

새로운 학습 하이브리드 실내 충격 응답 모델[#] New Learning Hybrid Model for Room Impulse Response Functions

신 민 철* · 왕 세 명†

Mincheol Shin and Semyung Wang

(2007년 12월 4일 접수 ; 2008년 1월 30일 심사완료)

Key Words : Room Impulse Response(실내 충격 응답), Learning Hybrid Model(학습 하이브리드 모델)

ABSTRACT

Many trials have been used to model room impulse responses, all attempting to provide efficient representations of room acoustics. The traditional model designs for room impulse response seem to fail in accuracy, controllability, or computational efficiency. In the time domain, room impulse responses are generally considered as combination of the three parts having different acoustic characteristics, initial time delay, early reflection, and late reverberation. This paper introduces new learning hybrid model for room impulse responses. In this proposed model, those three parts are modeled using different models with learning algorithms that determine the boundary of each model in the hybrid model. By the simulation with measured room impulse responses, the performance of proposed model shows the best efficiency in views of computational burden and modeling error.

기 호 설 명

$h[n]$: 실내 충격 응답 함수
 $h'[n]$: 실내 충격 응답 함수의 위치 민감도

1. 서 론

효과적인 실내 음향 특성의 표현을 위해 정확한 실내 충격 응답(room impulse response, RIR) 함수를 얻어내고자 많은 방법들이 시도되었으며 이는 실내 음향 연구 분야에 많이 이용되었다. 측정을 통해

얻어진 RIR 함수는 적당한 모델로 대체되어 음향 신호 처리 과정을 통해 3차원 게임, 실시간 입체 음향 등의 분야에 응용되고 있다. 또한 음향 공간의 RIR 함수를 실시간으로 예측하는 것을 주요 기술로 사용하는 잔향 제거기, 음향 피드백 제거기와 같은 응용 분야에서는 측정된 RIR 함수를 어떤 모델로 대체하느냐가 기기의 성능을 좌우한다.

현재까지 대표적으로 쓰이고 있는 RIR 함수의 모델은 자기회귀(auto-regressive, AR) 모델과 이동평균(moving average, MA) 모델, 그리고 둘을 복합한 모델(auto-regressive moving average, ARMA)을 대표적으로 사용해 왔다. 또한 통계적인 음향 모델⁽¹⁾도 소개 되었고 협대역 필터를 이용한 방법(subband technique)⁽²⁾, 부분 인공 재생 방법(partial artificial reproduction)⁽³⁾ 등, 원래 길이의 RIR 함수를 신호처리가 용이하도록 작은 신호 블록으로 나누어 처리하는 기술들이 개발 되었으며 최근

† 교신저자 : 정희원, 광주과학기술원 기전공학과
E-mail : smwang@gist.ac.kr
Tel : (062)970-2390, Fax : (062)970-3263

* 정희원, 광주과학기술원 기전공학과

이 논문은 2007 추계 소음진동 학술대회에서 우수논문으로 추천되었음.

에는 주파수 확대 모델링 기법⁽⁴⁾ 등 많은 연구들이 진행되어 왔다. 하지만 현재까지의 대표적인 모델링 기법들은 만족할 만한 정확도를 갖게 되면 이를 처리하는 계산량이 많아 실제 응용분야에 적용하는데 한계가 있고 계산량이 적은 모델의 경우에는 정확도가 문제가 된다. 그러므로 RIR 함수를 정확하게 표현할 수 있으며 처리해야 할 계산량을 기존의 모델과 비교하여 줄일 수 있는 모델을 개발하는 것이 필요하다.

이 논문에서는 계산량을 줄이며 만족할 만한 정확도를 유지시키기 위한 방법으로 학습 하이브리드 모델을 제안하였다. 일반적으로 RIR 함수는 시간 영역의 음향학적인 관찰을 통해서 세 개의 구간으로 나눌 수 있으며 이를 초기지연 구간, 초기반사 구간, 후기잔향 구간으로 명명하였다. 각각의 구간은 뚜렷이 구별되는 특성이 있으며 이를 효과적으로 표현하기 위해서 이 논문에서는 각 구간별 모델을 달리하는 하이브리드 모델을 제안하였다. 이때 실시간으로 변하는 각 구간의 경계는 학습 알고리즘을 이용하여 정의하였다. 이와 같이 제안된 모델은 기존의 RIR 함수 모델들과 비교한 결과, 정확도와 계산량 감소 측면의 평가 항목에서 동시에 향상된 결과를 얻을 수 있었다.

이 논문의 2장에서는 실내 충격 응답의 특성에 대해 고찰하였고 3장은 이 논문을 통해 제안된 학습 하이브리드 실내 충격 응답 모델에 대해 소개하였다. 4장에서는 제안된 새로운 실내 충격 응답 함수 모델을 기존에 많이 쓰이던 전통적인 모델과 비교하여 성능 평가를 실시하였다. 마지막으로 5장을 통해 이 논문을 마무리 하였다.

2. 실내 충격 응답의 특성

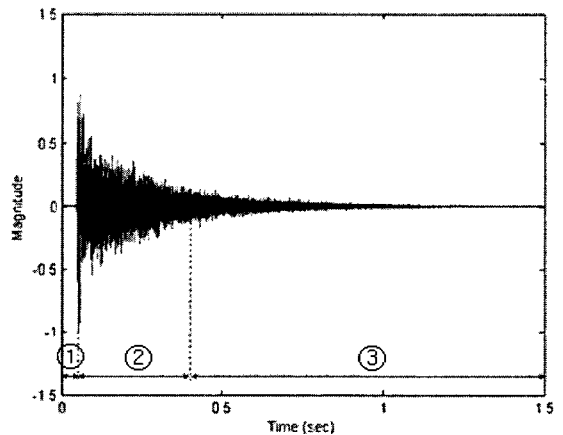
이 장에서는 실내 충격 응답 함수의 시간 및 주파수 영역 특성을 한 예를 통하여 알아보려고 한다. 뉴 멕시코 주립 대학에서 48 kHz의 표본화율로 측정된 RIR 함수를 Fig. 1에서 시간 영역과 주파수 영역으로 표현하고 있다⁽⁵⁾.

Fig. 1의 (a)는 RIR의 시간 영역에서의 표현을 나타내고 있다. 시간 영역 RIR에서는 현저히 다른 시간 영역 음향 특성과 함께 ① 초기지연 구간, ② 초기반사 구간, ③ 후기잔향 구간으로 각각 구분된다.

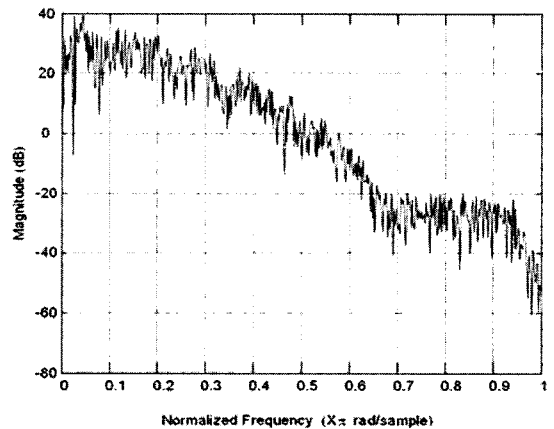
여기서 초기지연 구간은 직접 경로 즉, 음원과 수음점 사이의 거리에 비례하므로 굉장히 작은 값을 가지게 된다. 이에 반해 초기반사 구간은 실내의 음향 반사 특성을 민감하게 반영하므로 대부분의 에너지를 이 구간에 포함한다. 이로 인한 갑작스런 RIR 함수값의 변화로 두 구간은 확연히 구분된다. 또한 시각적으로 구분이 용이하지 않으나 음원과 수음점의 위치변화에 민감한 부분과 위치변화보다는 실내 공간의 음향 특성에 더 많은 영향을 받는 부분으로 나뉘어지며 각각 초기반사, 후기잔향 구간이라고 명명한다.

반면 주파수 영역의 표현은 (b)에서와 같이 분리 기준이 모호하고 주파수 영역별로 음향학적 의미를 내포하고 있지 못하다.

결과적으로 RIR은 시간 영역 표현에서 확연하게



(a) Time-domain representation



(b) Frequency-domain representation

Fig. 1 An example of room impulse response

구분된 영역을 음향학적 특성을 고려하여 각기 다른 모델을 적용하여 모델링하는 것이 효과적이다.

3. 학습 하이브리드 실내 충격 응답 모델

이 장은 앞 장에서 언급한 RIR의 시간영역 음향학적 특성을 고려하여 초기지연, 조기반사, 후기잔향의 세 개의 구간으로 나누어 효과적으로 모델링하기 위한 방법을 제안하였다. 각 구간의 특성을 잘 반영할 수 있는 각기 다른 모델을 적용하였다. 이는 다음의 Fig. 2에 묘사된 것과 같이 초기지연 구간에는 시간 지연 모델(time delay model), 초기반사 구간에는 MA (all-zero) 모델, 그리고 후기잔향 구간에는 AR (all-pole) 혹은 ARMA(pole-zero) 모델을 적용하였다. 이 모델에 시간에 따라 변하는 각 구간의 경계를 정의하는 학습 알고리즘을 구현하여 이를 학습 하이브리드 (learning hybrid) 모델이라고 명명하였다.

3.1 하이브리드 실내 충격 응답 모델

언급한 바와 같이 시간영역 RIR의 세 부분의 음향학적 특성은 확연히 구분되어 있다. 초기지연 구간은 음원과 수음점 사이의 거리를 나타내고 시간 지연 이외의 정보를 가지고 있지 않다. 초기반사 구간의 경우 음원과 수음점의 위치 변화에 굉장히 민감한 부분이며 후기잔향 구간의 경우는 공간 자체의 음향학적 특성을 잘 반영하며 위치의 변화에는 그리 민감하지 않다.

이와 같은 각 부분의 특성을 반영하여, 초기지연 부분에 대해서는 다음의 식(1)과 같은 시간 지연 모델을 적용하였다. 여기서 시간 지연은 음원과 수음점 사이의 거리에 비례한다.

$$H(z) = z^{-D} \tag{1}$$

여기서 D 는 직접 경로의 시간 지연에 해당하는 샘플의 개수를 의미한다.

조기반사 구간은 음원에서 수음점까지 직접 경로를 포함하는 간접 반사 경로로 구성되어 있다. 이러한 반사 경로들은 음원과 수음점의 위치 변화에 상

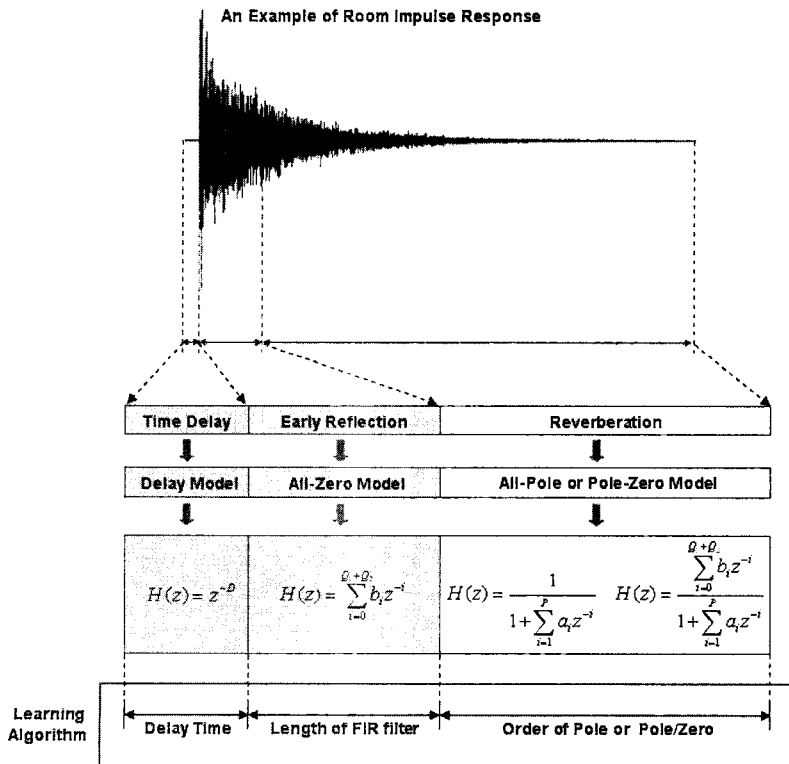


Fig. 2 Proposed learning hybrid model for RIR

당히 민감하다. 그러므로 이러한 음향학적 특성 때문에 이러한 변화를 민감하게 반영할 수 있는 식 (2)와 같은 MA(All-zero) 모델을 적용하였다.

$$H(z) = Cz^{-Q_1} \prod_{i=1}^{Q_2} (1 - q_i z^{-1}) = \sum_{i=0}^{Q_1+Q_2} b_i z^{-i} \quad (2)$$

여기서 C 는 상수이며, Q_1 과 Q_2 는 All-zero 모델의 차수를 나타낸다. q_i 는 제로의 값을, b_i 는 MA 모델의 계수값을 나타낸다.

후기잔향 구간은 초기반사 구간과 달리 전체적인 실내 음향 특성을 잘 나타낸다. 이러한 음향학적인 특성 때문에 후기잔향 구간을 효과적으로 표현하기 위해서는 음원과 수음점의 위치에 민감한 모델보다 이러한 위치 변화에 강인하면서 전체적인 특징을 잘 나타내는 모델이 적당하다. 결과적으로 식 (3)으로 표현되는 AR (All-pole) 혹은 식 (4)의 ARMA (pole-zero) 모델이 이러한 특성을 효과적으로 반영한다.

$$H(z) = \frac{C}{\prod_{i=1}^P (1 - p_i z^{-1})} = \frac{1}{1 + \sum_{i=1}^P a_i z^{-i}} \quad (3)$$

$$H(z) = Cz^{-Q_1} \frac{\prod_{i=1}^{Q_2} (1 - q_i z^{-1})}{\prod_{i=1}^P (1 - p_i z^{-1})} = \frac{\sum_{i=0}^{Q_1+Q_2} b_i z^{-i}}{1 + \sum_{i=1}^P a_i z^{-i}} \quad (4)$$

여기서 P 는 All-pole 모델의 차수를 나타내며 p_i

는 폴의 값을, a_i 는 AR 모델의 계수값을 나타낸다. 이 논문에서는 식 (4)의 ARMA 모델을 적용하였다.

3.2 학습 알고리즘

음원과 수음점의 움직임 또는 내부 구조물이나 사람의 움직임에 따라 RIR 함수 역시 변하게 된다. 이렇게 시간에 따른 환경 변화에 따라 변하는 RIR을 앞서 언급한 하이브리드 실내 충격 응답 모델로 표현하기 위해서는 초기지연과 초기반사 구간, 초기반사와 후기잔향 구간의 경계에 대한 명확한 정의가 필요하다. 이를 위해 학습 알고리즘이 도입되었다.

(1) 초기지연 구간과 초기반사 구간의 경계

RIR 함수의 초기지연 구간과 초기반사 구간 사이의 경계는 두 구간 사이를 지날 때 발생하는 RIR 함수값의 급격한 변화로 쉽게 찾아낼 수 있다. 이는 Fig. 3에 나타낸 바와 같이 적당한 문턱값(threshold, THD)을 설정하여 RIR함수, h 와 샘플간 비교를 통해 초기지연과 초기반사 구간 사이의 경계를 용이하게 찾아낼 수 있다. 이 논문의 THD는 0.005로 하였다.

(2) 초기반사 구간과 후기잔향 구간의 경계

초기반사 구간은 앞서 언급한 바와 같이 음원과 수음점의 위치 변화에 가장 민감한 부분이고 후기잔향 구간은 전체적인 실내 음향 특성을 반영하며 음원과 수음점의 위치 변화에 덜 민감한 부분이다. 그러므로 위치 변화에 따른 민감도를 정의하면 두 구간의 경계를 찾는 데 효과적이다. 일반적인 구조 해

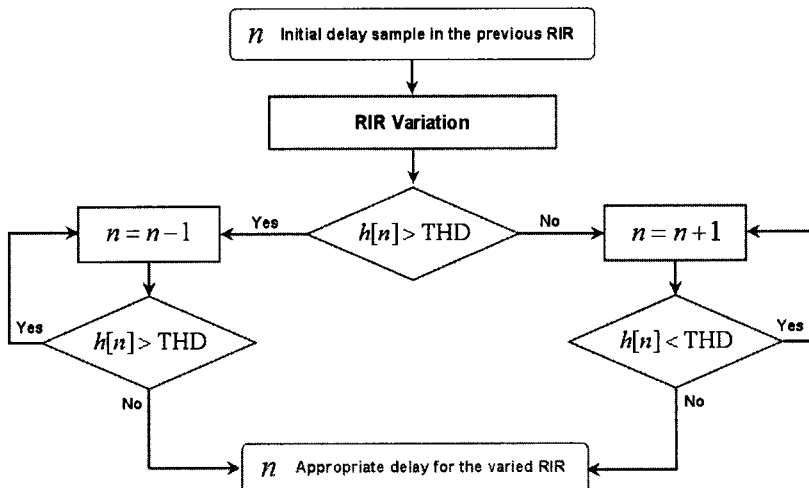


Fig. 3 Learning algorithm I for determining boundary between initial delay and early reflection

석 분야에서 민감도는 식 (5)와 같이 정의된다.

$$\psi' = \frac{\psi(r_0 + \Delta r) - \psi(r_0)}{\Delta r}, \quad (5)$$

여기서 ψ' 는 구조의 민감도, ψ 는 주어진 구조의 함수, r_0 와 Δr 은 각각 목적 변수의 초기 상태와 단위 변화량을 나타낸다. 이러한 개념을 이용하여 목적 변수를 음원과 수음점의 위치 변화로 하여 실내 충격 응답 함수에 적용하기 위해 Fig. 4와 같이 단위 위치 변화량을 정의하였다.

Fig. 4를 통해 정의된 단위 위치 변화량, Δl 을 이용하여 RIR의 위치 변화에 따른 민감도, h' 을 식 (6)과 같이 정의한다.

$$h'[n] = \frac{|h_{l_0+\Delta l}[n] - h_{l_0}[n]|}{\Delta l}, \quad (6)$$

여기서 h_{l_0} 와 $h_{l_0+\Delta l}$ 은 각각 초기 위치의 RIR과 단위 위치를 변화시킨 후 RIR를 의미한다. 이렇게 얻어낸 민감도를 각기 다른 음향 특성을 갖는 실내 공간에 적용하기 위해서 다음의 식 (7)과 같이 정규화된 민감도를 제안하였다.

$$h'_N[n] = \frac{h'[n]}{\max(h'[n])} \times 100 (\%) \quad (7)$$

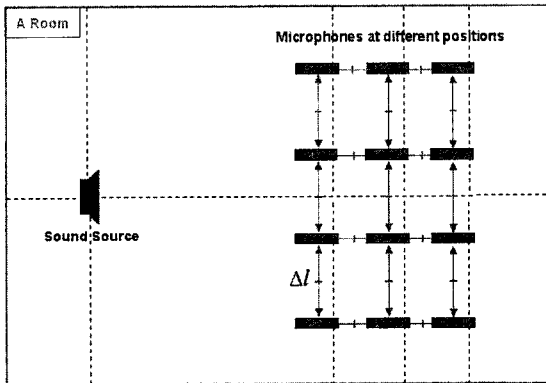


Fig. 4 Fixed sound source and microphones at various locations with unit location difference, Δl

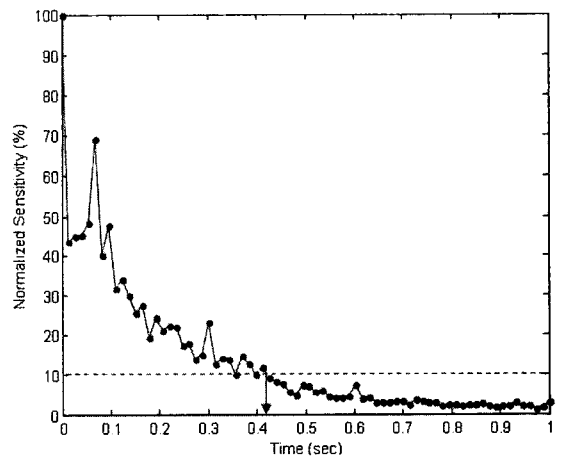


Fig. 5 Normalized sensitivity plot of a RIR with the change of unit location

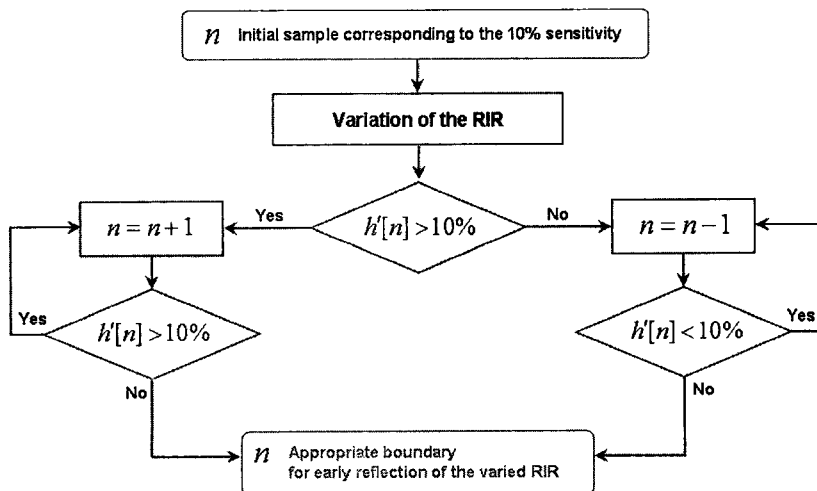


Fig. 6 Learning algorithm II for determining boundary between early reflection and late reverberation

Fig. 5는 단위 위치 변화, Δl 에 따른 RIR의 정규화된 민감도의 한 예를 보여주고 있다. 이 그림은 초기지연 구간을 제외한 부분을 나타낸 것으로 점선은 민감도가 10%일 때를 표시한 것이다. 정규화된 민감도를 이용하여 조기반사 구간과 후기잔향 구간을 나누는 문턱값은 응용 분야에 따라 그리고 목적에 따라 제어할 수 있으며, 이 논문에서는 편의상 민감도가 10%인 지점을 기점으로 이보다 큰 곳을 조기반사 구간으로 이보다 작은 구간을 후기잔향 구간으로 정의하였다.

Fig. 6은 조기반사 구간과 후기잔향 구간의 경계를 정의하는 학습 알고리즘을 소개하였다. 이는 학습 알고리즘 I과 비슷하지만 RIR의 함수값인 h 대신 RIR의 단위 위치에 따른 민감도, h' 을 미리 정의한 경계값인 10%와 비교하여 시간에 따라 변하는 RIR의 조기반사 구간과 후기잔향 구간의 경계를 찾도록 하였다.

4. 제안된 모델의 성능 평가

이 논문에서 제안된 학습 하이브리드 RIR 모델의 성능을 평가하기 위해 다음의 Fig. 7과 같이 실제 RIR을 광주과학기술원(GIST) 대강당에서 측정하였다.

Fig. 7을 통해 측정 시 실내음향 공간의 크기 및 음원과 마이크론의 위치를 묘사하였고 단위 위치 변화량, Δl 은 2미터로 하였다. 또한, 성능을 평가의

척도로서 실제 측정된 RIR과 예측된 RIR의 오차를 출력 신호간의 차이를 이용하여 검출해 내며, 모델의 정확도를 나타내는데 많이 사용되는 모델링 에러(modeling error, J)를 선정하였다.

Fig. 8과 같은 단일 채널 음향 시스템을 가정하면 모델링 에러, $J(\%)$ 는 식 (8)과 같이 정의된다.

$$J(\%) = \frac{E[e^2(t)]}{E[d^2(t)]} \times 100, \quad (8)$$

입력신호를 $x(t)$ 라고 가정하면 $d(t) = h(t) * x(t)$, $e(t) = \hat{h}(t) * x(t) - d(t)$ 이며 $h(t)$ 는 측정된 RIR을, $\hat{h}(t)$ 는 특정 모델을 이용하여 표현한 RIR을 나타낸다. 여기서 *는 컨볼루션 연산을 의미한다. Fig. 7의 공간에서 측정된 RIR을 기존의 모델들(MA, AR, ARMA)을 이용하여 추정된 RIR의 모델링 에러와 이때 각각의 모델들에 사용된 가변 파라미터의 개수를 제안된 학습 하이브리드 모델(learning hybrid, LH)과

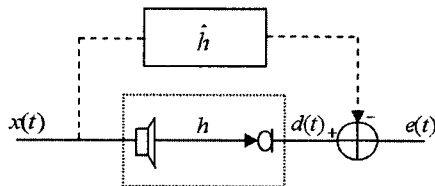


Fig. 8 Single channel acoustic system and estimated model

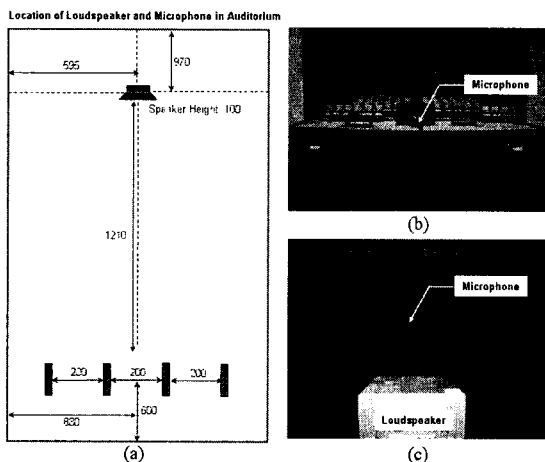


Fig. 7 RIR measurement setup in GIST auditorium; (a) measurement schematics, (b) front view, (c) rear view

Table 1 Number of variable parameters and modeling error for different RIR models (MA: number of MA coefficients, AR: number of AR coefficients, Delay: number of delay coefficients, MA1: number of MA coefficients in early reflection part of hybrid model, MA2: number of MA coefficients in late reverberation part of hybrid model)

RIR model	Variable parameters		Modeling error (%)
	Total	Detail	
MA	4000	MA: 4000	0.1
AR	3000	AR: 3000	99.01
ARMA	3000	AR: 1500, MA: 1500	52.37
LH	2626	Delay: 1, MA1: 1745, AR: 430 MA2: 450	3.28

비교한 결과를 Table 1에 나타내었다. 이때 사용된 각 모델의 파라미터의 개수는 모델링 에러를 최소화할 수 있는 최소의 개수이며 이보다 늘렸을 경우 모델링 에러가 개선되지 않았다. 각각의 모델이 실시간 응용 분야에 사용되었을 때 업데이트가 필요한 가변 파라미터의 개수가 전체 계산량에 미치는 영향이 지배적이기 때문에 이 논문에서는 가변 파라미터의 개수를 성능 평가에 사용하였다.

Table 1에 나타난 바와 같이 MA 모델의 경우 가장 작은 모델링 에러를 갖지만 또한 많은 가변 파라미터를 갖기 때문에 이를 처리하는 과정에서 계산량이 상당히 증가하게 된다. AR과 ARMA 모델의 경우 3000개의 상당한 가변 파라미터를 사용했음에도 불구하고 모델링 에러는 만족할만한 수준이 되지 못하였다. 하지만 이 논문에서 제안된 학습 하이브리드 모델의 경우 가장 적은 개수의 가변 파라미터를 가지며 모델링 에러 또한 비교적 작았다. 결과적으로 가변 파라미터의 개수와 모델링 에러를 동시에 고려한다면 제안된 학습 하이브리드 모델이 가장 효과적으로 RIR을 표현한다고 할 수 있다.

5. 결 론

이 논문에서는 학습 알고리즘을 이용한 새로운 하이브리드 RIR 모델을 소개하였다. RIR은 각각 다른 음향 특성으로 인해 초기지연, 조기반사, 후기잔향의 세 구간으로 나눌 수 있다. 각각의 구간은 서로 다른 모델을 이용하여 표현되었으며 시간에 따라 변하는 각 구간의 경계는 학습 알고리즘을 통해 정의되었다. 기존에 많이 사용되던 RIR 모델들(MA, AR, ARMA)과 제안된 학습 하이브리드 모델의 계산량을 결정하는 가변 파라미터의 개수와 모델의 정확도를 나타내는 모델링 에러가 실제로 측정된 RIR을 이용하여 비교되었다. 모델링 에러의 측면에서는 가장 많은 가변 파라미터의 개수에도 불구하고 MA 모

델이 가장 정확한 모델로 판명되었다. 또한 AR과 ARMA 모델의 경우 현저하게 큰 모델링 에러로 인해 RIR전체를 표현하는데 적합하지 않은 모델로 판단되었다. 제안된 학습 하이브리드 모델은 가변 파라미터의 개수로 대변되는 처리 시의 계산량과 모델링 에러를 동시에 고려할 때 가장 효과적인 모델이었다. 또한 이 모델은 각 구간의 경계를 정의하기 위해 사용된 학습 알고리즘의 문턱값과 민감도를 조절하면 계산량과 정확도를 응용분야에 따라 적절히 조절할 수 있는 장점도 가지고 있다.

참 고 문 헌

- (1) Martin, J., Van Maercke, D. and Vian, J. P., 1993, "Binaural Simulation of Concert Halls: A New Approach for the Binaural Reverberation Process", *Journal of the Acoustical Society of America*, Vol. 94, No. 6, pp. 3255~3264.
- (2) Laroche, J., 1989, "A New Analysis/synthesis System of Musical Signals Using Prony's Method - Application to Heavily Damped Percussive Sounds", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing* 89, pp. 2053~2056.
- (3) Jot, J. M. and Chaigne, A., 1991, "Digital Delay Networks for Designing Artificial Reverberators", in *Proc. 90th Audio Engineering Society Convention, Paris, France*, Preprint 3030.
- (4) Karjalainen, M., Esquef, P. A. A., Antsalo, P., Akivirta, A. M. and Valimaki, V., 2002, "Frequency-zooming ARMA Modeling of Resonant and Reverberant Systems", *Journal of Audio Engineering Society*, Vol. 50, No. 12, pp. 1012~ 1029.
- (5) http://www.ece.nmsu.edu/~pdeleon/BSS/Impulse_Responses/README_Impulse_Responses.pdf.