

맞춤형 머리전달함수 구현을 위한 모델링 기법

Modeling HRTFs for Customization

신기훈† · 박영진* · 박윤식**

Ki H. Shin, Youngjin Park and Yoonshik Park

Key Words : HRTF, Pole-Zero Modeling, Customization, Virtual Audio

ABSTRACT

This study reveals some recent attempt in modeling empirically obtained B&K HATS (Head and Torso Simulator) HRTFs (Head Related Transfer Functions) to isolate parameters that stimulate lateral and elevation perception. Localization using non-individual HRTFs often yields poor performance in synthesizing virtual sound sources when applied to a group of individuals due to differences in size and shape of head, pinnae, and torso. For realization of both effective and efficient virtual audio it is necessary to develop a method to tailor a given set of non-individual HRTFs to fit each listener without measuring his/her HRTF set. Pole-zero modeling is applied to fit HRIRs (Head Related Impulse Responses) and modeling criterions for determining suitable number of parameters are suggested for efficient modeling. Horizontal HRTFs are modeled as minimum-phase transfer functions with appropriate ITDs (Interaural Time Delay) obtained from RTF (Ray Tracing Formula) to better fit the size of listener's head for usage in simple virtualizer algorithms without complex regularization processes. Result of modeling HRTFs in the median plane is shown and parameters responsible for elevation perception are isolated which can be referred to in the future study of developing customizable HRTFs.

1. 서론

가상음장 재현을 위한 머리전달함수에 관한 연구는 MIT의 Bill Gardner에 의한 KEMAR HRTF 공개(1994)를 비롯하여 이후 University of Wisconsin(미국)과 ISVR Lab.(영국)등 여러 곳에서 다양하게 이루어졌다. 머리전달함수란 인간 머리 주변에 위치한 음원에서 청취자 고막까지의 음파의 전달 경로를 수학적 전달함수로 표현한 것으로서 머리와 몸통 및 외이에 의한 음파의 산란과 회절로 비롯되는 음파의 왜곡 정보를 모두 담고 있다. 대상 인간의 머리 주변 임의의 곳에 스피커를 두고 이에 의해 고막에서 느껴지는 음향 신호와 스피커 입력신호를 채취하여 둘 사이의 전달함수를 계산하고, 이를 역으로 이용하여 묘사하고자 하는 소리에 convolution 형태로 곱하여 헤드폰으로 들려주면 청취자는 임의의 방향에서 소리가 들리는 것을 느끼게 된다. 머리전달함수의 이 같은 성질을 이용하여 근래에는 기존의 5.1 내지 7.1 채널 홈씨어터 시스템을 임의의 위치에 가상 스피커를 생성하여 전면 2 채널 서라운드 시스템으로 대체하는 virtualizer(1)가 국내외산 디지털 TV에 탑재되어 출시되고 있다. 휴대폰과 MP3 Player에도 적용되어 이어폰으로 듣는 음악에 다양한 공간감 및 방향감을 실어주어 음질의 향상과 live band 앞에서 듣는 듯한 실감음질 구현이 이루어지고 있는 추세다. 이 밖에 on-line game의 가상현실의 몰입감 향상에도 머리전달함

수가 쓰이는 등 이 기술의 응용은 앞으로 꾸준히 발전될 전망이다.

위와 같은 머리전달함수의 응용에 있어 아직까지 해결되지 않은 많은 문제점들 중 몇 가지만 서술하자면 우선 더미헤드를 이용하여 측정된 머리 전달함수가 개개인의 신체와 맞지 않음으로써 생기는 앞뒤 및 위아래 혼동, 또는 음원의 고도를 전혀 느끼지 못하며 머리전달함수의 방위각조차도 다르게 인지되는 문제점들이 있다. 머리중심점과 양 귀를 포함하는 수평면에서의 방향성은 양 귀간의 음파 도달 시간차인 ITD와 양 귀간의 음압차인 ILD(Interaural Level Difference)에 의해 지배적으로 영향을 받으며 외이에 의한 음파 왜곡은 상대적으로 훨씬 덜 영향을 준다. 이에 반해 방위각 0도에서 머리의 양 옆을 가르는 median plane에서는 ITD와 ILD가 각각 0이며 따라서 이 평면에서의 방향성은 외이의 영향이 지배적이라 할 수 있다. 바로 이 때문에 더미헤드의 머리 전달함수를 다수의 청취자들에게 들려줄 경우, 수평면에서의 방향성은 어느 정도 맞게 인지하나 대부분 고도를 인지하지 못한다.(2) ITD만 어느 정도 맞으면 수평면에서의 방향성은 대부분 인지하나 방위각 0도에서의 고도 변화는 더미헤드의 외이와 개개인의 외이가 그 크기와 모양이 현저하게 다르기 때문에 인지하지 못하는 것이다. 이 같은 문제점들을 해결하는 가장 분명한 방법은 개개인의 머리전달함수를 모두 측정하여 사용하는 것이나 실용적이지 못하며 이를 바탕으로 한 virtual audio의 구현은 사실상 불가능하므로 non-

individual HRTF 를 개개인에 맞추는 기법이 필요하다. 맞춤형 HRTF 개발은 최근 몇 년간 매우 다각적인 측면에서 이루어져 왔으나 아직 이렇다 할 성공을 거두고 있지 않다. 대표적인 방법으로는 HRTF 의 numerical modeling(3)과 frequency scaling(4), 또는 database matching(5)이 있다. 본 연구의 목적은 numerical modeling 의 한 방법인 pole-zero modeling(6)을 적용하여 더미헤드의 머리전달함수를 모델링하고 방향성 인지에 필요한 최소한의 핵심 요소들을 분리하여 이를 토대로 개개인에게 보다 잘 맞는 맞춤형 머리 전달함수를 개발하는 것이다. 이의 첫 단계로서 본 논문에서는 수평면상의 머리전달함수와 머리 양 옆을 가르는 중앙면상의 머리전달함수의 모델링 결과와 각각의 경우 방향성 인지에 필요한 최소한의 핵심 요소들을 머리전달함수 분석(7)에 의거하여 결정하고 이를 토대로 효율적인 virtualizer 구현에 쓰일 수 있는 맞춤형 수평면 머리 전달함수를 구하는 기법을 소개한다.

2. B&K HATS 머리전달함수

맞춤형 머리전달함수의 모태로 본 연구실에서 측정된 B&K HATS 머리전달함수를 사용하였다. 머리중심점에서 음원까지의 거리를 0.2 m 에서 1 m 까지 0.2 m 간격으로 변화시키며 각 거리 별로 음원의 방위각과 고도각을 5 도씩 바꾸어 측정하였고 이 중 거리가 1 m 에 해당되는 원거리 머리 전달함수를 모델링하였다.(7) Figure 1 은 이때 사용된 측정 장치의 모습이다. 측정된 머리전달함수를 IFT(Inverse Fourier Transform)하여 얻어진 임펄스 응답을 각도별로 이어서 그려보면 시간영역에서 각도에 따른 변화를 쉽게 관찰할 수 있다. Figure 2 는 음원의 거리가 1 m 일 때 수평면 머리전달함수의 왼쪽 귀 임펄스 응답을 각도별로 이어 위에서 바라본 모습이다. 더미헤드 마이크로폰 시스템을 구성하는 각 부위별로 임펄스 응답에 미치는 영향을 살펴보면 다음과 같은 사실을 유추할 수 있다. 음파가 스피커에서 방사되어 외이 안에 장착된 마이크로폰에 먼저 도달하고 이어서 외이, 머리, 몸통 순으로 귀에서 가까운 부위에 반사되거나 회절, 산란된 음파가 마이크로폰에

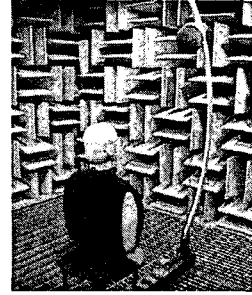


Figure 1: HRTF measurement setup.

도달한다. 따라서 Figure 2 에 보이는 임펄스 응답을 각 부위별 응답으로 나눌 수 있다. 임펄스 응답이 시작되는 지점으로부터 0.0002 초 안에 있는 부분은 외이에 의한 산란과 회절의 영향이며 그 다음 0.0004 초는 머리와 어깨, 이후는 어깨 아래 상반신의 응답을 나타낸다. 턴테이블에 의한 반사는 대략 각 임펄스 응답의 시작점에서 0.0017 초 떨어진 지점에 나타나므로 이후 보이는 작은 응답들은 더미헤드 마이크로폰 시스템과 무관한 머리전달함수 측정 장치 하단과 바닥에 의한 영향이라고 볼 수 있다. 아울러 음원이 전방에서 후방으로 시계반대방향으로 돌아감에 따라 왼쪽 외이 응답의 변화를 살펴보면 음원이 전방에서 후방으로 이동해감에 따라 첫 번째와 두 번째 peak 사이의 ridge 가 점차 사라지며 두 개의 peak 들이 합쳐지고 첫 번째 peak 의 모양도 변한다. 이는 음원이 전방에 위치할 경우 음파가 외이의 fossa 와 concha 와 같은 pinna cavity 에 의해 변조되고 음원이 후방에 위치할 경우엔 이러한 pinna cavity 들이 pinna 에 의해 가려져 그 영향이 사라지는 것으로 해석할 수 있다.

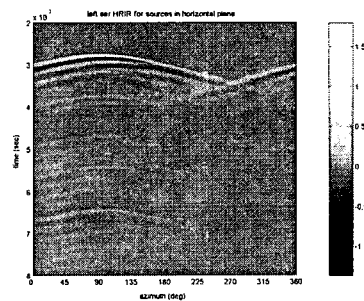


Figure 2: Horizontal HRIR (Head Related Impulse Response).

3. HRTF Modeling

3.1 HRTF Cleansing

모델링에 들어가기에 앞서 머리전달함수의 임펄스 응답에서 중요한 부분만을 살리고 나머지 불필

† Ph.D. Candidate, NoViC, KAIST
E-mail: ks007b@kaist.ac.kr
Tel: (042) 869-3076, Fax : (042) 869-8220

* Professor, NoViC, KAIST

** Professor, NoViC, KAIST

요한 부분에서의 응답은 삭제하는 이른바 cleansing 작업이 필요하다. 임펄스 응답의 전체 길이에서 외이 응답 전에 보이는 신호와 외이 응답 시작점에서 0.0017 초 지난 시점에 보이는 턴테이블에 의한 반사 이후에 보이는 신호는 모두 불필요한 부분들이므로 이를 삭제하고 0 으로 대체한다. 이는 우리가 필요한 더미헤드 마이크로폰 시스템에 의한 응답만 모델링하기 위해서이다. Cleansing 처리된 머리전달함수는 Figure 3 에서 볼 수 있듯이 그 차이가 확연하다.

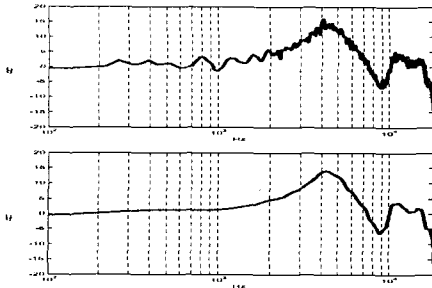


Figure 3: Empirical HRTF (top panel) and cleansed HRTF (bottom panel)

3.2 수평면 HRTF 의 최소위상화

1 장에서 언급한 바와 같이 수평면상의 머리전달함수를 다수의 사람들에게 적용해 보면 개인마다 차이는 있으나 대부분 음원의 각도를 어느 정도 맞게 인지한다. 이는 곧 수평면상에서는 음원의 각도에 따른 외이 응답의 변화보다 양 귀사이의 전체적인 위상차인 group delay 가 더 큰 영향을 준다는 것을 의미한다. 따라서 cleansing 된 수평면 HRIR 의 cepstrum 을 구하여 이로부터 다음 식을 만족하는 최소위상을 갖는 머리전달함수의 임펄스 응답을 구한다.(8)

$$H(z) = H_{\min}(z)H_{ap}(z) \quad (3.1)$$

여기서 $H_{\min}(z)$ 은 pole 과 zero 가 모두 단위원안에 있는 최소위상 시스템이며 $H_{ap}(z)$ 은 크기 응답이 1 인 all-pass 시스템이다. Figure 4 에 보여지듯이 머리전달함수의 임펄스 응답과 이의 최소위상판은 첫 번째 peak 의 모양이 다르고 바로 이어지는 ridge 의 깊이도 다르다. 그러나 이러한 외이 응답이 수평면상에서의 방향성 인지에 큰 영향을 주지 않기 때문에 이들 머리전달함수의 최소위상판에 Ray Tracing Formula(9)에 의거하여 청취자의 머리크기에 맞는 ITD 를 계산하여 다음 식에서와 같이 곱해주면 기존의 수평면 머리전달함수에 비해 개개인에게 더 잘 맞는 맞춤형 수평면 머리전달함수를 구할 수 있다.

$$H_{\text{custom}}(z) = H_{\min}(z) \cdot e^{i\omega d/c} \quad (3.2)$$

여기서 d 는 RTF 를 사용하여 구한 양 귀간의 ITD 이다(Figure 5). 또한 이렇게 얻어진 수평면 머리전달함수는 그 역필터도 마찬가지로 안정하기 때문에 1 장에서 언급된 virtualizer 의 필수 요소인 cross-talk cancellation filter 를 보다 손쉽게 구현하는데 쓰일 수도 있다.

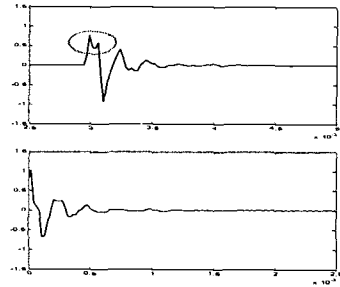


Figure 4: Cleansed left ear HRIR at 0° azimuth (top panel) and its minimum-phase reconstruction (bottom panel).

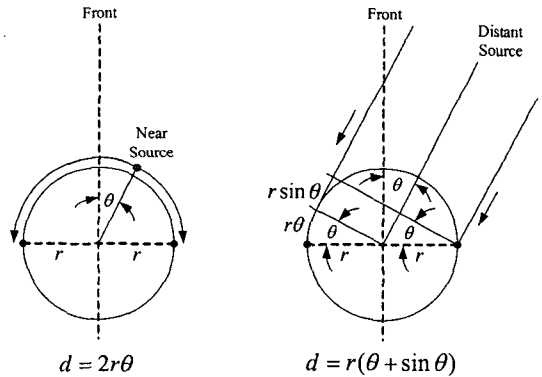


Figure 5: Ray Tracing Formula for nearby sources (left) and far-field sources (right). The corresponding ITD is d .

3.3 Pole-Zero Modeling

더미헤드의 머리전달함수를 아래와 같이 p 개의 pole 과 q 개의 zero 를 갖는 LTI(Linear Time Invariant) system 으로 모델링하여(6) 음원의 위치에 따라 이들이 변하는 정도를 살펴봄으로써 개개인에게 더 맞는 맞춤형 머리전달함수 개발의 실마리를 얻을 수 있다.

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^q b_q(k)z^{-k}}{1 + \sum_{k=1}^p a_p(k)z^{-k}} \quad (3.3)$$

4. 결 론

수 차례에 걸쳐 모델링을 수행해본 결과, 외이와 머리 그리고 어깨에 의한 응답을 정확히 모델링하려면 최소한 20 개의 pole 이 필요하고 외이에 의한 응답까지만 모델링하는데에는 단 10 개의 pole 이 필요하며 zero 의 개수는 많고 적음의 차이가 크지 않음을 알 수 있었다. Figure 6 에서 볼 수 있듯이 수평면 머리전달함수의 최소위상판을 각각 20 개와 5 개의 complex conjugate pole 과 zero 로 모델링하면 각 방위각마다 512 개에 달하는 임펄스 응답을 13 개의 파라미터로 정확히 모델링할 수 있다.

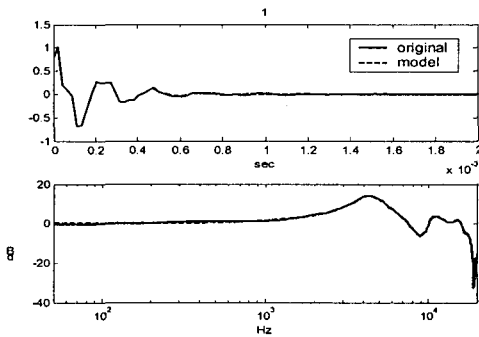


Figure 6: Horizontal min. phase HRTF modeling result ($p=20, q=5, \text{azimuth}=0^\circ, \text{left ear}$).

이에 반해 중앙면상에서의 머리전달함수는 외이의 응답이 가장 중요하고 나머지 머리나 어깨에 의한 응답이 덜 중요하므로 아래의 그림에서처럼 10 개의 pole 과 5 개의 zero 만으로 모델링 가능하다. 모델링하기에 앞서 임펄스 응답의 시작점까지의 delay 를 모두 없애주어야 낮은 차수로 모델링이 가능함을 명시해 둔다.

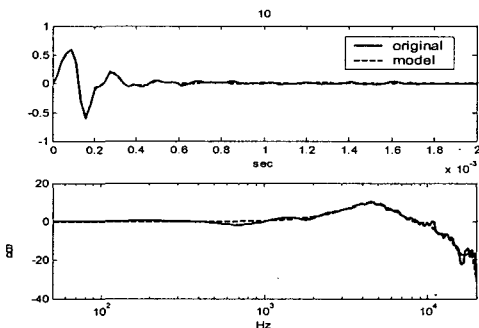


Figure 7: Median HRTF modeling result ($p=10, q=5, \text{azimuth}=0^\circ, \text{elevation}=60^\circ, \text{left ear}$).

끝으로 Figure 8 에서 볼 수 있듯이 중앙면상의 머리전달함수는 음원의 고도가 바뀌에 따라 pole 과 zero 의 위치가 화살표방향으로 점차 옮겨가는 것을

관찰할 수 있다. 이러한 사실을 토대로 향후 개개인의 머리와 외이에 맞는 맞춤형 머리전달함수를 개발할 수 있을 것이다.

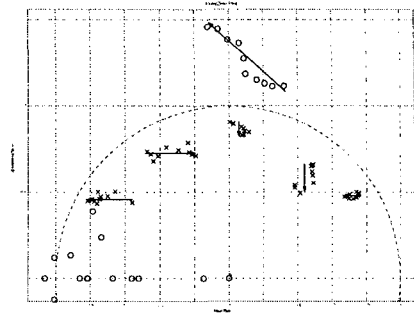


Figure 8: Pole-zero plot of median HRTFs. Arrows indicate the direction of elevation change ($\text{azimuth}=0^\circ, \text{elevation}=0\sim 90^\circ$).

후 기

본 연구는 NRL 사업과 BK21 사업의 후원으로 이루어졌음을 밝혀둔다.

참고문헌

- (1) 신 기훈, 박 영진, 2005, "Development of a quantitative evaluation technique to assess virtual audio systems on their source positioning capability," *Internoise*.
- (2) Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., 1993, "Localization using nonindividualized head-related transfer functions," *Journal of the Acoustical Society of America* 94(1), pp. 111~123.
- (3) Kahana, Y., Nelson, P. A., Petyt, M., and Choi, S., 1999, "Numerical modeling of the transfer functions of a dummy-head and of the external ear," *AES 16th Int. Conf. Spatial Sound Reproduction*.
- (4) Middlebrooks, J. C., 1999, "Virtual localization improved by scaling nonindividualized external-ear functions in frequency," *Journal of Acoustical Society of America* 106(3), pp. 1493 ~ 1509.
- (5) Zotkin, D. N., Duraiswami, R., Davis, L., Mohan, A., and Raykar, V. C., 2002, "Virtual audio system customization using visual matching of ear parameters," *Proceeding of the 2002 Int. Conf. on Pattern Recognition*, pp. 1003~1006.
- (6) Hayes, M. H., 1996, *Statistical Digital Signal Processing and Modeling* (John Wiley & Sons, INC., NY).
- (7) 신 기훈, 박 영진, 2004, "가상음장 재현을 위한 근거리 머리전달함수 측정 및 해석," *한국음향학회 논문집* 23(2), pp. 335~338.
- (8) Oppenheim, A. V., 1989, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
- (9) Woodworth, R. S. and Schlosberg, G., 1962, *Experimental Psychology* (Holt, Rinehard and Winston, NY).