

잔향 생성기에서 심리 음향 필터를 이용한 고속 컨벌루션 방법

Fast Convolution Method using Psycho-acoustic Filters in Sound Reverberator

신민철*· 왕세명†

Mincheol Shin and Semyung Wang

Key Words : Fast Convolution (고속 컨벌루션), Psycho-acoustic Filter (심리 음향 필터), Sound Reverberator (잔향 생성기)

ABSTRACT

With the advent of sound field simulator, many sound fields have been reproduced by obtaining the impulse responses of specific acoustic spaces like famous concert hall, opera house. This sound field reproduction has been done by the linear convolution operation between the sound input signal and the impulse response of certain acoustic space. However, the conventional finite impulse response based linear convolution operation always makes real-time implementation of sound field generator impossible due to the large amount of computational burden. This paper introduces the fast convolution method using perceptual redundancy in the processed signals, input audio signal and room impulse response. Temporal and spectral psycho-acoustic filters considering masking effects are implemented in the proposed convolution structure. It reduces the computational burden of convolution methods for real-time implementation of a sound field generator. The conventional convolutions are compared with the proposed one in views of computational burden and sound quality. In the proposed method, a considerable reduction in the computational burden was realized with acceptable changes in sound quality.

———— 기호설명 —————

$x[n]$: Input signal, $h[n]$: RIR, $y[n]$: Output signal

1. 서 론

특정한 음향 공간들, 예를 들면 유명한 콘서트 홀이나 오페라 하우스의 음향 충격 응답함수(Room Impulse Response, RIR)를 측정 한 뒤 이를 음원에 적용하여 음장을 재현하려는 많은 시도가 있었으며 이는 음장 시뮬레이터 또는 잔향 생성기로 구현되었다. 잔향 생성기는 대표적인 두 가지 방법으로 설계하는데, IIR (Infinite Impulse Response) 필터 구조 또는 FIR (Finite Impulse Response) 필터 구조를 이용하는 방법이 있다. IIR 필터 구조를 활용하는 방법은 Comb 필터 또는 All-pass 필터와 같은 회귀적 네트워크를 사용하는 방법으로 비교적 적은 수의 가변 파라미터를 가지며 신호 처리 시에 적은 계산량을 장점으로 한다. 이와 같은 방법은 기존의 전통적인 Schroeder 의 방법 이래로 많은 알고리즘이 개발되었다 [1, 2, 3]. 하지만 이와 같은 IIR-based 방법들은 민감한 인간의 청각 특성을 제대로 반영하지 못하며 이에 따라 음질이 부자연스럽게 들리며 원음 또한 왜곡시키는 경향이 있다. 이러한 IIR 필터 구조를 기초로 하는 잔향 생성기의 단점은 FIR 필터 구조로 극복할 수 있다. FIR 필터 구조의 설계는 입력 신호와 충격 응답 신호의 직접적인 컨벌루션으로 인해 음질의 왜

곡이 없다. 하지만 이 방법은 어떤 특정한 음향 공간의 실내 충격 응답의 길이에 비례한 가변 파라미터가 증가로 인해 컨벌루션 시 많은 양의 계산이 필요하게 된다. 이로 인해 FIR 필터 구조를 기초하는 잔향 생성기는 실시간 연산을 필요로 하는 응용 분야에 사용하기에 많은 어려움이 있다. 이를 해결하기 위해 많은 연구자들은 FIR 필터 구조를 기초로 하는 컨벌루션 연산의 계산량을 감소시키기 위한 시도를 계속해 왔다. 그 중 대상 신호들을 작은 여러 개의 블록으로 나누어 주파수 영역에서 처리한 블록 컨벌루션 방법은 계산량을 현저히 감소시켜 컨벌루션 알고리즘의 실시간 처리의 가능성을 보여주었다 [4, 5]. 또한 청각의 시간과 주파수 영역에서의 특성을 이용하여 계산량을 줄이려는 시도가 있었다 [6, 7].

본 논문에서는 잔향 생성기를 위한 FIR 필터 구조를 기초로 하는 고속 컨벌루션 방법을 소개하였다. 제안된 방법은 기존의 블록 컨벌루션 알고리즘에 인간의 청각 특성을 고려한 시간 영역, 주파수 영역심리 음향 필터를 설계하여 신호 블록에 적용하였다. 기존의 컨벌루션 방법과 계산량, 음질 두 가지 항목에서 성능을 비교해 본 결과 제안된 방법이 비교적 음질 왜곡이 없었으며 상당한 계산량의 감소를 보였다.

본 논문의 2 장에서는 기존의 블록 컨벌루션 방법에 대해 고찰하였고 3 장은 본 논문을 통해 제안된 고속 컨벌루션 방법을 소개하였다. 4 장에서는 제안된 고속 컨벌루션 방법을 기존에 많이 쓰이던 전통적인 컨벌루션 방법들과 계산량, 음질의 항목에서 비교하여 성능을 평가하였다. 마지막으로 5 장을 통해 본 논문을 마무리 하였다.

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

* 광주과학기술원, 기전공학과

2. Conventional Block Convolution

기존의 시간 영역 선형 컨벌루션 방법의 많은 계산량과 주파수 영역 컨벌루션 방법의 처리 시간 지연의 문제를 해결하기 위한 방법으로 블록 컨벌루션 방법이 고안되었다. 이 방법은 컨벌루션의 대상이 되는 두 신호를 처리가 용이한 작은 블록으로 나누어 이를 주파수 영역의 곱셈과 결과 신호의 적절한 배열을 통해 컨벌루션을 구현한다. 컨벌루션 대상 신호를 각각 무한 입력 신호, $x[n]$ 과 유한 신호, $h[n]$ 로 하고 이를 수식으로 나타내면 다음과 같이 표현된다.

$$x[n] = \sum_{r=0}^{\infty} x_r[n - rN], h[n] = \sum_{s=0}^{M-1} h_s[n - sN], \quad (1)$$

$$\text{여기서 } x_r[n] = \begin{cases} x[n + rN] & 0 \leq n \leq N - 1 \\ 0 & \text{otherwise} \end{cases},$$

$$h_s[n] = \begin{cases} h[n + sN] & 0 \leq n \leq N - 1 \\ 0 & \text{otherwise} \end{cases}.$$

위의 수식으로 분리된 신호 블록을 이용하여 선형 컨벌루션의 결과 신호, $y[n]$ 을 다음 식 (2)로 쓸 수 있다.

$$y[n] = x[n] * h[n] = \left\{ \sum_{r=0}^{\infty} x_r[n - rN] \right\} * \left\{ \sum_{s=0}^{M-1} h_s[n - sN] \right\}$$

$$= \sum_{r=0}^{\infty} \sum_{s=0}^{M-1} y_{r,s}[n - rN - sN], \quad (2)$$

여기서 $y_{r,s}[n] = x_r[n] * h_s[n]$ $0 \leq n \leq 2N - 1$.

위의 식 (2)의 컨벌루션은 주파수 영역에서 곱셈으로 대체시킬 수 있으며 이는 p 번째 결과 신호 블록에서 다음의 식 (3)으로 나타낼 수 있다.

$$Y_p[k] = \sum X_{p-s}[k] \times H_s[k], \quad (3)$$

여기서 $p = r + s$ 이며, $Y[k]$, $X[k]$, $H[k]$ 는 각각 $y[n]$, $x[n]$, $h[n]$ 의 주파수 영역 표현이다. 이렇게 각 결과 신호 블록 안에서의 계산이 수행되고 난 후 이를 **Overlap-save** 방법으로 적절히 배열하면 컨벌루션 된 출력 신호를 얻게 된다. 이는 신호의 입력과 동시 실시간으로 이루어지는 컨벌루션으로 많은 응용 분야를 갖게 된다. 또한 그 결과는 시간 영역의 선형 컨벌루션의 그것과 일치한다. 블록 컨벌루션에서 시간 영역의 신호를 주파수 영역으로 바꿀 때, 주파수 영역의 결과 신호 블록을 시간 영역으로 바꿀 때, 계산량의 감소를 위해 각각 고속 푸리에 변환 (Fast Fourier Transform, FFT), 와 역 고속 푸리에 변환 (Inverse

Fast Fourier Transform, IFFT) 을 사용한다. 다음의 그림 1 은 본 장에서 설명한 블록 컨벌루션의 개략적 도식이다.

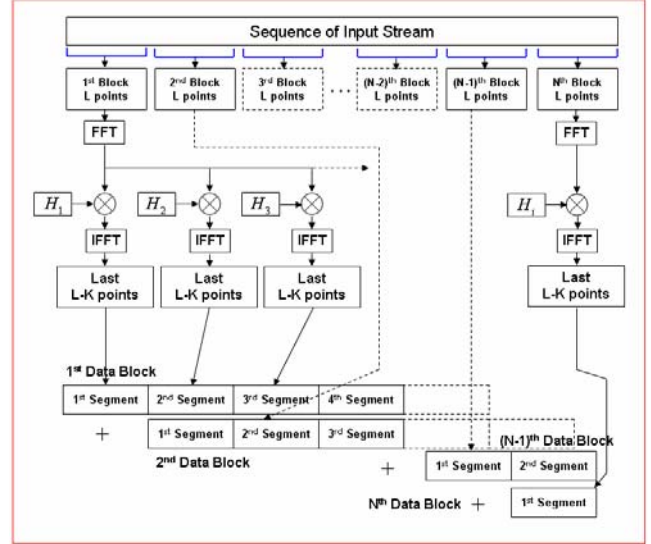


그림 1. Conventional Block Convolution

3. Proposed Fast Block Convolution

잔향 생성기에서 식 (1)의 $x[n]$ 은 음원 신호이고 $h[n]$ 은 실내 충격 응답 (Room Impulse Response, RIR) 이다. 그러므로 본 논문은 인간의 청감 특성, 즉 시간 영역과 주파수 영역의 마스킹 효과를 고려한 심리 음향 필터를 삽입하여 전체적인 계산량을 감소시키면서 결과 신호의 음질의 왜곡을 최소화하고자 하였다.

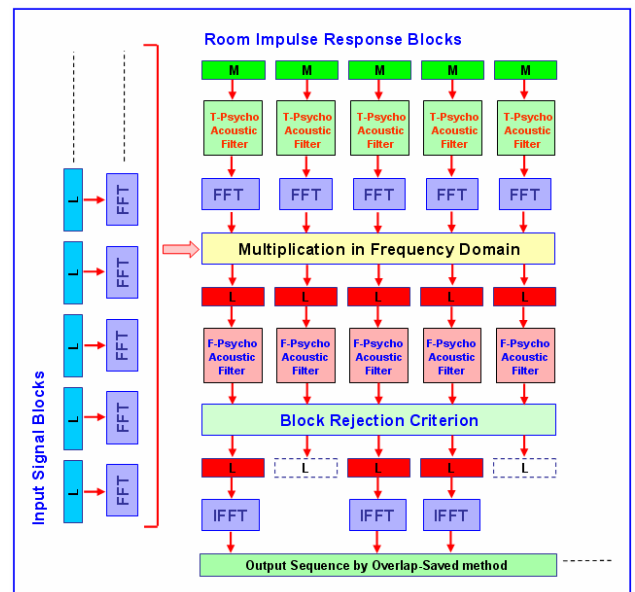


그림 2. 제안된 고속 컨벌루션 방법의 개략도

그림 2 는 본 논문에서 제안된 고속 컨벌루션 방법을 묘사하는 개략도이다. 이 방법은 기존의 블록 컨벌루션 방법의 장점을 그대로 유지하면서 음향 신호 자체가 가지는 청각 잉여분을 시간과 주파수 영역의 심리 음향 필터를 이용하여 제거하는 것을 설계의 주된 목표로 하고 있다. 그러므로 시간 영역과 주파수 영역의 심리 음향 필터의 설계와 이의 삽입 위치가 가장 중요하다. 다음은 전체적인 신호 처리의 과정을 요약하였다.

1. 입력 신호와 RIR 을 적당한 신호 블록으로 나눈다.
2. RIR 의 신호 블록을 시간 영역의 심리 음향 필터를 통과시켜 추후에 계산되는 데이터의 수를 감소시킨다.
3. 입력 신호 블록과 심리 음향 필터를 통과한 RIR 의 신호 블록들을 FFT 를 이용하여 주파수 영역으로 변환시킨다.
4. 각 블록에 해당하는 주파수 영역에서의 곱셈 연산을 실시한다.
5. 각 결과 신호 블록을 주파수 영역의 심리 음향 필터를 통과시킨다.
6. 주파수 영역의 심리 음향 필터를 통과 시킨 신호 블록에 블록 제거 알고리즘을 적용한다.
7. 제거되지 않은 결과 신호 블록을 IFFT 한 뒤 Overlap-save 방법을 이용하여 최종 결과 신호를 생성한다.

3.1 시간 영역 심리 음향 필터의 설계

실내 충격 응답 함수는 충격 신호(Impulse Signal)의 조합이다. 그러므로 일반적인 음향 신호에 대한 청각 특성이 아닌 충격 신호에 해당하는 것을 반영하여야 한다. 본 논문에서는 음향 충격 신호의 가장 대표적인 청각 특성으로 Post-Masking (PM) 효과와 Gap Detection Threshold (GDT) 에 주목하였다 [8].

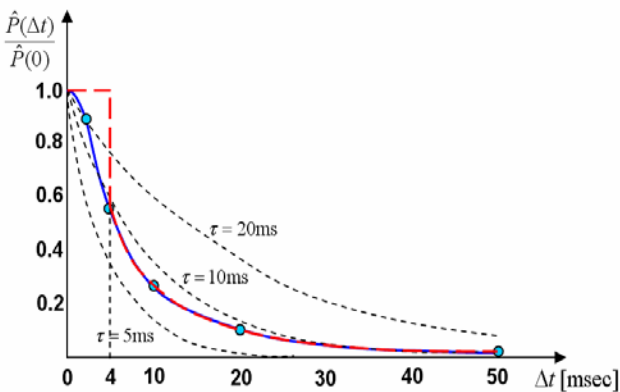


그림 3. 시간 영역 심리 음향 필터

그림 3 은 시간 영역의 심리 음향 필터를 나타내고 있다. 파란색 실선은 충격 신호의 PM 특성을 나타내고 있으며 이 때 \hat{P} 는 충격 신호의 Peak 음압을 나타낸다. GDT 는 대부분의 주파수 성분을 갖고 있는 충격 신호의 경우 0.005 초에 해당하며 이는 0.001 초의 여유분을 고려하여 그림의 초기 0.004 초의 영역으로 반영되었다. 결국 제안된 시간 영역의 심리 음향 필터는 빨간색 점선으로 표시하였으며 이는 초기 0.004 초의 GDT 와 다음 식 (4)로 근사화된 PM 을 동시에 반영하였다. 이 때 τ is 0.0075 로 정하였다.

$$a_{apx} = \exp\left(-\frac{t}{\tau}\right), \quad (4)$$

이 때 a_{apx} 는 근사화된 PM 을 나타내며 τ 는 시간 상수이다.

3.2 주파수 영역 심리 음향 필터의 설계

주파수 영역의 심리 음향 필터의 설계에는 Threshold in Quiet (TQ) 특성을 반영하였고 블록 제거 알고리즘을 고안하여 적용하였다.

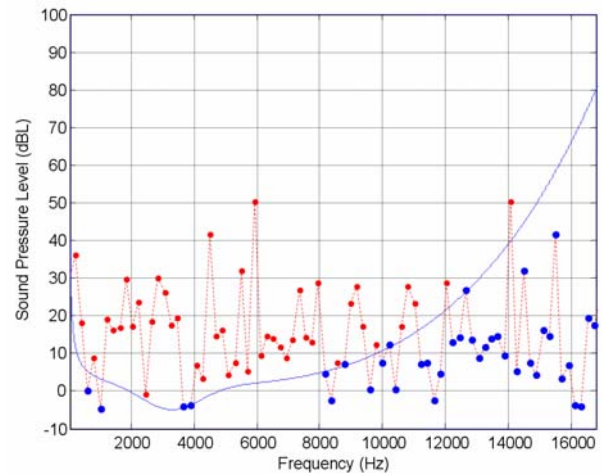


그림 4. Threshold in Quiet 곡선과 음향 신호

위의 그림 4 는 인간 청각의 주파수 영역 TQ 특성을 파란색 실선으로 나타내고 있고 임의의 음향 신호를 점선으로 나타내고 있다. TQ 곡선의 위쪽에 위치한 빨간색 점으로 나타낸 데이터는 사람들이 들 수 있는 데이터지만 파란색 데이터는 가청 영역을 벗어난 데이터이다. 이러한 특성은 다음의 식 (5)로 주파수 영역 심리 음향 필터 설계에 반영되었다.

$$\begin{cases} Y_p^{aud}[k] = Y_p[k] & Y_p[k] > T_q[k] \\ Y_p^{aud}[k] = 0 & Y_p[k] < T_q[k] \end{cases} \text{ [dB],} \quad (5)$$

여기서 $T_q(f)$ 는 TQ 곡선의 함수이며 Y_p 는 P 번째 결과 블록의 주파수 영역 데이터를 나타내며 Y_p^{aud} P 번째 결과 블록의 가청 데이터를 나타낸다. 다음의 식 (6)은 블록 제거 알고리즘을 소개하고 있다.

$$\begin{cases} Y_p^{out}[k] = Y_p^{aud}[k], & \frac{1}{N} \sum_{k=0}^{N-1} Y_p^{aud}[k] > \frac{1}{N} \sum_{k=0}^{N-1} BN[k] \\ Y_p^{out}[k] = 0, & \frac{1}{N} \sum_{k=0}^{N-1} Y_p^{aud}[k] < \frac{1}{N} \sum_{k=0}^{N-1} BN[k] \end{cases} \quad (6)$$

BN 은 결과 신호 블록과 같은 길이, N 을 갖는 배경 잡음 블록을 의미하고 Y_p^{out} 는 최종적으로 IFFT를 하기 위해 얻어진 P 번째 출력 신호 블록을 의미한다. 그러므로 식 (6)은 식 (5)로부터 얻어낸 Y_p^{aud} 를 같은 길이를 가진 BN 의 평균값과 비교하여 이보다 크면 최종 출력 신호 블록으로 선택되어 그 이후의 과정들이 수행된다. 선택되지 못한 블록들은 제거되어 다음의 신호 처리 과정에서 고려하지 않기 때문에 계산량을 현저히 줄일 수 있다.

4. 제안된 알고리즘의 성능 평가

본 장에서는 제안된 고속 컨벌루션 방법은 기존의 컨벌루션 방법과의 성능 비교를 통해 평가되었다. 이를 위해서 모든 신호들의 Sampling rate는 48kHz로 하였다. 성능 평가의 세부 항목은 계산량과 음질로 하였다.

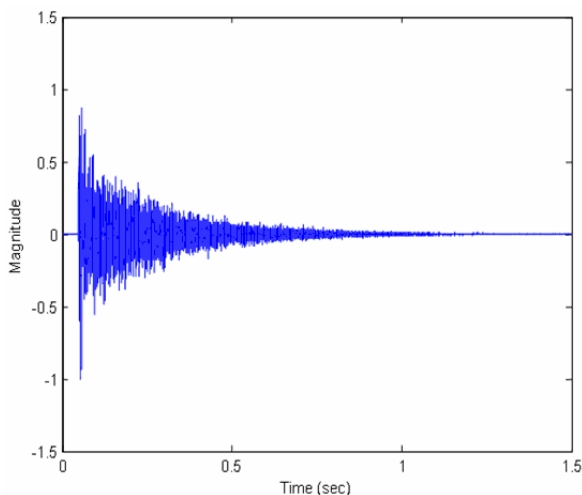
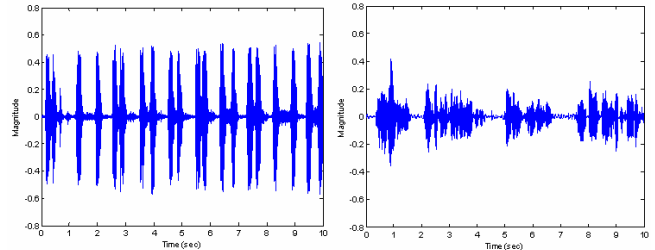


그림 5. 실내 충격 응답의 예

그림 5는 성능 평가를 위해 도입된 RIR이며 이는 New Mexico State University의 한 욕실에서 측정되었다[9].



(a) A dog's bark signal (b) A voice signal

그림 6. 입력 신호의 예

위의 그림 6은 성능 평가를 위해 사용된 입력 신호들이며 (a)는 개 짖는 소리, (b)는 사람의 목소리를 무향실에서 10초 동안 녹음한 것이다. 본 장의 성능 평가에서 입력 신호 블록의 길이는 10000 Samples, RIR 신호 블록의 길이는 1000 Samples로 하였다.

4.1 계산량의 평가

본 절은 제안된 알고리즘의 계산량을 기존의 알고리즘과 비교하여 평가하였다. 계산량은 단위 결과 신호 블록 길이당 신호 처리시 가장 많은 CPU의 부담으로 작용하는 곱셈의 수로 정의하였다. 표 1은 5가지 컨벌루션 방법에 따른 계산량을 나타내었다.

표 1. 컨벌루션 방법에 따른 계산량

Convolution Type	Input Signals	
	Dog's Bark	Voice
Linear Convolution	720000000	
Block Convolution	29421459	
BC with T-PAF	153237	
BC with F-PAF	13068494	10184944
Proposed Method	78105	55657

위의 표 1에서는 기존의 선형 컨벌루션 방법, 블록 컨벌루션 방법, 시간 영역 심리 음향 필터만 적용한 블록 컨벌루션 방법(Block Convolution with Time-domain Psycho-Acoustic Filter, BC with T-PAF), 주파수 영역 심리 음향 필터만 적용한 블록 컨벌루션 방법(Block Convolution with Frequency-domain Psycho-Acoustic Filter, BC with F-PAF), 그리고 본 논문에서 제안된 고속 컨벌루션 방법에 따른 계산량을 나타내고 있다. 제안된 심리 음향 필터는 단독으로 사용될 때, 기존의 방법과 비교하여 상당한 계산량을 감소시키는 역할을 하였고, 시간 영역과 주파수 영역의 심리 음향 필터를 모두 사용하는 제안된 고속 컨벌루션 방법이 단위 길이당 가장 적은 계산량을 보여주고 있다.

4.2 음질의 평가

제안된 알고리즘의 음질을 평가하기 위해 주관 음질 평가를 실시하였다. 선형 컨벌루션 결과 신호를 기준 신호로 정의하고 시간 영역 심리 음향 필터만 적용한 블록 컨벌루션 방법(Block Convolution with Time-domain Psycho-Acoustic Filter, BC with T-PAF), 주파수 영역 심리 음향 필터만 적용한 블록 컨벌루션 방법(Block Convolution with Frequency-domain Psycho-Acoustic Filter, BC with F-PAF), 본 논문에서 제안된 고속 컨벌루션 방법의 결과 신호를 비교 신호로 정의한다.

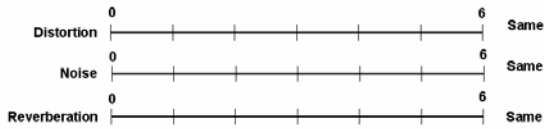
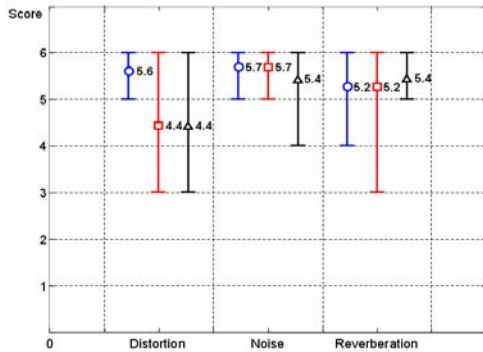
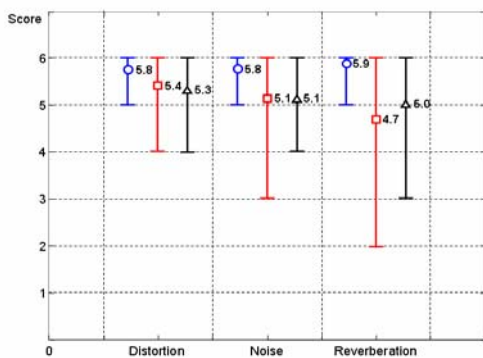


그림 7. 주관 음질 평가표

위의 그림 7은 주관 음질 평가표를 나타낸다. 음질 평가는 헤드폰을 사용하여 기준 신호를 10 초 동안 들려주고 2 초의 공백을 둔 뒤 비교 신호를 10 초 동안 들려주어 기준 신호와 비교하여 음질의 왜곡 여부(Distortion), 잡음의 생성(Noise), 잔향의 왜곡(Reverberation)의 항목으로 평가하였다. 두 신호가 가장 유사하면 6 점에 완전히 다르다면 0 점에 표시하도록 하였다. 평가의 대상은 각 평가 지표에 대해 사전 교육이 이루어진 11 명의 정상 청각을 지닌 남자로 하였다.



(a) Dog's bark convolved with bathroom RIR



(c) Voice convolved with bathroom RIR

그림 8. 주관 음질 평가 결과

그림 8은 각각의 경우에 대해 주관 음질 평가의 결과를 나타내고 있다. 파란색 실선과 원으로 나타낸 것은 BC with F-PAF의 결과이고, 빨간색 실선과 사각형으로 나타낸 것은 BC with T-PAF의 결과이다. 제안된 고속 컨벌루션의 음질 평가 결과는 검은색 실선과 삼각형으로 표시하였다. 여기서 실선은 점수의 분포를 나타내고 각 도형과 숫자는 평균값을 나타낸다. 이 결과는 각 항목별 평가 결과 상당한 유사성을 나타내는 3 점 이상의 점수를 얻었고 이를 통해 음질 측면에서 기준 신호와 비교 신호 사이의 많은 차이를 나타내지 않았음을 알 수 있었다.

5. 결론

잔향 생성기에서 실시간 신호 처리를 목표로 하는 고속 컨벌루션 알고리즘이 소개되었다. 인간의 청각 특성을 시간 영역과 주파수 영역에서 고려하여 처리되는 데이터의 양을 줄임으로써 계산량의 감소를 도모하였다. 제안된 알고리즘은 기존의 블록 컨벌루션 방법을 도입하였고 심리 음향 필터를 설계하여 적용하였다. 이를 통해 상당한 계산량의 감소를 실현하였고 또한 주관 음질 평가를 통해 음질의 왜곡이 거의 없다는 것을 알 수 있었다. 본 알고리즘은 대상 신호에 따라 심리 음향 필터의 디자인을 적절히 조절하여 사용할 수 있으며 또한 음향 신호를 대상으로 하는 실시간 응용 분야에 효과적으로 사용될 수 있다.

참고문헌

- [1] J. M. Jot, A. Chaigne, 1991, "Digital delay networks for designing artificial reverberators," in Proc. 90th. Audio Engineering Society Convention, preprint 3030.
- [2] J. Dattorro, 1997, "Effect design Part 1: Reverberator and other filters," Journal of Audio Engineering Society, vol. 45, pp. 660-684.
- [3] W. G. Gardner, 1992, "The virtual acoustic room," dissertation for the degree of Master of Science, Massachusetts Institute of Technology, Media Lab.
- [4] M. Schonle, U. Zolzer, N. Fliege, 1992, "Modeling of room impulse responses by multirate systems," in Proc. 93rd Audio Engineering Society Convention, Preprint 3447.
- [5] A.V. Oppenheim, R.W. Schaffer, 1989, Discrete Time Signal Processing, Englewood Cliffs, NJ, Prentice Hall.
- [6] L. Wen-Chieh, L. Chi-Min, Y. Chung-Han, G. Jiun-In, 2003, "Fast perceptual convolution for room reverberation," in Proc. 6th International Conference on Digital Audio Effects, London, UK.
- [7] K. Iida, K. Mizushima, Y. Takagi, T. Suguta, 1995, "A new method of generating artificial reverberant sound," in Proc. 99th Audio Engineering Society Convention.
- [8] E. Zwicker, R. Feldtkeller, 1967, The Ear as a Communication Receiver, Hirzel Verlag, F.R.G..
- [9] http://www.ece.nmsu.edu/~pdeleon/BSS/Impulse_Responses/README_Impulse_Responses.pdf.