

# 멀티채널 룸임펄스 응답 기반 외재화 알고리즘

장인선 이 용 주 장 대 영 강 경 옥

한국전자통신연구원

{jinsn, draball, dyjang, kokang}@etri.re.kr

## Sound Externalization using Multichannel Room Impulse Response

Inseon Jang, Yong-Ju Lee, Dae Young Jang, Kyeongok Kang

Electronics and Telecommunications Research Institute

### 요약

헤드폰 또는 이어폰으로 오디오 청취 시 흔히 음상이 머리 내부에 맺히는 현상이 발생하게 되며, 이러한 현상을 음상 내재화(Inside Head Localization; IHL)라 한다. 오디오의 음상이 머리 주변 혹은 머리 내부에 맺히게 되면 공간감이나 입체감이 떨어지게 되어 음향의 현실감을 저하시키게 되며 또한 청취에 따른 피로도가 증가하게 된다. 이러한 음상 내재화 현상을 제거하여, 헤드폰/이어폰을 통해 오디오 청취 시 음상이 머리의 외부에 맺히도록(Out of Head Localization; OHL) 하는 기술을 음상 외재화(Sound Externalization) 기술이라 한다.

룸 임펄스 응답이 방향 큐와 연계하여 생성되었을 경우 외재화가 가능하다는 실험적 사실을 바탕으로 기존의 음상 외재화 방법은 일반적인 HRTF (Head Related Transfer Function)를 이용하여 외재화 필터를 구성해왔다. 본 논문에서는 구체마이크로폰을 이용하여 녹음한 멀티채널 룸 임펄스 응답을 기반으로 모델링 된 외재화 필터를 이용한 음원 외재화 방법을 제안한다. 또한 실험 및 결과 분석을 통하여 본 알고리즘의 전방 음원 외재화 성능의 우수성을 입증하고, 외재화 알고리즘 적용 후의 원 신호 음상 보존 성능을 확인한다.

## 1. 서론

최근 멀티미디어 기술의 발전으로 인해 MP3 플레이어, PMP, DMB 등과 같은 휴대 단말을 통한 멀티미디어 콘텐츠의 소비가 증가하고 있다. 특히, 휴대폰에 MP3 플레이어가 내장되고, DMB 수신 기능의 추가되는 등의 복합화, 다기능화 추세에 따라 이러한 경향은 더욱 증가하고 있다. 이러한 휴대 단말이 소형화, 경량화 되면서 일반적으로 사용자들은 휴대 단말에 내장되어 있는 스피커보다는 헤드폰/이어폰을 이용하여 오디오를 청취하고 있다. 이 경우 음상이 머리 내부에 맺히는 현상이 발생하게 되는데, 이러한 현상을 음상 내재화(Inside Head Localization; IHL)라 한다. 이는 공간감이나 입체감을 떨어지게 하여 음향의 현실감을 저하시키는 요인이 되며, 또한 청취자로 하여금 피로를 쉽게 느끼게 한다[1].

이러한 문제를 해결하여 헤드폰/이어폰을 통한 청취 시 음상이 머리의 외부에 맺히도록 (Out of the Head Localization; OHL) 하는 기술을 음상 외재화(Sound Externalization) 기술이라 하며 공간의 반사 및 잔향에 의한 공간 음향 특성, 개인의 머리 및 귓바퀴 등 인체에 의한 음향 전달 특성, 머리 움직임에 의한 음향 전달 특성 변화 등을 이용한 접근들이 연구되어 오고 있다. 이들 중 공간 음향 특성 기반 외재화 기술은 반사 및 잔향이 사람이 느끼는 현장감에 큰 영향을 미친다는 연구 결과에 근거한다[2, 3]. 또한 반사 및 잔향이 HRTF(Head Related Transfer Function)와 같이 개인화가 필요한 정보가 아니며 구현하는데 있어 계산량이 적다는 장점 등으로 인하여 현재 상용화되고 있는 대부분의 상용 외재화 기술에서 이를 이용하고 있다.

기존의 공간 음향 특성 기반의 외재화 기술들은 반사음과 잔향을 적용하여 위해 다양한 방법들을 사용하고 있으며 예로서, 여러 개의 HRTF 및 이득/지연을 이용하여 반사음을 추가하는 방법, 임의의 각도에서 반사음을 가정하고 인공 잔향을 추가하는 방법 등이 있다[4, 5].

본 논문에서는 구체마이크로폰을 이용하여 측정된 멀티채널 룸 임펄스 응답을 기반으로 외재화 필터를 모델링하여 음상 외재화 성능을 개선한다. 이는 일반적인 더미헤드(dummy head)를 이용하여 무향실에서 녹음한 기존의 HRTF 대신 특정 공간에서 5채널의 구체마이크로폰을 이용하여 녹음한 멀티채널 룸 임펄스 응답을 이용함으로써 우수한 외재화 성능과 자연스러운 잔향을 제공한다.

본 논문의 구성은 다음과 같다. 2장에서는 멀티채널 룸 임펄스 응답 기반 음상 외재화 방법을 설명한다. 3장에서는 외재화 알고리즘의 전방 음원 외재화 성능 및 음상 보존 성능 평가를 위한 청취평가 환경 및 방법을 설명하고 그 결과를 분석한다. 또한 기존의 상용 음상 외재화 기술과의 성능 평가를 통해 제안한 기술의 우수성을 입증한다. 마지막으로, 4장에서는 결론 및 향후 연구 방향에 대해 논한다.

## 2. 멀티채널 룸 임펄스 응답 기반 외재화 알고리즘

그림 1과 같이  $0^\circ, \pm 75^\circ, \pm 105^\circ$ 의 마이크로폰 레이아웃을 갖는 5채널 구체마이크로폰을 이용하여 전방  $\pm 30^\circ$ 에 위치하고 있는 음원으로 부터의 룸 임펄스 응답을 측정하였다. 여기에서, 전방  $\pm 30^\circ$ 의 가상 음원 위치를 가정한 것은 기존 스테레오 콘텐츠 음상의 자연스러운 외재

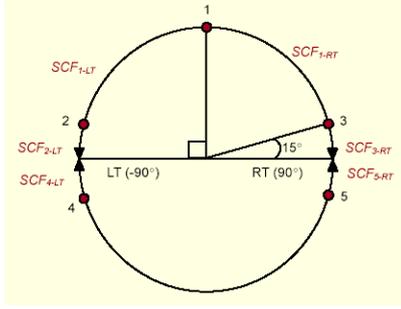


그림 1 구체마이크로폰의 마이크로폰 배치

를 위함이다.

측정된  $-30^\circ$  음원으로부터의 립임펄스 응답  $h_{L1}, \dots, h_{L5}$  과  $+30^\circ$  음원으로부터의 립 임펄스 응답  $h_{R1}, \dots, h_{R5}$  을 이용하여 외재화를 위한 channel-to-ear 필터  $H_L^L, H_L^R, H_R^L, H_R^R$  을 식 (1)과 같이 모델링할 수 있다.

$$\begin{bmatrix} H_L^L & H_L^R \\ H_R^L & H_R^R \end{bmatrix} = \begin{bmatrix} SCF_{1-LT} & SCF_{2-LT} & 0 & SCF_{4-LT} & 0 \\ SCF_{1-RT} & 0 & SCF_{3-RT} & 0 & SCF_{5-RT} \end{bmatrix} * \begin{bmatrix} h_{L1} & h_{R1} \\ h_{L2} & h_{R2} \\ h_{L3} & h_{R3} \\ h_{L4} & h_{R4} \\ h_{L5} & h_{R5} \end{bmatrix} \quad (1)$$

여기에서  $[A * B]_{ij} = \sum_{k=1}^N a_{ik} * b_{kj}$  (단,  $A \in R^{M \times N}, B \in R^{N \times L}$  및  $i=0, \dots, M$  이며  $j=0, \dots, L$ ) 그리고,  $H_i^j$  는  $i$  channel-to- $j$  ear 필터이며, SCF(Sphere Conversion Filter)는 구체마이크로폰의 멀티채널 출력을 스테레오 신호로 변환하기 위한 필터로써,  $SCF_{0-355} = conv(SIR_{0-355}, SIR_0^{-1})$  (SIR: Sphere Impulse Response)로 계산된다[6].

이와 같이 계산된 channel-to-ear 필터를 기반으로 한 음상 외재화 알고리즘의 구조도는 그림 2와 같다. 스테레오 오디오 신호가 외재화 필터에 입력되면, 각 채널 신호는 멀티채널 립임펄스 응답을 이용하여 모델링한 channel-to-ear 필터에 필터링 된 후 L, R ear에 해당하는 신호들의 합을 통하여 외재화 스테레오 신호를 생성한다.

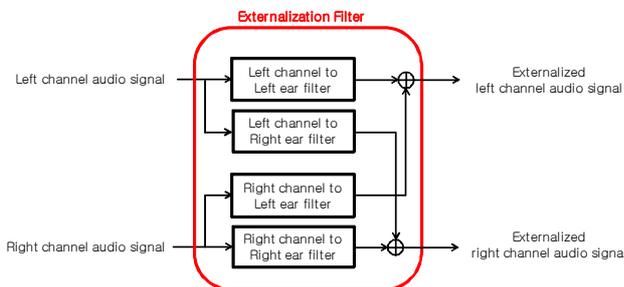


그림 2 멀티채널 립임펄스 응답 기반 외재화 방법 구조도

제한된 외재화 알고리즘은 구체마이크로폰을 이용하여 녹음한 멀티채널 립임펄스 응답을 기반으로 하여 외재화 필터를 모델링함에 의

해 전방 음상 강조되며 또한 사람의 머리에 따른 반사 및 회절 등과 자연스러운 크로스 톡(cross-talk) 및 잔향이 반영되어 효과적인 외재화 성능을 제공한다.

### 3. 청취 실험

본 장에서는 제안된 기술의 음상 외재화 및 음상 정위 성능에 대한 실험 방법 및 결과를 정리분석하고 상용기술인 SRS Headphone과 비교청취 평가 결과 통해 제안된 기술의 성능을 검증한다.

#### 가. 실험 환경 및 방법

제한한 음상 외재화 알고리즘의 성능 평가를 위해, 음상 외재화 거리에 대한 평가와 음상 변화를 측정하는 실험을 수행하였다. 일반적으로 오디오 신호에 다양한 신호처리를 수행하면 음상이 변화하는 문제가 발생할 수 있으므로, 음상의 변화를 측정하는 실험은 외재화 알고리즘의 음상 보존 성능을 검증하기 위해서 필요하다. 음상 외재화 알고리즘의 성능 평가는 총 15명의 피험자를 대상으로 수행되었다.

음상 외재화 거리를 측정하기 위한 실험은 모노 음원과 스테레오 음원에 대해 각각 수행하였다. 이는 일반적으로 좌우 신호의 상관성이 높은 스테레오 신호의 경우 외재화 효과가 모노 신호에 비해 크지 않으므로, 서로 다른 외재화 성능을 나타낼 수 있을 것으로 예상되기 때문이다.

모노 음원의 외재화 거리 측정 실험은 아래와 같은 절차에 따라 이루어졌다. 먼저, 피험자의  $30^\circ$ , 1m 전방에 스피커를 위치시키고, 모노의 백색잡음 신호를 스피커를 통해 들려주어서, 실제  $30^\circ$  각도의 1m 전방에서 재생되는 소리를 피실험자로 하여금 인지하도록 한다. 다음으로 피실험자에게 이어폰을 착용하게 한 후, 원 신호인 모노의 백색잡음 신호를 청취하도록 하여 음상 내재화 현상을 인지하도록 하였다. 충분히 숙지하도록 한 후, 모노의 백색잡음 신호에 외재화 알고리즘을 적용한 신호를 청취하도록 하고, 피험자가 인지하는 음상 외재화 거리 정도를 손으로 가리키도록 한 후 이 거리를 측정하였다.

스테레오 음원의 외재화 거리 측정 실험도 모노 음원의 외재화 거리 측정 실험과 유사한 방법으로 이루어졌는데, 그 절차는 아래와 같다. 먼저, 피험자의  $\pm 30^\circ$ , 1m 전방에 스피커들을 위치시키고, 스테레오 백색잡음 신호를 스피커를 통해 들려주어서, 실제 1m 전방에서 재생되는 소리를 인지하도록 한다. 이때, 좌우 모두 백색 잡음 신호이므로, 음상은 피험자의 전방에 맺히게 된다. 다음으로 피실험자에게 이어폰을 착용하게 한 후 원 신호인 스테레오의 백색잡음 신호를 청취하도록 하여 음상 내재화 현상을 인지하도록 하였다. 1m 전방에서 좌우 스피커에 의해 재생되는 소리와 이어폰 착용 시에 음상이 내재화되는 정도를 충분히 숙지하도록 한 후, 스테레오의 백색잡음 신호에 외재화 알고리즘을 적용한 신호를 청취하도록 하고, 피험자가 인지하는 음상 외재화 거리 정도를 손으로 가리키도록 한 후 이 거리를 측정하였다.

외재화 알고리즘에 따라 음상이 변화하는 정도를 측정하는 실험은  $0^\circ \sim 180^\circ$  사이에  $30^\circ$  간격으로 총 7개의 각도에 대해 이루어 졌는데, 이는 음상의 각도에 따라 외재화 신호처리 후에 음상이 변화하는 정도가 다르게 나타날 수 있기 때문이다. 음상 변화에 대한 실험은 다음과 같은 절차에 따라 이루어졌다. 먼저, 백색 잡음에 위에서 제시한 해당 각도의 HRTF를 이용하여 렌더링한 신호(reference)를 이어폰을 통해 청취하도록 하여 피험자가 해당 각도에 대해 충분히 인지하도록

한다. 이 후, reference 신호에 외재화 알고리즘을 적용한 신호를 피험자에게 들려주고, 피험자가 인지하는 음상 정위 각도를 지시하도록 하여, 그 각도를 측정하였다.

이 외에 기존 상용기술과의 외재화 성능 비교를 위하여 총 8명의 피험자를 대상으로 SRS Headphone 과의 비교청취를 실시하였다. 본 비교청취평가에서는 기준이 되는 신호(reference)는 주어지지 않았으며, 피험자로 하여금 오직 두 시스템 (A: 제안된 방법, B: SRS Headphone)의 출력 신호에 대한 외재화 정도를 비교하도록 하여 상대적인 외재화 성능을 평가하였다. 제안한 외재화 알고리즘의 경우 실측된 롬 임펄스 응답을 이용하였고 이를 외재화 필터로 모델링하는 과정에서 발생하는 음색 변화에 대해 추가적인 전처리 후처리 작업을 하지 않았기 때문에, 비교 청취 시에 피험자에게 실험 콘텐츠의 음색은 고려하지 않고, 외재화 정도만을 평가하도록 하였다.

표 1은 비교 평가를 위한 점수표이다. 실험 콘텐츠로는 약 20~25 초 길이의 오디오 클립(44.1kHz 샘플링 율)을 이용하였고, 클래식, 가요 및 음성 3 종류 콘텐츠를 이용하였다.

표 1 비교청취평가를 위한 점수표

1	A is more external B
0	There is little difference between A and B
-1	B is more external A

### 나. 실험 결과

제안된 알고리즘의 음상 외재화 거리에 대한 청취평가 결과는 그림 3과 같다. 그림 보는 바와 같이, 청취자의 이마 전방으로 모노 신호는 평균 16.74 cm, 스테레오 신호는 평균 13.43 cm 의 외재화 성능을 제공하는 것으로 나타났다. 즉, 머리 내에 정위되던 음상이 제안된 방법을 이용하면, 머리 밖 10cm 이상에 정위된다는 것을 알 수 있다. 또한, 모노 신호에 비해 스테레오 신호에 대한 외재화 성능이 떨어지는 것을 볼 수 있는데, 이는 앞서 설명하였던 것과 같이, L, R 신호 간의 유사도(correlation)가 높은 경우 음상이 피험자 가까이에 맺히기 때문으로 판단된다.

외재화 알고리즘에 따라 음상이 변화하는 정도를 측정하는 실험의 결과를 그림 4와 같다. 그림 4의 (가)는 각도에 따라 음상이 변화한 값의 평균과 95% 신뢰구간의 값을 나타낸 것이고, (나)는 전체 평균값을 나타낸 것이다. 청취평가 전 훈련(training)을 통해 충분히 reference를 인지하도록 하여, 피험자가 전방과 후방을 혼동(front back confusion)하는 경우는 없었다. 결과로써, 전방 보다는 후방에 있는 소리의 음상 변화가 크다는 것을 알 수 있었는데, 이는 사람의 음상 인지에 대해 전방 해상도가 후방 해상도에 비해 높다는 기존 사실에 기반하는 것으로 이해할 수 있다. 각도에 따라 차이는 있지만, 음상이 평균적으로  $-0.75^\circ$  에  $\pm 10.65^\circ$  정도로 변화함을 확인하였으며, 이는 제안한 외재화 알고리즘이 원래의 음상에 변화를 초래하지만, 그 변화량은 크지 않다는 것을 나타낸다.

마지막으로, SRS headphone과의 음상 외재화 정도에 대한 비교 청취 평가 결과는 표 2와 같다. 실험 결과 클래식, 가요, 음성에 대해 75%, 87.5%, 87.5%의 피험자들이 제안된 기술이 SRS Headphone

보다 더 우수하거나 유사한 음상 외재화 현상을 인지하는 것으로 나타났다.

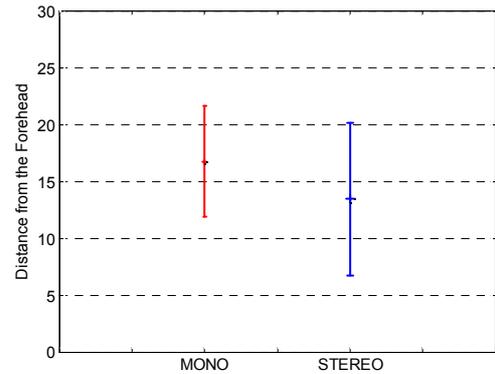
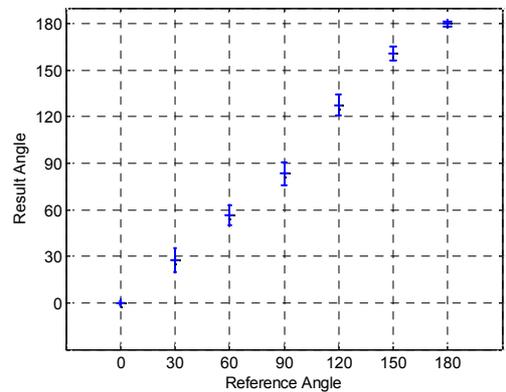
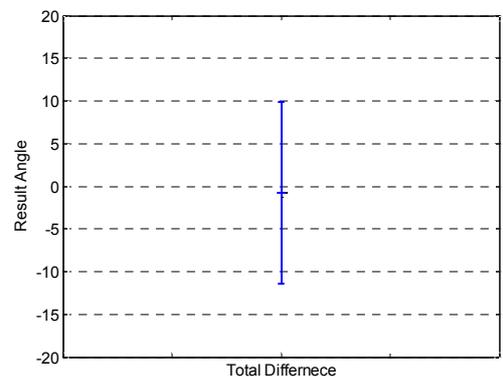


그림 3 음상 외재화 청취 평가 결과



(가)



(나)

그림 4. 음상 정위 청취평가 결과: (가) 각도별 결과, (나) 평균 결과

표 2 외재화 성능 비교청취 평가결과

Relative Distance	Percentage (%)
A>B	75
A=B	8.33
A<B	16.67

#### 4. 결론

본 논문에서는 구체마이크로폰을 이용하여 녹음한 멀티채널 룸 임펄스 응답 기반의 음성 외재화 알고리즘을 제안하였다. 제안된 기술은 기존에 사용되어 온 일반적인 HRTF 대신 특정 공간에서 5채널의 구체마이크로폰을 이용하여 녹음한 멀티채널 룸 임펄스 응답을 이용하여 외재화 필터를 구성하여 전방 음성 강조는 물론이고 사람의 머리에 따른 반사 및 회절 등과 자연스러운 크로스 톡(cross-talk) 및 잔향이 반영되어 효과적인 외재화 성능을 제공한다. 이에 대한 전방 외재화 성능 및 음장 보존 정도를 청취평가를 통하여 확인하였으며, 상용 외재화 기술인 SRS Headphone 과의 비교청취를 통하여 제안한 기술의 외재화 성능에 대한 우수성을 입증하였다.

이를 기반으로, 제안한 알고리즘의 음색 변화에 대한 고려 및 전후 처리 기술 개발을 통한 음질 보안을 진행하고 있으며, 향후 휴대단말에서 동작하는 음성 외재화 기술의 최적화 구현을 계획하고 있다.

#### 5. 감사의 글

본 연구는 정보통신부 및 정보통신연구진흥원의 IT신성장동력핵심기술개발사업의 일환으로 수행하였음. [2007-S-004-01, 무안경 개인형 3D 방송기술]

#### 6. 참고 문헌

- [1] N. Sakamoto, T. Gotoh and Y. Kimura, "On 'Out-of-Head Localization' in Headphone Listening," *JAES*, vol. 24, pp. 710-716, Nov. 1976.
- [2] Durlach, et. al. *On the externalization of auditory images*, Presence, vol.1, number 2, 1992.
- [3] Youngsik Yoon, Youngjin Park, "Externalization of sound images in headphone-based 3D sound system", *Proceeding of International Conference on Control, Automation and Systems (ICCAS)*, Muju Resort, Jeonbuk, Korea, pp. 586-588, 2002.
- [4] P. Rubak, "Headphone signal processing system for Out-of-head localization," *90th AES Convention*, 1991.
- [5] T. Choi, Y. Park and D. Youn, "Efficient out of head localization system for mobile applications," *120th AES Convention*, 2006.
- [6] T. Lee, D. Jang, K. Kang, J. Kim, D. Jeong, and H. Hamada, "A Spatial Audio System Using Multiple Microphones on a Rigid Sphere," *ETRI Journal*, vol.27, no.2, pp.153-165, Apr. 2005.