

공간감 인자로서의 고주파 대역 포락선 양이 시간차의 유효성

Saliency of Envelope Interaural Time Difference of High Frequency as Spatial Feature

서정훈*, 전상배*, 성평모*

(Jeong-Hun Seo*, Sang Bae Chon*, Koeng-Mo Sung*)

*서울대학교 전기컴퓨터공학부 & 서울대학교 뉴미디어 통신 공동 연구소

(접수일자: 2010년 4월 22일; 채택일자: 2010년 7월 5일)

다채널 오디오 코딩 시스템을 평가함에 있어서 음색 요소뿐만 아니라 공간감 요소 역시 큰 중요성을 갖는다. 이러한 이유로 양이 시간차 왜곡 (Interaural Time Difference Distortion, *ITDDist*), 양이 크기차 왜곡 (Interaural Level Difference Distortion, *ILDDist*), 양이 상관관계 왜곡 (Interaural Cross Correlation Distortion, *IACCDist*) 과 같은 공간감 요소를 추가하여 기존의 음질 객관 평가 권고안인 ITU-R Rec. BS, 1387-1 을 다채널 오디오 시스템에 적용하기 위한 시도가 있었다. 이 다채널 오디오 시스템을 위한 확장 모델에서는 Duplex 이론에 따라, 양이 시간차 왜곡은 1.5kHz 이하의 저주파 영역에 대해서만 계산되었으며, 양이 크기차 왜곡은 2.5kHz 이상의 고주파 영역에 대해서만 계산되었다. 하지만 고주파 영역에 있어서 포락선의 양이 시간차는 공간감 인자에 있어서, 특히 음상 정위에 있어서 중요한 역할을 한다. 이러한 고주파 영역 포락선의 양이 시간차 공간감 인자에 미치는 영향을 정량적으로 분석하기 위해 해당 인자를 계산하는 방법이 본 논문에서 소개된다. 또한, 이렇게 계산된 고주파 영역 포락선의 양이 시간차는 다채널 오디오의 주관평가 결과와 높은 상관관계를 가짐을 확인한다.

핵심어: 고주파 포락선, 양이 시간차, 다채널 오디오 코딩 시스템, 공간감 인자, 객관 평가

투고분야: 음악음향 및 음향심리 분야 (8.4)

Both timbral features and spatial features are important in the assessment of multichannel audio coding systems. The prediction model, extending the ITU-R Rec. BS, 1387-1 to multichannel audio coding systems, with the use of spatial features such as *ITDDist* (Interaural Time Difference Distortion), *ILDDist* (Interaural Level Difference Distortion), and *IACCDist* (InterAural Cross-correlation Coefficient Distortion) was proposed by Choi et al. In that model, *ITDDist*s were only computed for low frequency bands (below 1500Hz), and *ILDDist*s were computed only for high frequency bands (over 2500Hz) according to classical duplex theory. However, in the high frequency range, information in temporal envelope is also important in spatial perception, especially in sound localization. A new model to compute the ITD distortions of temporal envelopes in high frequency components is introduced in this paper to investigate the role of such ITD on spatial perception quantitatively. The computed ITD distortions of temporal envelopes in high frequency components were highly correlated with perceived sound quality of multichannel audio sounds.

Keywords: High Frequency Envelope, Interaural Time Difference, Multichannel Audio Coding System, Spatial Features, Objective Assessment

ASK subject classification: Musical Acoustics and Psychoacoustics (8.4)

I. 서론

객관음질 평가는 심리음향 분야의 중요한 응용분야 중

하나로, 이에 대한 많은 연구가 이루어지고 있다. 이러한 객관음질 평가는 모노와 스테레오 음원의 압축 기법에 대한 품질 평가에 널리 사용되고 있다. 하지만 최근 들어 다채널 시스템이 보편화되고 다채널 압축 기법이 발전함에 따라 다채널 오디오 신호에 대한 객관음질 평가에 대한 필요성이 대두되고 있다. 이에 따라 기존 ITU-R Rec.

BS. 1387-1의 음질 예측 모형을 다채널 신호로 확장하기 위한 연구가 진행되었다 [1]. 이 확장 모형은 ITU-R Rec. BS. 1387-1에서 사용된 10개의 음색 요소와 함께 양이 시간차 왜곡 (*ITDDist*), 양이 크기차 왜곡 (*ILDDist*), 양이 상관관계 왜곡 (*IACCDist*)과 같은 3개의 추가적인 공간감 요소를 사용하였다 [2]. 특히, [1]의 연구에서는 음상 정위에서의 오차를 예측하기 위한 요소로 양이 시간차 왜곡에 대한 연구가 이루어졌다. 이 예측 모델에서 양이 시간차 왜곡은 양이 위상차 (Interaural Phase Difference, IPD)가 뚜렷한 저주파 대역에 대해서만 계산되었다. 하지만, 최근 연구들에 따르면 저주파 대역의 양이 시간차뿐만 아니라 고주파 대역의 포락선에 대한 양이 시간차 역시 음상 정위에 중요하다는 결과가 밝혀지고 있다 [3-5].

본 논문에서는 고주파 대역의 포락선 양이 시간차 (Envelope Interaural Time Differences, EITDs)와 다채널 오디오 신호에 대한 주관적 청취 평가 결과의 상관도를 분석함으로써 다채널 오디오 신호의 객관 평가 인자로서의 고주파 포락선의 양이 시간차의 효용성을 검증한다.

본 논문의 구성은 다음과 같다. 2.1에서는 생리학과 심리음향학에서의 연구결과를 바탕으로 고주파 포락선 정보의 유용성에 대해 설명하며, 이를 바탕으로 2.2와 2.3에서 고주파 포락선의 양이 시간차 왜곡을 계산하는 방법에 대해 설명한다. 최종적으로 고주파 포락선의 양이 시간차 왜곡과 다채널 오디오 신호의 주관청취평가 결과의 상관관계를 분석하고 기존의 음색 인자 및 공간감 인자가 주관평가 결과와 갖는 상관계수를 비교함으로써, 고주파 포락선의 양이 시간차 왜곡이 다채널 오디오 신호에 대한 객관 품질 평가 모형의 한 인자로서 사용될 수 있음을 검증한다.

II. 고주파 포락선의 양이 시간차 계산

2.1. 고주파 포락선 정보의 유용성

일반적으로 인간이 저주파 음원과 고주파 음원의 위치를 인지할 때, 뇌에서는 서로 다른 처리과정을 거치게 된다. 저주파 음원의 위치 인지에 대해서는 양이 시간차 (Interaural Time Difference, ITD)를 사용하게 된다. 저주파 음원 자극에 의해 기저막 (basilar membrane)에서 발생하는 자극 패턴 (excitation pattern)은 중앙 상올리브핵 (Medial Superior Olive, MSO)으로 전달되고, 전달된 신호는 일치 검출 뉴런 (coincidence detection neurons)에 의해 처리되어 양이 시간차를 계산하고, 인간은 이를

이용해 음원의 위치를 인식하게 된다.

반면, 고주파 음원에 대해서는 기저막의 자극 패턴이 측면 상올리브핵 (Lateral Superior Olive, LSO)으로 전달되고, 이로 인해 양측 측면 상올리브핵에서는 다른 크기의 전기신호가 발생하게 되고, 발생된 전기신호의 차이에 의해 인간은 고주파 음원의 위치를 인식하게 된다. 하지만, 이러한 양이 크기차 외에, 고주파 음원의 포락선 정보 역시 고주파 음원의 음상 정위에 이용된다는 연구 결과들이 발표되고 있다 [3, 4]. 특히, Joris 외 다른 연구자들에 의한 연구 결과에 따르면, 측면 상올리브핵에 존재하는 뉴런은 고주파 전치 신호 (high frequency transposed tones)에 대한 민감도를 가진다. 또한, Bernstein과 Trahiotis의 연구결과에 따르면, 고주파 전치 신호 (high frequency transposed tones)에 대한 청신경 섬유 (Auditory Nerve Fiber, ANF)의 신경신호 격발 (neural firing) 확률은 저주파 음원에 대한 신경섬유의 격발 확률과 유사하다 [5]. 또한, Griffin 외 다른 연구자들에 의해 고주파 포락선의 양이 시간차에 대한 민감도는 저주파 음원의 양이 시간차에 대한 민감도와 유사함이 밝혀졌다 [6]. 이러한 연구 결과에 비추어 보았을 때, 고주파 영역 포락선의 양이 시간차는 저주파 영역의 양이 시간차와 고주파 영역의 양이 크기차와 함께 음상 정위에 많은 영향을 끼친다고 판단할 수 있다.

고주파 포락선의 양이 시간차가 뇌에서 어떻게 처리되는지 역시 또 다른 중요한 쟁점이다. [6]에서 밝혔듯이 고주파 신호 포락선의 양이 시간차에 대한 민감도를 매개하는 뇌에서의 처리 과정은 저주파 신호의 세부 구조에 기반한 양이 시간차에 대한 민감도를 매개하는 처리 과정과 유사하다. 특히, 고주파 전치 신호와 사인 진폭 변조된 신호에 대해 중뇌의 하구 (Inferior Colliculus, IC)의 뉴런에서 발생하는 위상 결속 (phase locking)은 유사한 패턴을 가진다. 중뇌의 하구에서 발생하는 위상 결속은 저주파 신호의 양이 시간차 발생에 대한 중요한 근거 중 하나이다. 이러한 위상 결속이 진폭 변조에 의한 고주파 신호의 포락선에 대해서도 유사하게 발생한다는 것은 고주파 음원의 위치 인지에 있어서 고주파 포락선의 양이 시간차가 이용됨을 시사하는 것이며, 또한 고주파 포락선의 양이 시간차 계산에 대한 근거를 제시한다고 해석할 수 있다. 저주파 신호에 대한 양이 시간차와 고주파 포락선에 대한 양이 시간차를 매개하는 중앙 처리 과정이 유사하다면, 고주파 영역의 음상 정위에 대한 양이 단서 (binaural cues)가 측면 상올리브핵 (Lateral Superior Olive, LSO)에서 추출된다 하더라도 고주파 포락선의 양이 시간차는

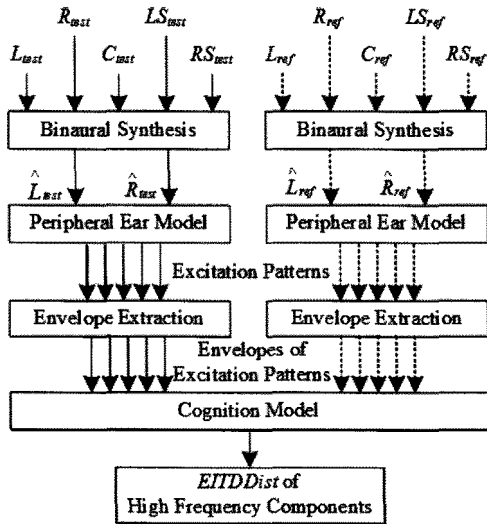


그림 1. 고주파 포락선의 양이 시간차왜곡의 계산
Fig. 1. Calculation of High Frequency Envelope ITD Distortion.

중앙 상올리브핵에서의 일치 검출 뉴런에 의한 양이 시간차와 같은 방법으로 계산될 수 있다. 이에 따라 고주파 포락선의 양이 시간차를 [1]에서 제안된 인지 모형에서 저주파 양이 시간차와 같은 방법으로 계산할 수 있다.

2.2. 고주파 포락선 추출

그림 1은 고주파 포락선의 양이 시간차 왜곡 계산 과정을 블록선도로 나타낸 것이다. 우선 머리전달함수 (Head Related Transfer Functions, HRTFs) 를 이용하여 다채널 음원을 양이 신호로 합성하는데, 이를 각각 \hat{L}_{test} , \hat{R}_{test} , \hat{L}_{ref} , \hat{R}_{ref} 로 표시하였다. 다채널 음원 및 양이 신호에서 아래첨자 *test*와 *ref*는 각각 평가 신호와 참조신호를 의미한다.

양이 신호 합성에 사용된 머리전달함수는 ITU-R Rec. BS. 1116-1 에 권고된 것과 같은 표준 환경 시청실에서 녹음된 것이며, LFE 채널은 모든 음원에 대해 0으로 조정되었다 [7]. 이 양이 신호 합성 과정은 식 (1)과 같이 표현할 수 있다.

식 (1)에서, H_{CL} , H_{LR} , H_{RR} , H_{Ld} , H_{Rd} , H_{CR} , H_{Lp} , H_{Rp} , H_{Lr} , H_{Rr} 은 각 스피커에서 양쪽 귀까지의 음파 전달 경로를 나타내는 10개의 양이 공간 전달함수 (Binaural Room Transfer Functions, BRTFs)이며, \hat{L} 과 \hat{R} 은 각각 양쪽 귀의 입력 신호를 의미한다.

$$\begin{pmatrix} \hat{L}_{test} \\ \hat{R}_{test} \end{pmatrix} = \begin{pmatrix} H_{Ld} & H_{Rp} & H_{CL} & H_{Ld} & H_{Rd} \\ H_{Lr} & H_{Rp} & H_{CR} & H_{Lr} & H_{Rr} \end{pmatrix} \begin{pmatrix} LF_{ref} & LF_{test} \\ RF_{ref} & RF_{test} \\ C_{ref} & C_{test} \\ LS_{ref} & LS_{test} \\ RS_{ref} & RS_{test} \end{pmatrix} \quad (1)$$

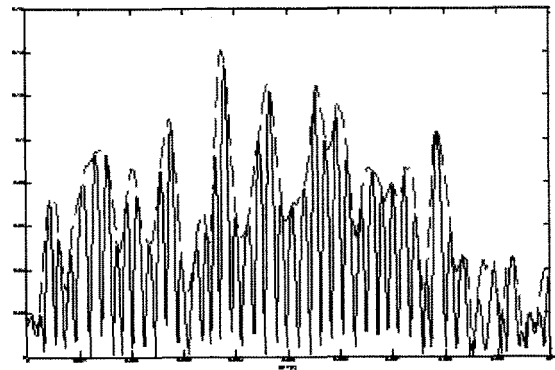


그림 2. 이산 힐버트 변환을 통한 포락선 추출. (실선: 정파 정류된 자극패턴, 점선: 추출된 포락선)
Fig. 2. Envelope extraction using discrete Hilbert transform (solid: full-rectified excitation pattern, dashed: extracted envelope).

이와 같은 방식으로 합성된 양이 신호는 말초 청각 모형 (peripheral ear model)에 의해 처리된다. 실제 양쪽 귀로 들어온 입력신호는 중이를 거쳐 달팽이관에서 처리되는데, 이 과정을 모사한 것이 말초 청각 모형이다. 말초 청각 모형 내의 달팽이관 모의장치 (cochlea simulator)는 양이 입력신호를 인간의 기저막 (basilar membrane)에서 유모세포 (hair cell)를 자극하는 신호로 변환한다. 달팽이관 모의장치는 ERB (Equivalent Rectangular Bandwidth) 스케일에 의해 결정된 중심주파수를 갖는 24개의 대역 통과 필터로 구성된 필터 뱅크로 간주할 수 있고, 이 모의장치를 통과한 신호는 각 대역 통과 필터를 거친 신호의 자극패턴 (excitation pattern)으로 변환된다.

변환된 자극 패턴 중 고주파 영역의 성분들에 대해 이산 힐버트 변환을 적용하여 자극 패턴의 포락선을 추출할 수 있는데, 그림 2는 포락선 추출의 예를 나타낸 것이다.

그림 2에서 실선은 정파 정류된 자극패턴 신호를 의미하며, 점선은 이 신호로부터 추출된 포락선 신호를 나타낸다.

2.3. 고주파 포락선의 양이 시간차 왜곡 (Envelope Interaural Time Difference Distortion, EITDDist)

2.1.에서 기술했듯이 고주파 포락선과 저주파 성분의 양이 시간차는 청각 기관에서 유사한 방식으로 처리되기 때문에 고주파 포락선의 양이 시간차는 [1]에서 기술된 저주파 성분의 세부 구조에 기반한 양이 시간차와 동일한 방법으로 계산할 수 있다. EITD는 시간-주파수 분할된 양쪽 귀의 입력 신호에 대해 정규화된 교차 상관 함수 (Normalized Cross-correlation Function, NCF)를 이용

하여 계산할 수 있는데, 이는 식 (2)와 같이 표현할 수 있다. 식 (2)에서 $X_{L,k,n}$ 과 $X_{R,k,n}$ 은 양쪽 귀에서 발생하는 자극 패턴의 포락선 신호를 의미하며, d 는 샘플단위의 시간 지연, k 는 주파수 대역 지수 (index), 그리고 n 은 시간 프레임 지수를 의미한다.

$$NCF_{k,n}[d] = \frac{\sum_l X_{L,k,n}[l]X_{R,k,n}[l+d]}{\sqrt{\sum_l X_{L,k,n}^2[l]X_{R,k,n}^2[l]}} \quad (2)$$

고주파 포락선의 양이 상관계수 (Envelope InterAural Cross-correlation Coefficient, ELACC) 는 NCF의 최대 값으로 정의되고, 양이 시간차는 NCF가 최대값을 가질 때의 시간 지연값으로 정의될 수 있다. 교차 상관은 10ms 길이의 사각 창함수 (Rectangular Window) 를 적용한 후 각 시간 프레임 별로 계산되며, 각 창함수는 서로 7/8씩 겹치도록 설계되었다. EITD와 ELACC 는 시간-주파수 분할된 신호에 대해 식 (3)과 식 (4) 와 같이 계산된다. 프레임 길이와 겹침 비율은 [1]의 결과에 의해서 결정된 것인데, LSO가 관여하는 ILD 왜곡과 주관청취평가 결과 사이의 상관계수가 가장 높은 값을 동일하게 사용하였다. 또한 Sine-based 창함수를 사용할 경우 기저막에서 발생하는 자극 패턴을 온전하게 유지할 수 없고, 이로 인해 포락선의 형태가 변하게 되어 측면 상올리브핵에서의 포락선에 대한 반응을 모사하지 못하게 되므로, 각 프레임에는 사각 창함수를 사용하여 EITD를 계산하였다.

$$EITD[k, n] = \arg \max_d [NCF_{k,n}[d]]_{d=-N}^{d=N} \quad (3)$$

$$ELACC[k, n] = \max [NCF_{k,n}[d]]_{d=-N}^{d=N} \quad (4)$$

매개변수 N 은 d 의 범위인데, 이론적으로 가능한 양이 시간차의 값을 의미한다. EITD와 ELACC는 각각 참조 신호와 평가 신호에 대해 계산되고, 이 역시 아래첨자 *ref*와 *test*로 나타내었다. 또한, 고주파 포락선에 기반한 음원 방향의 인지적 차이는 단위 원 상에서 두 점 사이의 거리로 근사할 수 있기 때문에 참조신호와 평가신호의 EITD 차이는 식 (5)와 같이 계산할 수 있다. 식 (5)는 f_s 는 표본화 주파수 (sampling frequency)를 나타낸다.

$$\Delta EITD[k, n] = \sqrt{2 - 2 \cos(\pi \cdot f_s \cdot (EITD_{ref}[k, n] - EITD_{test}[k, n]) / N)} \quad (5)$$

이와 같은 방법으로 $\Delta EITD$ 를 계산한 후, 인지적으로 양이 시간차에 의한 음상 정위에 실패할 경우가 고려되어야 한다. ELACC가 매우 낮은 경우, EITD로 인해 인지되는 음원의 방향은 뚜렷하지 않기 때문에 인지되는 음원 방향 검출에 대한 확신도 (certainty)를 고려하기 위해 EITD의 차이값에 결정 인자 (decision factor)를 적용한다. 확신도를 모형화 하기 위해 탄젠트 S자 결정 함수 (Tangential Sigmoid Function)을 사용하였는데, 이 함수는 ELACC 값을 비선형적으로 변환하게 된다. 참조 신호와 평가 신호에 대해 ELACC 값은 식 (6)과 (7)에 의해 비선형 변환된다.

$$p_{ref}[k, n] = \left\{ 1 + e^{-S(ELACC_{ref}(k, n) - \tau_k)} \right\}^{-1} \quad (6)$$

$$p_{test}[k, n] = \left\{ 1 + e^{-S(ELACC_{test}(k, n) - \tau_k)} \right\}^{-1} \quad (7)$$

결정 인자가 $\Delta EITD$ 에 적용된 후의 EITD 왜곡은 식 (8)과 같이 계산된다.

$$EITDDist[k, n] = \frac{1}{2} (p_{ref}[k, n] + p_{test}[k, n]) \cdot \Delta EITD[k, n] \quad (8)$$

최종 EITD 왜곡은 식 (9)와 같이 전체 주파수 밴드와 시간 프레임에 걸쳐 평균값을 취하게 된다. 즉, EITD 왜곡은 EITD 차이로 인한 참조신호와 평가 신호에 대한 음원 위치의 인지적 거리를 의미하는 평균 EITD 왜곡을 나타낸다.

$$EITDDist = \frac{1}{N} \sum_{n=1}^N \left(\frac{1}{Z} \sum_{k=1}^Z EITDDist[k, n] \right) \quad (9)$$

EITD 왜곡의 성능을 다른 공간감 인자와 비교하기 위해 다른 공간감 인자인 ITD 왜곡, ILD 왜곡, IACC 왜곡은 [1]에서 계산된 것과 동일하게 계산되었다.

III. 공간감 인자로서 EITD 왜곡의 성능 비교

3.1. 주관청취평가 데이터베이스

다채널 음원에 대한 객관적 음질 평가에 있어서 공간감 인지들의 영향을 고찰하기 위해 우선적으로 주관청취평가가 시행되었다. 본 논문에 사용된 주관청취평가 데이터베이스는 ISO/MPEG 오디오 그룹에 의해 배포된 것

로서 ITU-R Rec. BS. 1534-1 “Multiple Stimulus with Hidden Reference and Anchor (MUSHRA)”에 권고된 바에 따라 시행되었다 [8-10]. 주관청취평가에는 11개 종류의 음원이 사용되었다. 각 음원은 11개의 다른 다채널 오디오 코딩 기법을 이용하여 부호화 및 복호화되어, 결과적으로 주관청취평가를 통해 121개의 아이템이 생성되었다. 주관청취평가에 사용된 음원 및 다채널 오디오 코딩 기법에 대한 세부사항은 [1]에 자세히 기술되어 있다.

3.2. EITDDist의 중요성

그림 3과 표 1은 주관청취평가 결과와 객관평가에 사용되는 14개의 평가인자 사이의 상관계수를 보여준다. 각 상관계수 $\rho_{x,y}$ 는 식 (10)과 같이 계산되었다.

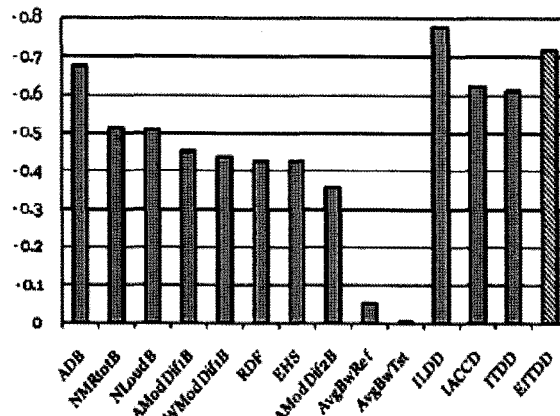


그림 3. 주관청취평가 결과와 14개 인자들 사이의 상관계수
Fig. 3. Correlation coefficients between Mean Opinion Scores (MOS) and fourteen features.

표 1. 주관청취평가 결과와 14개 인자들 사이의 상관계수
Table 1. Correlation coefficients between Mean Opinion Scores (MOS) and fourteen features.

평가인자	상관계수
ADB	-0.68
NMRtoB	-0.51
NLoudB	-0.51
AModDif1B	-0.45
WModDif1B	-0.44
RDF	-0.43
EHS	-0.43
AModDif2B	-0.36
AvgBwRef	-0.06
AvgBwTst	-0.00
ILDD	-0.78
IACCD	-0.62
ITDD	-0.61
EITDD	-0.72

$$\rho_{x,y} = \frac{COV(X,Y)}{\sigma_x \sigma_y} = \frac{E[XY] - E[X]E[Y]}{\sigma_x \sigma_y} \quad (10)$$

식 (10)에서 X는 MOS, Y는 각 인자의 데이터를 의미하며, 14개의 인자와 주관청취평가 결과의 상관계수는 양이 신호로 합성된 121개의 신호에 대해서 계산되었다. 14개의 평가인자 중 첫 10개는 현재 ITU-R Rec. BS. 1387-1에서 사용되고 있는 모형 출력 변수 (Model Output Variables, MOVs)이다. 10개의 모형 출력변수와 4개의 공간감 인자는 표 2와 표 3에 각각 요약하였다.

모든 인자들은 주관청취평가 결과와 음의 상관관계를 가지기 때문에, 그림 3 및 표 1에서 상관계수의 절대값이 클수록 더 좋은 음질 예측 성능을 가진다고 간주할 수 있다. 표 1에서 볼 수 있듯이, EITDDist는 주관청취평가 결과와 상관계수 0.72의 매우 높은 상관관계를 가짐을 알 수 있다. 특히, EITDDist는 주관청취평가 결과와의 상관계수 0.62를 갖는 IACCD나 상관계수 0.61을 갖는 ITDD보다 더욱 높은 상관 관계를 갖고 있으며, 기존

표 2. ITU-R Rec. BS. 1387-1에서 음색 열화에 대한 요인으로 사용되는 모형 출력 변수

Table 2. Model Output Variables (MOVs) for measuring timbral degradation in ITU-R Rec. BS. 1387-1.

인자	설명
ADB	평균 왜곡 블록 (Averaged distortion block). 왜곡된 블록의 전체 개수에 대한 전체 왜곡의 비
NMRtoB	기반 에너지 (masker energy) 대 평균 잡음 비의 로그 값
EHS	오차의 배음구조
BWRef	참조 신호의 대역폭
BWTest	평가 신호의 대역폭
AModDif1B	평균 변조 차이
AModDif2B	참조 신호에 변조가 거의 없는 경우 변조 변화에 가중치를 준 평균 변조 차이
WinModDif1B	창 함수를 적용한 평균 변조 차이
RDF	잡음 대 기반 에너지 비가 기준값 이상인 프레임의 개수
NLoudB	평균 잡음 세기

표 3. 공간감 열화를 나타내는 양이 요소
Table 3. Binaural features for measuring spatial degradation.

요소	설명
ITDDist	양이 시간차로 인해 발생하는 평가 신호와 참조 신호의 음원 방향 차이에 대한 인지적 거리
ILDDist	양이 크기차로 인해 발생하는 평가 신호와 참조 신호의 음원 방향 차이에 대한 인지적 거리
IACCDist	양이 상관계수차로 인해 발생하는 평가 신호와 참조 신호의 음원 넓이감 차이에 대한 인지적 거리
EITDDist	고주파 영역 포락선의 양이 시간차로 인해 발생하는 음원 방향 차이에 대한 인지적 거리

10개의 음색 왜곡 인자들보다도 높은 상관관계를 가짐을 확인할 수 있다. 이러한 결과들로부터 고주파 포락선 정보는 다채널 오디오 신호에 의한 공간감 인지와 전체적인 음질에 중요한 역할을 하고 있음을 확인할 수 있다. 또한 기존 ITU-R Rec. BS. 1387-1에서 사용되고 있는 음색 요소와 비교했을 때, 주관청취평가 결과와 4개의 공간감 요소는 유사하거나 더 높은 상관계수를 가짐을 확인할 수 있다. 이러한 결과를 바탕으로 다채널 오디오의 품질 평가에 있어서, 음색뿐만 아니라 공간감 요소 역시 중요함을 확인할 수 있다.

이와 같은 방법으로 계산된 각각의 인자는 다채널 오디오의 객관적 음질 평가를 위한 예측 모델의 입력 인자로서 사용될 수 있다. 현재 사용되고 있는 오디오 신호의 객관 품질 평가 모델에서는 10개의 음색 요소만을 사용하고 있지만, [1]에서 제안된 것과 같이 다채널 오디오 코딩 시스템의 객관적 음질 예측 모델에 주관청취평가 결과와 높은 상관계수를 갖는 공간감 왜곡 요소를 입력 인자로 추가했을 때 더욱 좋은 예측 성능을 보임을 확인할 수 있다. 이러한 다채널 오디오의 객관적 품질 예측에 있어서, *EITDDist*는 기존의 공간감 왜곡 인자와 더불어 객관적 음질 예측 모델에서 공간감 왜곡을 평가하는 인자로서 사용될 수 있다. 특히, *EITDDist*는 주관적 청취평가 결과와 높은 상관관계를 가지기 때문에, *EITDDist*를 음질 예측 모델의 입력 인자로 추가함으로써 다채널 오디오 코딩 시스템의 객관적 음질 예측 모델의 성능을 향상시킬 수 있을 것으로 기대된다.

IV. 결론

본 논문에서는 공간감 인자로서 고주파 포락선의 양이 시간차의 음원 위치 인지에 대한 효용성을 심리 음향적, 생리학적 이론에 기반하여 설명하였다. 일반적으로 알려져 있는 저주파 대역의 양이 시간차, 고주파 대역의 양이 크기차와 더불어 고주파 포락선의 양이 시간차는 인간의 음상 정위에 또 다른 중요한 요소 중에 하나임을 확인하였다. 또한 머리전달함수를 이용하여 합성된 양이 신호로부터 자극 패턴을 추출하고 이로부터 고주파 포락선 신호의 양이 시간차 왜곡을 계산하였다. 계산된 고주파

포락선의 양이 시간차 왜곡이 다채널 오디오 코딩 시스템의 객관평가 인자로서 효용성이 있는지를 확인하기 위하여 주관청취평가 결과와 상관계수를 계산하였다. 본 상관관계 분석으로부터 고주파 포락선의 양이 시간차 왜곡은 기존 10개의 음색 요인보다 높은 상관계수를 가짐을 확인할 수 있었으며, 다른 공간감 인자들과 비교하였을 경우에도 유의한 수준 또는 더 높은 상관계수를 가짐을 역시 확인할 수 있었다. 이러한 결과를 보았을 때, 고주파 포락선의 양이 시간차 왜곡 역시 다채널 코딩 시스템의 객관 음질평가에서 중요한 인자로 사용될 수 있음을 확인하였다.

참고 문헌

1. I. Choi, B. G. Shinn-Cunningham, S. B. Chon, K. -M. Sung, "Objective measurement of perceived auditory quality in multi-channel audio compression coding systems," *J. Audio Eng. Soc.*, vol. 56, pp. 3-17, 2008.
2. ITU-R Recommendation BS. 1387-1, *Method for objective measurement of perceived audio quality*, International telecommunication union, Geneva, Switzerland, 1999.
3. B. C. J. Moore, "An introduction to the psychology of hearing," Academic press, London, 1997.
4. P. X. Joris and T. C. Yin, "Envelope coding in the lateral superior olive: I. sensitivity to interaural time differences," *J. Neurophysiol.*, vol. 73, pp. 1043-1062, 1995.
5. L. R. Bernstein and C. Trahiotis, "Enhancing sensitivity to interaural delays at high frequencies by using "Transposed stimuli",
J. Acoust. Soc. Am. 112: 1026-1036, 2002.
6. S. J. Griffin, L. R. Bernstein, N. J. Ingham, and D. McAlpine, "Neural sensitivity to interaural envelop delays in the inferior colliculus of the guinea pig," *J. Neurophysiol.*, vol. 93, pp. 3463-3478, 2005.
7. ITU-R Recommendation BS. 1116-1, *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, International telecommunication union, Geneva, Switzerland, 1997.
8. ISO/IEC JTC1/SC29/WG11 (MPEG) Document N6813, *Report on spatial audio coding RMO selection tests*, Palma de mallorca, Spain, 2004.
9. ISO/IEC JTC1/SC29/WG11 (MPEG) Document N7138, *Report on MPEG spatial audio coding RMO listening tests*, Busan, Korea, 2005.
10. ITU-R Recommendation BS. 1534-1, *Method for the subjective assessment of intermediate quality levels of coding systems*, International telecommunication union, Geneva, Switzerland, 2003.

저자 약력

•서 정 훈 (Jeong-Hun Seo)

2003년 서울대학교 전기컴퓨터공학부 졸업
 2004년 ~ 현재 서울대학교 전기컴퓨터공학부 석박사통합과정
 ※ 관심분야: 음향 신호처리

•전 상 배 (Sang Bae Chon)

2000년 전북대학교 전기전자제어공학부 졸업
 2005년 서울대학교 전기컴퓨터공학부 석사 졸업
 2009년 서울대학교 전기컴퓨터공학부 박사 졸업
 2009 ~ 현재 서울대학교 전기컴퓨터공학부 BK사업단 박사후 연구원
 ※ 관심분야: 음향학, 음향 및 음성 신호처리

•성 평 모 (Koeng-Mo Sung)

1965년 ~ 1971년: 서울대학교 전자공학과
 1971년 ~ 1973년: 독일 아헨공대 Vordiplom
 1973년 ~ 1977년: 독일 아헨공대 전자통신공학 Dipl.-Ing.
 1977년 ~ 1978년: 독일 아헨공대 음향공학 Dr.-Ing.
 1977년 ~ 1983년: 독일 아헨공대 음향공학연구소 연구원
 1983년 ~ 현재: 서울대학교 공과대학 전기컴퓨터공학부 교수