

사용자의 음장선호도에 따른 오디오 콘텐츠 적응 기술

Audio Contents Adaptation Technology According to User's Preference on Sound Fields

서 정 일*, 강 경 옥*, 홍 재 근**

(Jeongil Seo*, Kyeongok Kang*, Jaegun Hong**)

*한국전자통신연구원 방송미디어연구그룹, **경북대학교 전자전기공학부

(접수일자: 2004년 5월 20일; 채택일자: 2004년 8월 3일)

본 논문에서는 사용자의 음장 선호도를 이용하여 오디오 콘텐츠를 사용자가 원하는 음장으로 변환하는 기술에 대하여 설명한다. 오디오 신호가 재생되는 공간의 음장을 사용자가 원하는 음장으로 변환시켜주는 음장효과 기술은 실감있고 현장감있는 음악재생에 필수적인 요소이다. 그러나, 음장효과를 실시간으로 처리하기 위해서는 막대한 연산량이 필요하므로 MP3 플레이어와 같은 휴대용 오디오 단말에서는 구현하기 힘들다. 본 논문에서는 사용자로부터 전달된 음장 선호도를 이용하여 서버에서 음장효과를 처리하도록 하여, 단말의 성능에 구애받지 않고 음장효과를 제공할 수 있는 기술을 제안한다. 사용자가 선호하는 음장을 표현하기 위하여 선호하는 음장을 실내응답신호의 URI 주소를 이용하여 표현하는 방법 뿐만 아니라 음향공간에 대한 심리적 파라미터를 이용할 수 있게 하였다. 또한, 실내응답신호와 복직분 연산을 통한 음장효과 처리 방법을 실시간 응용에 적용하기 위하여 고속 복직분 알고리즘을 제안하였으며, 실험을 통하여 실시간 응용에도 적용이 가능함을 확인하였다. 본 논문에서 제안한 음장 선호도 서술구조의 효용성을 검증하기 위하여, 일반인을 대상으로 음장을 구분하는 능력과 음장효과가 처리된 음악에 대한 선호도에 대한 주관듣기평가를 실시하여 제안된 음장 선호도가 일반인들에게 적용이 가능함을 확인하였다.

핵심용어: 실내응답신호, 심리적 음향 파라미터, 고속 복직분 알고리즘

투고분야: 건축음향 분야 (7.1)

In this paper, we describe a novel method for transforming audio contents according to user's preference on sound field. Sound field effect technologies, which transform or simulate acoustic environments as user's preference, are very important for enlarging the reality of acoustic scene. However huge amount of computational power is required to process sound field effect in real time, so it is hard to implement this functionality at the portable audio devices such as MP3 player. In this paper, we propose an efficient method for providing sound field effect to audio contents independent of terminal's computational power through processing this functionality at the server using user's sound field preference, which is transferred from terminal side. To describe sound field preference, user can use perceptual acoustic parameters as well as the URI address of room impulse response signal. In addition, a novel fast convolution method is presented to implement a sound field effect engine as a result of convolving with a room impulse response signal at the realtime application, and verified to be applicable to real-time applications through experiments. To verify the evidence of benefit of proposed method we performed two subjective listening tests about sound field descriptive ability and preference on sound field processed sounds. The results showed that the proposed sound field preference can be applicable to the public.

Keywords: Sound field, Preference, Room impulse response signal, Perceptual acoustic parameter, Fast convolution algorithm

ASK subject classification: Architectural acoustics (7.1)

I. 서론

지난 수십년간의 멀티미디어 콘텐츠를 위한 데이터 압축기술의 급속한 발전과 무선 인터넷, IMT-2000과 같은 새로운 전송환경의 등장으로 인하여 멀티미디어 콘텐츠를 언제 어디서나 (Anytime and Anywhere) 접속하고 소비하는 기술에 대한 요구가 날로 증가하고 있다. 또한 대화형 방송기술과 광대역 통신 기술의 발전은 방송과 통신기술의 융합을 촉진하고 있다. 이와 관련하여, 멀티미디어 콘텐츠에 대한 제작, 분배, 소비과정을 모두 포함하는 광의의 프레임워크에 대한 국제표준화가 MPEG에서 2000년 초부터 MPEG-21 이란 이름으로 진행중이다. MPEG-21은 디지털 콘텐츠의 소비에 있어서 안전하고 편리한 환경을 제공하기 위한 일반적인 구조와 관련 도구들을 제공하는 것을 목표로 하고 있다[1].

MP3 플레이어와 같은 이동형 디지털 오디오 플레이어의 등장과 인터넷을 통한 오디오 방송 및 오디오 콘텐츠 다운로드 서비스가 가능해짐에 따라 시공간상의 제약없이 오디오 콘텐츠를 이용하는 것이 가능해지고 있으며, 막대한 연산량을 가지는 DSP가 출현함에 따라서 오디오 이퀄라이저나 음장제어기 등과 같이 사용자의 취향에 맞도록 음악의 색감이나 분위기를 제어하는 기능 또한 제공되고 있다[2]. 그러나, 핸드폰이나 PDA와 같은 이동형 단말에서는 배터리나 CPU의 연산능력의 한계가 있으므로 오디오 콘텐츠를 사용자의 취향에 맞도록 변환하는 모든 기능을 제공하기에는 무리가 있다. 2003년 표준화가 완료된 MPEG-21 DIA (Digital Item Adaptation) 표준은 사용자가 멀티미디어 콘텐츠를 사용할 때의 네트워크

크, 사용자 특성, 주변 환경 뿐만 아니라 단말의 능력등을 표현할 수 있는 사용 환경 도구 (Usage Environment Tool)를 제공하여, 서버나 라우터에서 단말의 환경에 맞게 콘텐츠를 적응 또는 변환하기 위한 도구들을 제공한 데[3]. 그림 1은 MPEG-21 DIA 기술을 이용하여 MP3로 압축된 오디오 콘텐츠를 다양한 오디오 단말에 맞게 변환하고 소비하는 흐름을 나타내고 있다. 또한, MPEG-21 DIA 사용 환경 도구에서는 오디오 콘텐츠를 재생할 때 사용자의 선호도를 기술할 수 있는 도구를 정의하고 있다.

본 논문에서는 MPEG-21 DIA 기술을 이용하여 오디오 콘텐츠를 사용자의 음장에 대한 선호도에 따라 적응 또는 변환하는 방법을 제안한다. 음장의 특성을 표현하는 방법은 음향공간에서 직접 취득한 실내응답신호 (room impulse response)를 직접 이용하는 방식, 실내응답신호로부터 잔향시간 (reverberation time)과 같은 음장 분석 파라미터들을 추출하여 표현하는 방식 및 음장에 대한 지각적인 파라미터를 이용하는 방식이 있다 [4]. 일반적인 사용자에게 친숙하고 편리한 방식은 자신이 원하는 음향공간을 실내응답신호를 이용하여 직접 지정하는 방식과 자신이 원하는 음장을 지각적인 파라미터를 이용하여 지정하는 방식이라고 할 수 있다. 따라서, 본 논문에서는 사용자가 선호하는 음장을 위 두가지 방식으로 기술할 수 있는 서술구조를 제안한다[5-7].

실내응답신호를 이용하여 음장효과를 처리하기 위해서는 입력 오디오 신호와 실내응답신호를 목적분하는 과정이 필요하다. 교회나 콘서트홀과 같이 넓은 공간에서의 실내응답신호는 3~4초 가량의 긴 응답시간을 가지므로 목적분과정에서 많은 연산량을 필요로 하며, 연산량을 줄이기 위하여 주파수대역에서 처리하려면 실내응답신호 만큼의 초기지연이 발생하므로 실시간 처리에 문제가 있다[4, 8]. 본 논문에서는 주파수 영역에서 필요한 FFT의 연산량을 줄이기 위해서 실내응답신호를 부단위로 분할하여 병렬로 처리하는 알고리즘을 제안하여 연산량을 감소시키는 방법을 제안한다. 본 논문에서 제안된 음장 선호도 서술구조와 고속 목적분 알고리즘을 이용하여 음장에 대한 선호도에 따라 오디오 신호를 변환하였을 때 오디오 신호에 대한 현장감과 사실감이 증대함을 주관들 기평가 실험을 통하여 확인하였다.

본 논문의 구성은 다음과 같다. 제 II장에서는 음장에 대한 사용자의 선호도를 기술하는 서술구조에 대하여 설명하고, 음장에 대한 선호도를 처리하였을 때 오디오 신호에 대한 현장감 또는 사실감이 증대됨을 확인하기위한

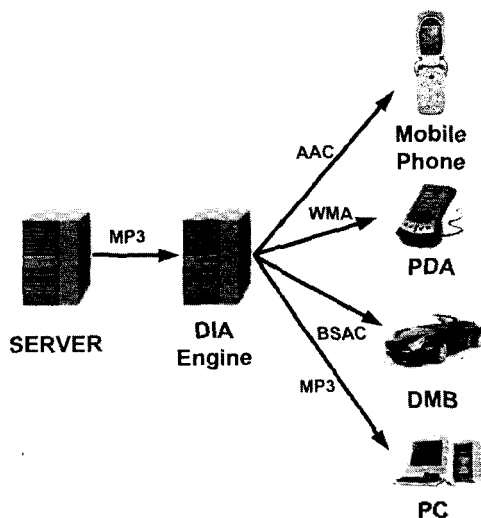


그림 1. MPEG-21 DIA를 이용한 오디오 콘텐츠 변환 서비스 개념도
Figure 1. Conceptual diagram of audio contents adaptation service using MPEG-21 DIA

주관듣기평가 실험의 결과를 기술한다. 제 III장에서는 실내응답신호를 이용하여 음장에 대한 선호도를 실시간으로 처리하기 위한 고속 복직분 알고리즘에 대하여 서술하고, 고속 복직분 알고리즘의 연산량 감소에 대한 실험결과를 기술한다. 끝으로 제 IV장에서는 본 연구의 요약 및 향후과제를 기술한다.

II. 음장 선호도 서술구조

멀티채널 오디오 기기의 보급과 더불어 DSP를 이용한 음향효과 기능이 추가되기 시작했으며, 그 중 가장 주요한 부분은 음장효과이다. 음장효과 (sound field effect)는 사용자가 원하는 음장을 모사함으로써 오디오 신호의 현장감과 사실감을 주는 효과외에, 음원의 위치감과 거리감을 증대시키는 부가적인 효과도 가지고 있다. 그러나, 충실한 음장재현을 위해서는 실내응답신호를 조절하여 처리해 주어야 하며, 실내응답신호와 오디오 신호와의 복직분 연산이 필요하므로 많은 양의 연산량이 요구된다. 상용 멀티채널 오디오 프로세서에서는 고성능 DSP를 이용하여 처리하고 있지만 휴대폰이나 PDA와 같이 제한된 성능을 가진 단말에서 음장효과를 구현하기에는 여러가지 제약점이 있다. 본 논문에서는 인터넷과 같은 서버-클라이언트 구조에서 오디오 콘텐츠를 소비할 때 단말에서 사용자의 음장에 대한 선호도를 기술하여 서버로 전송하면 서버에서 음장효과 과정을 처리한 후 오디오 콘텐츠를 전송함으로써 임의의 단말에서도 사용자가 원하는 음장대로 오디오 콘텐츠를 이용할 수 있는 기술을 제안한다. 그림 2는 본 논문에서 제안한 음장 선호도 서술구조를 이용하여 구성된 오디오 콘텐츠 서비스 시나리오를 나타낸다. 사용자가 임의의 오디오 콘텐츠를 서버에게 요구하면 (1) 서버는 콘텐츠를 라우터나 스위치에 위치하고 있는 적응엔진으로 전송한다 (2). 단말에

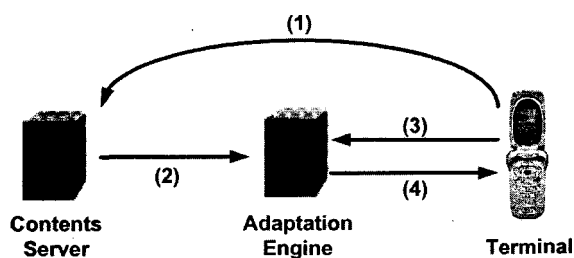


그림 2. 음장 선호도 서술구조를 이용한 오디오 콘텐츠 서비스 시나리오.
Figure 2. Audio contents service scenario using sound field preference description structure.

서 사용자가 원하는 음장으로 듣기 위하여 원하는 음장에 대한 정보를 음장 선호도 서술구조를 이용하여 적응엔진으로 전달하면 (3), 적응엔진에서 서버로부터 받은 오디오 콘텐츠에 단말에서 원하는 음장효과를 처리하여 단말로 전송함으로써 (4), 최종적으로 음장효과가 부가된 오디오 콘텐츠를 단말에서 이용할 수 있게 된다.

III. 음장특성 표현방법

음장 (sound field)이라고 불리는 실내의 음향특성은 음향공간의 크기 및 형태와 벽면의 재료에 의해서 결정된다[9, 10]. 음원으로부터 음파가 생성되면, 청취자에게 직접 도달하는 직접음 (direct sound) 외에, 그림 3과 같이 벽면이나 바닥, 천정등의 다양한 경로를 통해 반사된 음들이 반사경로에 따라 시간차를 가지며 도달하게 된다.

잔향 (reverberation sound)은 복잡한 반사경로를 가진 반사음들의 그룹에 의한 소리의 울림현상을 말하며, 음장의 특성을 규명하는 가장 중요한 파라미터이다. 최근들어서는 직접음과 잔향음 사이의 초기반사음 및 측면반사음 또한 음장의 특성에 많은 영향을 주는 것으로 판

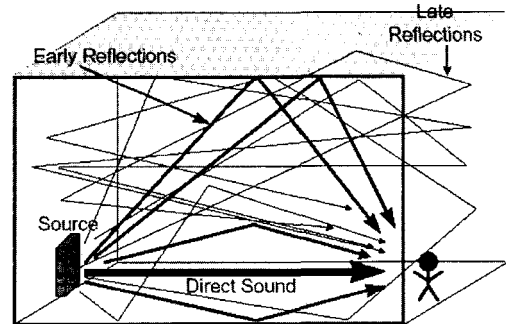


그림 3. 음파전달 모델.
Figure 3. Sound propagation model.

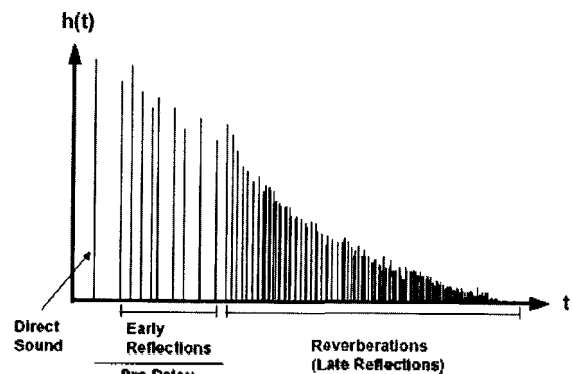


그림 4. 실내응답신호의 일반모델.
Figure 4. General room impulse response model.

명되어, 음장의 특성을 명확하게 규정하기 위해서 음향 공간에서 관측된 실내응답신호의 일부가 아닌 전체를 이용하고 있다. 그림 4는 임의의 음향공간에 대한 실내응답신호를 간략화하여 나타낸 것이다.

음장을 표현하는 방법은 상기에서 기술한 바와 같이 실내응답신호를 이용하는 방법과 잔향시간 (reverberation time)과 같은 음장 분석 파라미터를 이용하는 방법이 있다. 또한, 음장 분석 파라미터를 사람의 음장에 대한 느낌을 표현한 지각적 파라미터로 변환하여 나타내는 방법이 제안되어 있다[11].

IV. 음장 선호도 서술구조

일반적으로 음장효과를 처리하기 위하여 실내응답신호를 이용하는 방법은 많은 연산량을 필요로 하므로, 음장 분석 파라미터와 인공잔향기를 이용한 방법이 널리 사용되고 있으나, 음향학에 대한 지식이 부족한 일반적인 사용자가 자신이 원하는 음장을 음장 분석 파라미터로 표현하기에는 곤란한 면이 있다. 따라서, 자신이 원하는 음장특성을 가진 음장공간을 실내응답신호를 이용하여 직접 지정하는 방법과 지각적인 파라미터를 이용하여 자신이 원하는 음장을 표현하게 하는 것이 적절하다. MPEG-21 DIA 표준에서는 오디오 콘텐츠에 대한 사

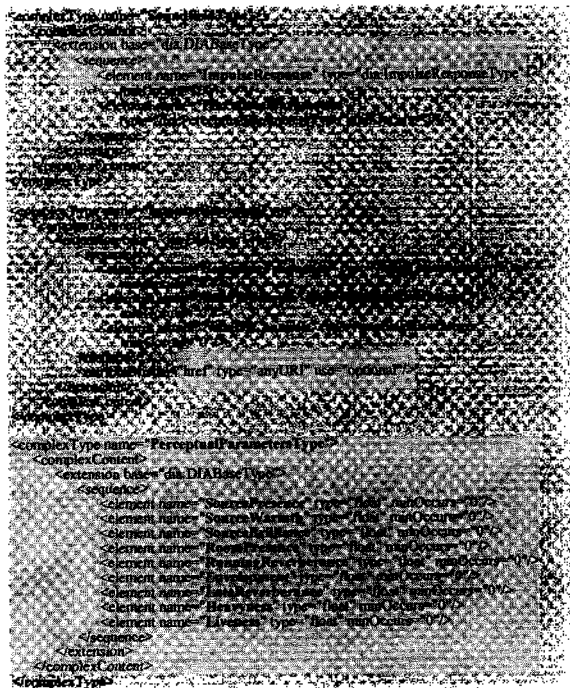


그림 5. 음장선호도 서술구조.
Figure 5. Sound field preference description scheme.

용자의 선호도를 표현하기 위한 서술구조를 XML언어 기반으로 AudioPresentationPreference 도구에서 정의하고 있다. 본 논문에서는 음장에 대한 선호도를 실내응답신호와 지각적인 음장 파라미터를 이용한 서술구조를 그림 5와 같이 표현될 수 있으며, MPEG-21 DIA 표준에 채택되었다[8]. 실내응답신호는 실내응답신호가 저장되어 있는 URI 주소를 직접기술하고, 표본화 주파수, 샘플당 비트수를 기술할 수 있게 함으로써 사용자가 편리하게 자신이 원하는 음장을 표현할 수 있게 하였다. 또한, 실내응답 신호에 대한 채널수를 기술할 수 있게 함으로써 멀티채널로 녹음된 실내응답신호 또한 적용할 수 있도록 하였다. 지각적인 음장 파라미터를 이용하는 방법에서는 MPEG-4 오디오 장면기술자(Audio BIFS) 2차버전에서 가상 음장 (virtual sound field)을 지각적인 파라미터로 표현하는 PerceptualParameters 노드에서 사용된 9개의 음장 지각 파라미터를 적용함으로써, MPEG-4 시스템과의 호환성을 유지하면서도 사용자의 느낌을 이용하여 음장을 표현하는 것이 가능하도록 하였다[12].

본 논문에서 제안된 음장선호도 서술구조를 이용하여 적응엔진에서 오디오 콘텐츠를 변환하는 과정을 간략화하여 나타내면 그림 6과 같다. 사용자가 오디오 콘텐츠를 서버로부터 요구하면 (1), 서버에서는 적응엔진이 위치한 라우터나 스위치로 오디오 콘텐츠를 전달한다 (2), 적응엔진은 사용자의 음장선호도를 음장 선호도 기술자를 통해 전달받아 (3) 서버로부터 받은 오디오 콘텐츠를 실내응답신호나 지각 음장 파라미터를 이용하여 음장효과를 부여하여 (4) 사용자에게 전달한다 (5).

V. 음장 구별 실험

본 논문에서 제안한 음장 선호도 서술구조의 효율성을

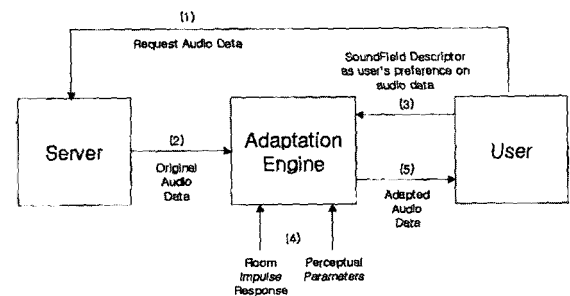


그림 6. 음장 선호도 서술구조에 따른 오디오 콘텐츠 적응과정 개념도.
Figure 6. The conceptual diagram of audio contents adaptation according to sound field preference description scheme.

표 1. 음장 구별 실험 결과.

Table 1. The result of sound field description test.

Sequence	Expected		ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	UoW	UoW	UoW	UoW	matched	# Tests	%
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16			
1	B	Church	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	16	16	100.0%
2	A	Church	A	A	A	A	A	A	A	A	A	B	A	A	B	A	A	A	14	16	87.5%
3	A	Church	A	A	A	A	A	A	A	A	A	B	A	A	A	B	B	B	12	16	75.0%
4	A	Church	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	16	16	100.0%
5	B	Church	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	16	16	100.0%
6	B	MusicHall	B	B	B	B	B	B	B	B	B	A	A	B	B	B	B	14	16	87.5%	
7	A	MusicHall	A	A	A	A	A	A	A	A	A	A	B	A	A	B	B	A	13	16	81.3%
8	B	MusicHall	A	B	B	B	B	A	B	B	B	A	A	B	A	B	A	B	10	16	62.5%
9	A	MusicHall	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	16	16	100.0%
10	A	MusicHall	A	A	A	A	B	A	A	A	A	A	A	A	B	A	A	A	14	16	87.5%
11	A	Stadium	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	A	15	16	93.8%
12	B	Stadium	B	B	B	B	B	B	B	B	B	B	A	B	A	A	A	B	12	16	75.0%
13	A	Stadium	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	16	16	100.0%
14	B	Stadium	B	B	B	B	B	B	B	B	B	B	A	B	B	B	B	A	14	16	87.5%
15	A	Stadium	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	16	16	100.0%
		matched	14	15	15	15	14	14	15	15	15	10	10	15	11	12	11	13	214	240	89.2%

표 2. 음장 선호도 실험 결과.

Table 2. The result of sound field preference test.

Sequence	Expected		ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	ETRI	UoW	UoW	UoW	UoW	sound field selected	# Tests	%
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16			
1	B	Church	A	A	B	A	B	B	A	B	B	B	A	B	A	A	B	B	9	16	56.3%
2	A	Church	B	B	A	B	A	A	B	A	A	B	B	A	A	B	B	A	8	16	50.0%
3	A	Church	A	B	A	B	A	A	A	A	A	A	B	A	A	B	A	A	12	16	75.0%
4	A	Church	B	B	A	B	A	A	B	A	A	A	B	A	B	B	B	A	8	16	50.0%
5	B	Church	A	A	B	A	B	A	A	B	B	B	A	B	A	A	B	B	8	16	50.0%
6	B	MusicHall	A	A	B	A	B	B	A	B	B	B	A	B	A	A	A	A	7	16	43.8%
7	A	MusicHall	B	B	A	B	A	A	B	A	A	A	B	A	B	B	B	A	8	16	50.0%
8	B	MusicHall	A	A	B	A	B	A	A	B	B	B	A	B	A	A	B	Z	7	16	43.8%
9	A	MusicHall	A	B	A	B	A	A	A	A	A	A	A	A	A	A	B	A	13	16	81.3%
10	A	MusicHall	A	B	A	B	B	A	A	A	A	B	B	A	A	A	B	B	9	16	56.3%
11	A	Stadium	B	B	A	B	A	A	B	A	A	B	B	A	B	A	A	A	9	16	56.3%
12	B	Stadium	A	A	B	A	B	B	A	B	B	B	B	B	A	A	A	B	9	16	56.3%
13	A	Stadium	A	B	A	B	A	A	A	A	A	A	B	A	A	B	A	A	12	16	75.0%
14	B	Stadium	A	A	B	A	B	B	A	B	B	B	A	B	B	A	B	B	10	16	62.5%
15	A	Stadium	B	B	A	B	A	A	B	A	A	A	B	A	A	A	A	A	11	16	68.8%
		sound field selected	4	0	15	0	14	13	4	15	15	12	2	15	7	4	8	12	140	240	58.3%

검증하기 위하여, 일반인을 대상으로 음장을 구분하는 능력과 음장효과가 처리된 음악에 대한 선호도에 대한 주관듣기평가를 실시하였다.

음장구분 실험을 위하여 무향실에서 녹음한 5섯가지의 음원 (성악, 바이올린, 첼로, 타악, 트럼펫)을 교회, 음악당, 종합운동장의 세가지 음장으로 처리하여 15개의 실험 음원을 작성하였다. 16명으로 구성된 듣기평가 그룹에서 각 음원을 듣고 음장을 선택한 결과는 표 1과 같으며, 실험결과 89.2%의 음장에 대한 변별력을 보여 일반인들이 음장을 비교적 정확하게 구별할 수 있음을 확

인하였다. 따라서, 음장의 특성에 따른 음질의 변화를 일반인들이 구별한다고 볼 수 있음을 확인하였다.

음장효과에 대한 일반인의 선호도를 측정하기 위하여, 동일한 음원과 평가 그룹으로 음장효과가 가미된 음원과 원음에 대한 비교 청취를 수행하였다. 본 실험에서는 피 청취자에게 두개의 음원 중 보다 자연스럽게 들리는 것을 선택하도록 하였으며 결과는 표 2와 같다. 음장효과 처리기에서의 삽입되는 잡음의 영향에도 불구하고 58.3%의 음원에 대해 선호도를 나타냄을 확인하였다. 또한 네명의 청취자는 음장효과가 처리된 음원을 모두

선호하는 결과를 나타내었으나, 한명의 청취자는 원음을 선호하는 것으로 나타내었다. 이는 음장효과에 대한 개인의 편차가 있음을 나타내며, 오디오 콘텐츠에 현장감과 사실감을 증대시키기 위한 음장효과를 모든 사용자에게 적용하기 보다는 음장효과를 선호하는 사람들에게 선택적으로 적용하는 것이 보다 효과적임을 나타낸다.

상기의 두가지 실험을 통하여 음장 선호도 서술구조가 일반인들에게 적용이 가능함을 확인하였으며, 이를 일률적으로 적용하는 것 보다는 사용자의 선호도에 따라 음장효과를 선별적으로 적용하는 것이 적절함을 알 수 있었다.

VI. 부단위 고속 복적분 알고리즘

음장 선호도를 실내용답신호를 이용하여 나타내었을 때 음장효과를 처리하기 위해서는 입력 오디오 신호와 실내용답신호를 시간 영역에서 선형복적분을 수행하여야 한다[13]. 시간 영역에서의 선형복적분을 이용한 음장효과 구현은 실내용답신호의 응답시간이 상당할 경우 많은 연산량이 필요하므로 실시간 구현에 어려움이 있다. 본 논문에서는 선형 복적분을 주파수 영역에서의 오버랩세이프 방법으로 수행하고, 실내용답신호를 FFT 연산량이 최소가 되는 개수의 서브 블록으로 나눈 후 병렬로 처리함으로써 연산량을 최소화하여 실시간 수행이 가능한 방법을 제안하였다. 제안된 고속 복적분 알고리즘을 이용하여 응답시간이 긴 실내용답신호를 처리할 수 있는 고속 음장효과 처리기를 구현하였다.

VII. 실내용답신호를 이용한 음장처리기

일반적으로 실내용답 신호의 응답 시간은 측정 공간이 방대하고 잔향시간이 길 경우 수초에 달한다. 그러므로 시간영역에서의 복적분 (convolution) 연산을 이용한 음장효과 구현은 실시간 처리가 불가능하고 그 연산량 또한 방대하다. 복적분 (convolution) 연산의 연산량 감소는 주파수 영역으로 변환 처리하는 오버랩세이프 (Overlap-save)방법에 의하여 가능하다. 예를들어 시간 영역에서 디지털 신호로 양자화된 N 개의 샘플신호를

복적분 (convolution)하기 위해서는 $O(N^2)$ 의 복잡도가 예상되어지나, N 개의 신호를 고속 푸리에 변환 (Fast Fourier Transform)을 이용하는 오버랩세이프 방법에서의 복적분 (convolution) 연산은 $O(N \log N) + O(N)$, 즉 $O(N \log N)$ 의 복잡도를 갖는다. 그러나 N의 길이가 상당하면 FFT 수행을 위한 블록 버퍼 크기의 증가로 인한 시간적 지연과, FFT 수행 자체를 위한 많은 연산량을 필요로 한다.

전자의 시간적 지연 문제를 해결하고자 Gardner는, 시간 영역에서의 선형복적분 방법과 주파수 영역에서의 오버랩세이프 방법을 효과적으로 연계한 방법을 제안하여 블록 버퍼 크기로 인한 시간적 지연은 해결하였으나, 최종적으로 처리하여야 하는 FFT 블록버퍼의 크기는 실내용답 신호의 길이만큼 커야 하므로 FFT 수행을 위한 연산량은 감소시키지 못 하였다[14].

VIII. 부단위 고속 복적분 알고리즘

지연시간이 상당한 실내용답신호를 주파수영역에서 오버랩세이프 방법을 이용하여 선형복적분을 수행할 때 필요한 FFT 수행을 위한 연산량을 줄이기 위하여, 본 논

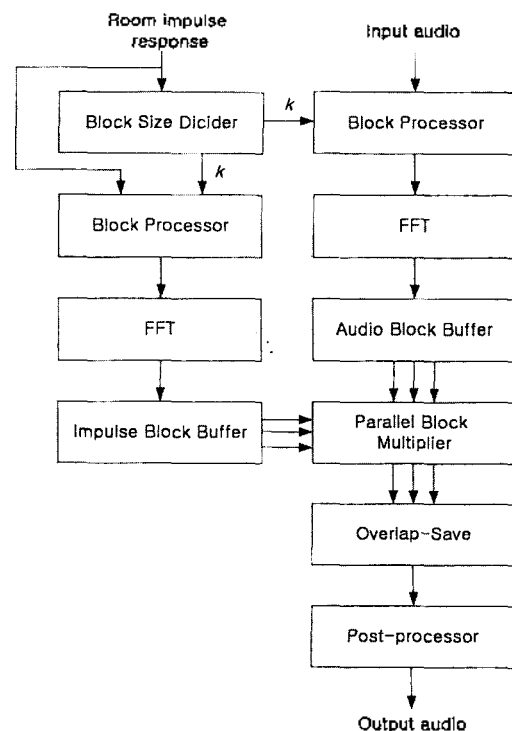


그림 7. 부단위 고속복적분 알고리즘 블록도.
Figure 7. The block diagram of parallel convolution method based on sub-block.

문에서는 실내용담신호를 FFT 연산이 최소가 되도록 블록수를 결정하고 실내용담신호를 분할한 뒤 병렬로 입력 오디오 신호와의 오버랩세이프 연산을 수행하는 부단위 고속 복적분 알고리즘을 제안한다. 본 논문에서 제안한 고속 복적분 알고리즘의 전체 블록도는 그림 7 과 같다.

먼저, 블록크기 결정부는 실내용담 신호를 부단위 블록으로 분할함에 있어서 주파수 영역의 복적분에 필요한 FFT 연산량을 최소화하도록 블록의 갯수 k 값을 결정한다. k 값은 식 (1)로부터 결정되며 이는 그림 8로부터 설명될 수 있다.

$$k = \arg \min_k (Cst_k - Rst_k),$$

$$Rst_k = 0.5 \times k \times L + L/2,$$

$$Cst_k = (2 \times k \times M \times \log M + k \times L + L) - (2 \times \log L + L),$$

$$M = L/k \tag{1}$$

식 (1)은 부분적으로 분할시 초기 L 개의 응답을 얻기 까지 소요되는 곱셈 연산량으로 그림 8의 (a) 영역과 (b) 영역으로부터 구한다. 이때 부분적으로 분할 시 곱셈 연산량의 감소는 (a) 영역으로부터 유도된다.

길이가 1 인 실내용담 신호를 벡터열 $A = [a_1, a_2, \dots, a_L, a_1]^T$ 로 표현할 때, 길이 1 에 대하여 $2q-1 < 1 < 2q$ 를 만족 하는 q 를 구하면, 부단위 블록의 길이 $L = 2q+1$ 로 정의한다. 여기서 블록의 개수 k 는 $k = 2j$ ($j=1,2,\dots$) 로 써, 식 (1)을 고려할 때 j 는 1과 3의 범위 안에 존재한다. 따라서, 실내용담신호에 대한 블록처리기의 출력 벡터는 식 (2)와 같이 표현된다.

$$a_j = \left[\underbrace{a_1 a_2 \dots a_1}_{2^j} \underbrace{0 \dots 00 \dots 0}_{2^j} \right]^T \tag{2}$$

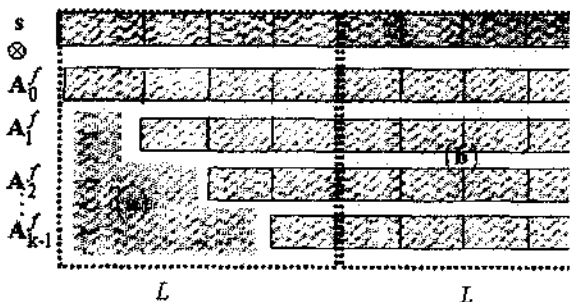


그림 8. 부분적으로 분할시 발생하는 곱셈연산 계산 영역 [(a): Rstk, (b): Cstk, ⊗샘플간 곱셈연산자]
Figure 8. Different computation regions in the process of uniformly splitting block overlap-save method. (a) is Rstk. (b) is Cstk

표 3. 선형 및 순환복적분 방법에 필요한 곱셈 연산 수
Table 3. Required multiplications for linear convolution and circular convolution.

N2 (N1 = 2N2)	Linear convolution (time domain)	Circular convolution (frequency domain)	Ratio (Circular/Linear)
8	64	480	7.5
128	16,384	10,240	0.625
512	262,144	46,080	0.175
2048	4,194,304	204,840	0.048

실내용담신호의 FFT 출력은 식 (3)의 형태로 실내용담 버퍼에 저장된다.

$$A_j^f = [a_1^f \ \Lambda \ a_{2^j}^f]^T \tag{3}$$

그림 7에서 오디오 블록처리의 출력인 임의의 블록 샘플시간 b 에서 음장감 부여를 위한 입력 오디오 신호 s(b) 는 식 (4)와 같이 정의한다.

$$s(b) = [s(b-M+1)L \ s(b)L \ s(b+M-1)]^T \tag{4}$$

FFT 연산을 통하여 s(b) 로부터 Sf(b) 를 생성하고, 오디오 블록 버퍼에서 각 블록샘플시간마다 한 블록씩 지연하여 k 개의 블록 버퍼가 채워지도록 한다. 블록 곱셈기에서는 각 부단위 블록의 샘플간의 곱셈을 수행하고, 블록 합산기에서는 각 부단위 블록을 합산하여 단일 부단위 블록을 생성한다. 블록 합산기를 거친 단일 부단위 블록은 블록 후처리기에서 알리어싱 (aliasing) 부분을 제거하고 역푸리에 연산을 수행하여 시간축 출력신호를 블록단위로 생성한다.

IX. 처리속도 측정 실험

본 논문에서 제안된 알고리즘의 성능을 분석하기에 앞서, 실내용담신호의 길이에 따라 시간 영역에서의 선형 복적분 방법과 주파수 영역에서의 순환복적분 방법간의 연산량을 비교해 보았다. 아래 표 3에 나타난 바와 같이 실내용담신호가 길어질수록 시간영역에서의 선형 복적분에 필요한 연산량이 급증하는 것을 알 수 있었다. 예를들면, 실내용담 신호의 샘플 수가 2048개 일때는 선형

표 4. 복적분 처리속도 실험 환경.

Table 4. The test condition for evaluating the processing time of convolution.

	오디오 신호	실내용답신호
녹음장소	무향실	장충체육관
표본화 주파수	44,100 Hz	44,100 Hz
재생시간	18.5 seconds	2.9 seconds
총샘플수	819,200	129,687
CPU 속도	Pentium III 500 MHz	

복적분 방법이 순환복적분 방법에 비하여 20배 이상의 곱셈연산이 필요함을 알 수 있었다.

본 논문에서 제안된 부단위 복적분 알고리즘의 연산량 감소 성능을 평가하기 위하여, 무향실에서 녹음한 오디오 신호와 장충체육관에서 취득한 실내용답신호를 복적분 하는데 소요되는 시간을 측정하는 실험을 수행하였다. 표 4는 실험에 사용된 두 음원의 특성과 사용된 컴퓨터의 사양을 기술한 것이다.

표 4에 기술한바와 같이 장충체육관의 실내용답신호의 샘플 수가 129,687개 이므로, 식 (1)을 이용하여 FFT 연산량이 최소가 되는 블록의 개수는 네개가 되며 한블록내의 샘플의 개수는 32,422개가 된다. FFT는 $2n$ (n 은 정수) 단위로 처리되므로 블록당 샘플 개수가 32,768개 일 때 FFT 연산량이 최소가 된다. 표 5는 실내용답신호를 부단위의 블록으로 분할하여 주파수 영역에서 복적분을 수행할 때 블록당 샘플개수에 따른 수행시간을 나타낸다. 식 (1)의 결과와 동일하게 블록당 샘플수가 32,768개 일 때 수행시간이 최소가 됨을 확인하였다. 또한, 입력 오디오 신호의 재생시간인 18.5초보다 빠른 14초만에 복적분 수행이 완료되었으므로 실시간 응용에도 적용 가능함을 확인하였다.

X. 결론

음장효과는 음향의 사실감과 현장감을 증대시킬 뿐만 아니라 거리감 및 방향감을 증대시켜 주므로, 실감음향을 재생하기 위해서 매우 중요한 요소이다. 또한, 멀티 채널 음향 재생 시스템의 급속한 보급에 따라, 오디오 기기에 대한 사용자의 요구사항이 오디오 데이터 압축 효율에서 현장감있는 음향공간 재현능력으로 집중되고

표 5. 처리속도 실험 결과.

Table 5. The results of convolution processing time test.

블록당 샘플수	수행시간
512	121 seconds
1,024	63 seconds
2,048	35 seconds
4,096	23 seconds
8,196	17 seconds
16,384	14 seconds
32,768	14 seconds
65,536	18 seconds

있으므로, 현장감 재현에 가장 큰 영향을 미치는 음장효과와 중요성이 강조되고 있다. 그러나, 음장효과는 사용자의 취향에 따라 효용성을 달리하며, 이를 재대로 처리하는데 많은 연산량을 필요로 하므로 보편적으로 적용되고 있지는 못하다.

본 논문에서는 음장에 대한 사용자의 취향을 기술하는 서술구조와 부단위 병렬 복적분 알고리즘 통한 고속 음장효과 처리기를 제안함으로써 오디오 콘텐츠를 사용자의 음장에 대한 취향에 따라 변환하여 제공하는 기술을 제안하였다. 음장효과가 처리된 오디오 콘텐츠에 대한 일반인들의 구별능력과 선호도를 조사한 결과 제안된 음장 선호도 서술구조가 다양한 어플리케이션에 응용 가능함을 확인하였으며, 본 논문에서 제안한 고속 복적분 알고리즘이 실시간 처리가 가능함을 실제 오디오 신호와 실내용답신호를 이용하여 확인하였다.

향후 과제으로써는 보다 정확한 음장을 표현하기 위하여 사용자의 재생환경을 보정하기 위한 기술에 대한 연구가 요구된다.

감사의 글

본 논문은 정보통신부의 연구사업인 "지능형 통합 정보 방송 기술 개발"과제의 일환으로 수행한 결과로서 정보통신부 담당자 및 관련 연구원들의 노력에 감사를 드립니다.

참고 문헌

1. ISO/IEC 21000-1, "Information Technology - Multimedia Framework. Part 1: Vision, Technologies and Strategy," International Standard, 2001.

2. Tim Kientzle, A Programmer's Guide to Sound, Addison Wesley, 1998.
3. ISO/IEC 21000-7, "Information Technology - Multimedia Framework, Part 7: Digital Item Adaptation," International Standard, 2004.
4. Udo Zolzer, *DAFX: Digital Audio Effects*, John Wiley & Sons, 2002.
5. ISO/IEC JTC1/SC29/WG11, Multimedia Description Schemes Group, "Additional Audio Description for MPEG-21 DIA," M8891, Oct., 2002.
6. ISO/IEC JTC1/SC29/WG11, Audio Group, "Report of the Core Experiment on Soundfield Description for Audio Presentation Preference in MPEG-21 DIA," M9173, Dec., 2002.
7. ISO/IEC JTC1/SC29/WG11, Audio Group, "Report of the Second Core Experiment on the Soundfield Description for MPEG-21 DIA," M9482, March 2003.
8. Ducksoo Lee and Kyeongmo Sung, "Sound Field Processor for Creating Virtual Concert Hall Impression in Home," IEEE Transactions on Consumer Electronics, 41(2), May 1995.
9. 강성훈, 강경욱, *입체음향*, 기전연구사, 1997.
10. Blauert J., *Spatial Hearing*, MIT Press, Cambridge Mass., 1983.
11. Eric D. Scheirer, Ritta Vaananen, and Jyri Huopaniemi, "AudioBIFS: Describing Audio Scenes with the MPEG-4 Multimedia Standard," IEEE Trans. on Multimedia, 1(3), Sep., 1999.
12. ISO/IEC 14496-1, "Information technology - Coding of audio-visual objects, Part 1: Systems," International Standard, 1999.
13. Alan V. Oppenheim and Ronald W. Schaffer, *Discrete-time Signal Processing*, Prentice Hall, 1999.
14. W. G. Gardner, "Efficient convolution without input-output delay," Journal of Audio Engineering Society, 43(3), 127-136, March, 1995.

• 홍재근 (Jae-Keun Hong)



1975년 2월: 경북대학교 전자공학과 공학사
 1979년 2월: 경북대학교 대학원 전자공학과 공학석사
 1985년 2월: 경북대학교 대학원 전자공학과 공학박사
 1979년 3월~1983년 2월: 경북전문대학 교수
 1983년 4월~현재: 경북대학교 전자전기공학부 교수
 *주관심분야: 음성신호처리, 음향신호처리

저자 약력

• 서정일 (Jeongil Seo)



1994년 2월: 경북대학교 전자공학과 공학사
 1996년 2월: 경북대학교 대학원 전자공학과 공학석사
 1996년 3월~현재: 경북대학교 대학원 전자공학과 박사과정
 1998년 3월~1999년 5월: LG반도체 근무
 1999년 6월~2000년 10월: 현대전자 근무
 2000년 11월~현재: 한국전자통신연구원 방송미디어연구그룹 선임연구원
 *주관심분야: 음성신호처리, 음성신호처리, 객체기반 3차원 오디오 신호처리, MPEG, DMB

• 강경옥 (Kyeong-Ok Kang)



1985년 2월: 부산대학교 물리학과 학사
 1988년 2월: 부산대학교 대학원 물리학과 석사
 1993년 3월~현재: 한국항공대학교 대학원 전자공학과 박사과정
 1991년~현재: 한국전자통신연구원 방송미디어연구그룹 3D미디어연구팀장 (책임연구원)
 *주관심분야: 디지털 방송, 에티데이터 신호처리, 음향신호처리, 객체기반 3차원 오디오 신호처리 기술