 이는 그림 3에 제시한 위상 왜곡 응답의 ROC 특성이 사람의 HRTF와 KEMAR HRTF 간에 큰 차이가 나타난 것으로도 입증될 수 있다. 또한 머리의 세부적인 모양과 재질에 따른 경로 차이가 HRTF의 차이를 가져올 수 있는데 예로서 KEMAR의 경우에는 머리카락이 없는 플라스틱의 재질의 인조 머리를 통해 음파가 전달되며 사람의 경우는 머리카락, 피부 등을 통하여 음파가 전달된다. 이러한 차이는 캡스트럼 계수에 영향을 끼치며 결과적으로 C_{MD} 에 대한 ROC 특성 곡선이 서로 다르게 나타나는 원인으로 작용된 듯 하다.

그림 4는 대표 왜곡 척도로서, 좌-우 채널의 HRTF 왜곡의 평균값을 사용하는 경우의 ROC 곡선을 나타낸다. 좌-우 채널의 최대 왜곡값을 대표 왜곡으로 사용한 결과인 그림 3과 대체적으로 비슷한 특성을 나타냄을 알 수 있다. 결론적으로, 모든 HRTF 데이터에 대해, 청취상 음원의 위치 변동을 비교적 잘 예측할 수 있는 왜곡 척도로서 주파수 크기 응답에 대한 왜곡 척도였으며, KEMAR HRTF를 제외한 HRTF 데이터에 대해서는 멜 캡스트럼 왜곡 척도가 다른 왜곡 척도에 비해 월등하게 우수한 예측 특성을 나타내었다.

최종적으로, 각 왜곡 척도가 ROC 특성 곡선에 부합하는 예측 성능을 나타내는 가에 대한 검증 실험을 수행하였다. 먼저 청취 테스트 결과에 사용된 576개의 근사 HRTF에 대해 왜곡 별로 청취상 방향감의 차이가 느껴질 확률 (P_1)과 차이가 느껴지지 않는 확률 (P_0)을 구하고, 청취상 차이가 느껴지지 않는 왜곡의 문턱치, 즉 $P(0|C \leq C_{th}) = 1 - P_1$ 인 C_{th} 값을 6개의 왜곡 척도에 대해 구하였다. 각 왜곡 척도에 대해 왜곡값이 C_{th} 보다 작은 경우 확률적으로 청취상 방향감의 차이가 느껴지지 않는다고 볼 수 있다. 각 왜곡 척도에 대해 C_{th} 를 구하고, 청취 테스트에 포함되지 않는 100개의 근사 HRTF에 대해 별도의 청취

테스트를 수행하였다. 100개의 근사 HRITF에 대한 왜곡 값의 분포는 이전에 사용된 576개의 근사 HRITF에 대한 왜곡값과 유사한 분포를 갖는다. 청취 테스트 결과, ROC

곡선 상으로 유의한 상관 관계를 나타내었던 C_{SDR} , C_A 및 C_{MCD} 에 대하여 P_0 는 각각 0.97, 0.98, 0.95를 나타내었다. 이 값은 4개 HRITF에 대한 총 확률 값이다. 다른

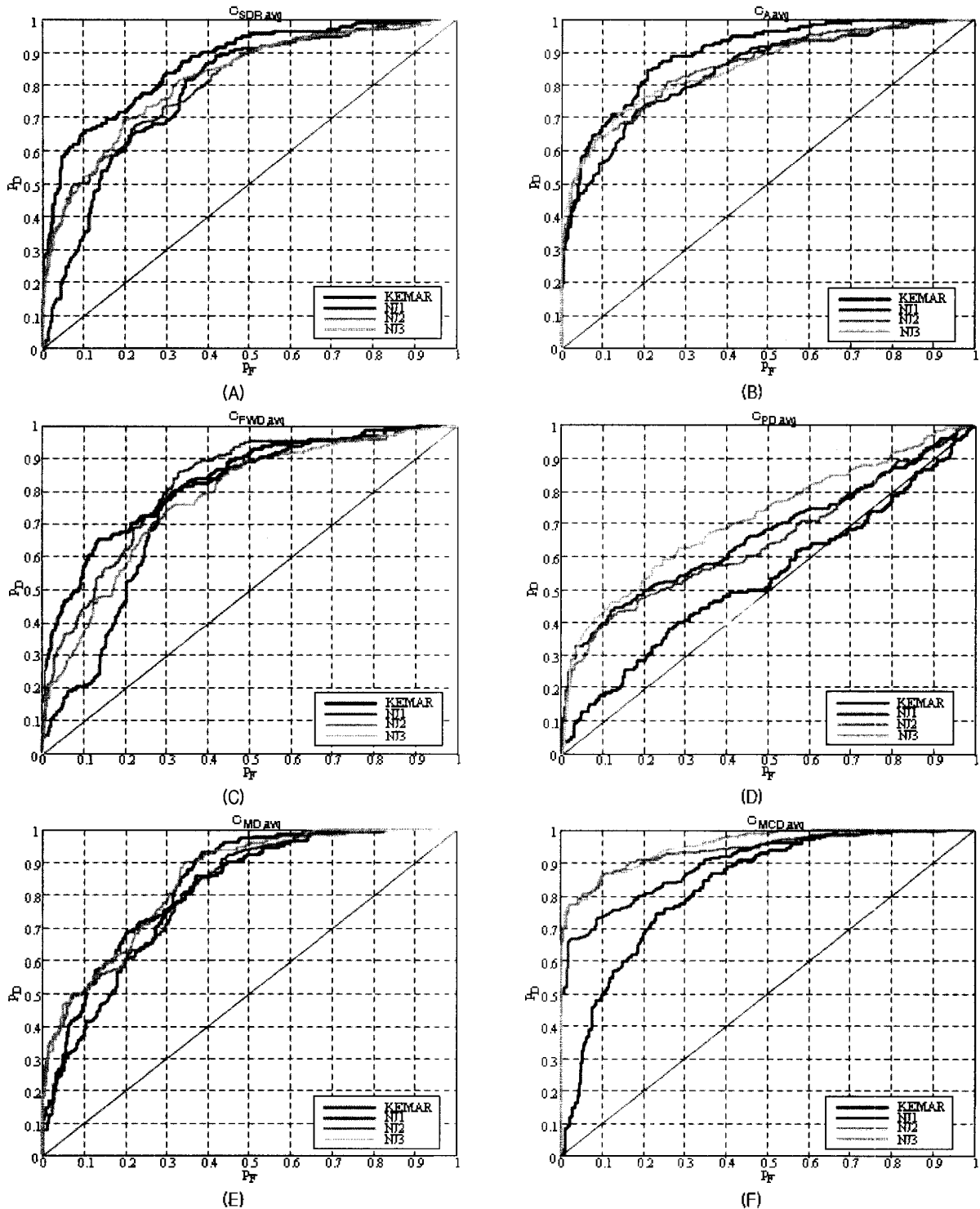


그림 4. 각 왜곡 척도 별 ROC 곡선 (채널간 평균 왜곡값을 대표 왜곡척도로 간주하는 경우) (A) HRIR의 신호 대 왜곡비, (B) Avedano 왜곡 척도, (C) 가중 주파수 왜곡 척도, (D) 위상 왜곡 척도, (E) 멜 밴드 스펙트럼 왜곡 척도, (F) 멜 캡스트럴 왜곡 척도
 Fig. 4. The ROC curves for each distortion criterion. (In the case of taking the averaged distortion as the representative distortion) (A) Distortion in HRIR-SDR, (B) Avedano distortion, (C) Weighted distortion in frequency domain, (D) Phase distortion, (E) Mel-band spectrum distortion, (F) Mel-Frequency Cepstral distortion.

왜곡 척도에 대한 I_0 는 이 보다 낮은 0.90~0.80 사이 값을 나타내었다. 이로부터, ROC 곡선 상으로 높은 예측성을 나타내었던 왜곡 척도가 청취상의 위치 차이를 비교적 잘 예측함을 알 수 있다.

VI. 결론

본 논문에서는 머리 전달 함수의 구획 분할, 보간 시 필요한 왜곡 척도로서, 청취상의 방향감 차이를 잘 예측할 수 있는 왜곡 척도의 선택 방법을 제안하였다. 왜곡 척도로서 시간 도메인에서 측정되는 머리전달 충격파 응답의 신호대 왜곡비, 주파수 도메인에서 표현되는 스펙트럼 크기의 왜곡 척도, 주파수 별 기중치가 고려된 왜곡 척도, 위상 왜곡 척도, 뱀-스케일로 표현되는 주파수 대역의 대역별 에너지 차이, 켈스트럼 왜곡 등이 고려되었다. 이와 같은 왜곡 척도와 청취상 음원 차이 간의 상관성 규명을 위해 근사 HRTF와 측정 HRTF의 조합을 통해 청취 테스트를 수행하였다. 사용자 응답 특성을 이용하여 왜곡 척도와 청취 테스트 결과를 연관시켰으며, 이에 대한 결과로 KEMAR HRTF를 제외한 3개 HRTF 데이터에 대해서는 뱀 켈스트럼 왜곡이 청취상 음원 위치 차이를 비교적 잘 예측하는 것으로 나타났으며, KEMAR HRTF를 포함한 4개의 HRTF 데이터에 대해서는 시간 도메인에서 측정된 신호대 왜곡비 그리고 주파수 도메인의 크기 응답에 대한 왜곡이 음원의 청취상 차이를 잘 예측할 수 있는 것으로 나타났다. 추 후 연구로는 본 논문에서와 같이 고도각을 일정하게 하고 방위각만을 변화시키면서 왜곡 척도의 예측성을 조사하는 것이 아니고, 고도각의 변화에 의해서도 예측성을 관찰하는 것이라 하겠다.

참고 문헌

1. J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1983.
2. D. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction", *J. Acoust. Soc. Am.*, Vol. 91, No. 3, pp. 1637-1647, 1992.
3. J. Breebaart and A. Kohlrausch, "The perceptual (ir) relevance of HRTF magnitude and phase spectra", *Audio Engineering Society Preprint, 110th Convention*, preprint no. 5406, 2001.
4. A. Kulkarni, S. K. Isabelle and H. S. Colburn, "On the minimum-phase approximation of head-related transfer functions", *Proceedings of IEEE ASSP workshop on applications of signal*

processing to audio and acoustics, pp. 84-87, 1995.

5. P. Minnaar and J. Plogsties and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis", *J. Audio Eng. Soc.*, Vol. 53, No. 10, pp. 919-929, 2005.
6. R. Tamura, Y. Hiraiwa, H. Hasegawa and M. Kasuga, "Research on the creation of a sound field using a 2-loudspeaker mobile phone", *Proceedings of IEEE region 10 conference TENCON 2004*, pp. 120-123, 2004.
7. F. Keyrouz and K. Diepold, "A new HRTF interpolation approach for fast synthesis of dynamic environmental interaction", *J. Audio Eng. Soc.*, Vol. 56, No. 1/2, pp. 28-35, 2008.
8. R. Nicol, V. Lemaire, A. Bondu and S. Busson, "Looking for a relevant similarity criterion for HRTF clustering: a comparative study", *Audio Engineering Society Preprint, 120th Convention*, preprint no. 6653, 2006.
9. A. Kulkarni, S. K. Isabelle and H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra", *J. Acoust. Soc. Am.*, Vol. 105, No. 5, pp. 2821-2840, 1999.
10. S. Shimada, N. Hayashi and S. Hayashi, "A clustering method for sound localization transfer functions", *J. Audio Eng. Soc.*, Vol. 42, No. 7/8, pp.577-584, 1994.
11. 구교식, 차형태, "스테레오 시스템을 위한 머리전달함수의 개선", *한국음향학회지 27권 4호* 207-214쪽, 2008.
12. J. R. Deller, J. G. Proakis and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, New York, 1993.
13. C. Avendano, R. O. Duda and V. R. Algazi, "Modeling the correlated HRTF", *AES 16th International Conference on Spatial Sound Reproduction*, 1999.
14. W. M. Hartmann, *Signal, sound, and Sensation*, Springer, 1998.
15. D. D. Rife and J. Vanderkooy, "Transfer-Function Measurements using Maximum-Length Sequences", *J. Audio Eng. Soc.*, Vol. 37, No. 6, pp. 419-444, 1989.
16. MIT media lab, *KEMAR HRTF data*, <http://sound.media.mit.edu/pub/Data/KEMAR>
17. R. Luce and C. Krumhansl, "Measurement, scaling and psychophysics", in *Handbook of Experimental Psychology*, S. Stevens, Ed, New York: Wiley, 1988.

저자 약력

•이 기 승 (Ki-Seung Lee)



1991년 2월: 연세대학교 전자공학과 (공학사)
 1993년 2월: 연세대학교 대학원 전자공학과 (공학석사)
 1997년 2월: 연세대학교 대학원 전자공학과 (공학박사)
 1997년 3월~1997년 8월: 연세대학교 신호처리 연구센터 선임 연구원
 1997년 10월~2000년 9월: AT&T Shannon Lab 연구원
 2000년 11월~2001년 8월: 삼성종합기술원 HCI Lab 전문연구원
 2001년 9월~2007년 8월: 건국대학교 정보통신대학 전자공학부 조교수
 2007년 9월~현재: 건국대학교 정보통신대학 전자공학부 부교수
 *주관심분야: 생체 신호 처리, 오디오 신호 처리 등.

•이 석 필 (Seok-Pil Lee)



1990년 2월: 연세대학교 전기공학과 (공학사)
1992년 8월: 연세대학교 대학원 전기공학과 (공학석사)
1997년 8월: 연세대학교 대학원 전기공학과 (공학박사)
1997년 7월 ~ 2002년 2월: 대우전자(주) 선임연구원
2002년 3월 ~ 현재: 전자부품연구원 방송통신융합 연구
센터장
※주관심분야: 비디오 신호 처리, 오디오 신호 처리