

컴퓨터 그래픽스 커뮤니티에 소개된 사운드 관련 연구들:

Part I. 사운드 합성과 공간화

유민준^o 이인권

연세대학교 컴퓨터과학과

debussy@cs.yonsei.ac.kr^o, iklee@yonsei.ac.kr

Sound Researches in Computer Graphics Community:

Part I. Sound Synthesis and Spatialization

Min-Joon Yoo^o In-Kwon Lee

Department of Computer Science, Yonsei University

요약

사운드는 가상현실 및 컴퓨터 애니메이션에서 사용자의 현실감 및 집중도를 높이기 위한 중요한 요소이다. 최근에 컴퓨터 그래픽스 분야에서도 연구자들이 사운드 관련 연구들에 관심을 가지고 여러 컴퓨터 그래픽스 학술지와 학회에 주목할만한 연구결과들이 발표되고 있다. 본 논문에서는 컴퓨터 그래픽스 분야에서 소개된 논문들을 중심으로, 컴퓨터 그래픽스에서 관심을 가지는 사운드 연구의 주제에 대하여 알아보고 중요성을 가지는 논문들을 소개한다. 특히 본 논문에서는 다음 두 가지 분야의 논문들을 소개한다. 첫째, 사운드를 물리적으로 합성함으로써 그래픽스의 물리 법칙에 동기화되는 사운드를 생성하는 것이다. 둘째, 음원에서 나오는 사운드가 사람에게 들리는 음향적 환경을 모델링하는 것이다. 사운드에 대한 연구에서 주된 관심은 이러한 음향학적인 물리 법칙을 더욱 효율적으로 모델링하여 적은 자원으로도 실제적인 사운드 효과를 발생시킬 수 있는지에 대한 것이다. 이러한 관점을 중점으로 다양한 사운드 관련 논문들을 살펴보고 컴퓨터 그래픽스와의 관계를 살펴보고자 한다.

Abstract

Sound is very important element to enhance and reinforce reality and immersion of users in virtual reality and computer animation. Recently, significant researches about sound modeling are presented in computer graphics community. In this article, main subjects are explained and major researches are reviewed based on the sound papers presented in computer graphics community. Specially, several papers about following two subjects are reviewed in this paper: 1) synthesizing sound using physically-based laws and generating sound synchronized with graphics. 2) spatializing sound and modeling sonic environment. Many research about sound modeling have been focused on more efficient modeling of real physical law and generate realistic sound with limited resources. Based on this concept, various papers are introduced and the relationship between researches about sound and graphics is discussed.

키워드: 사운드 연구, 사운드 합성, 사운드 공간화, 그래픽스 커뮤니티

Keywords: sound research, sound synthesis, sound spatialization, graphics community

1. 서론

가상 세계를 현실적으로 표현하는 것은 컴퓨터 그래픽스가 이루고자 하는 주된 목적 중 하나라고 할 수 있다. 그리고 그 가상 세계를 이루는 다양한 요소 중 그래픽스 못지 않게 중요하게 여겨지는 것은 사운드에 대한 것이다. 가상 세계에서 사운드는 존재감, 현실성을 더하기 위하여 존재하기 때문에, 이는 사실상 가상 세계에서 그래픽스의 역할과 동일하다고 할 수 있다. 여러 연구를 통하여 개발된 정확한 음향학 모델은 사용자에게 가상 현실에서 더욱 강한 실제감을 부여하고[1], 높은 수준의 사운드는 시각적 요소의 인식을 향상시킨다고[2] 보고되고 있다.

최근에 컴퓨터 그래픽스 분야에서도 주목할만한 수준의 사운드 관련 연구 결과들이 보고 되고 있다. 상대적으로 컴퓨터 그래픽스 학계에서는 사운드 연구에 대한 관심은 적은 편이었지만, 2000년대 이후로부터 컴퓨터 그래픽스 학계에서도 사운드의 중요성을 인식하는 연구자들의 수가 늘어나고 있고, 높은 수준의 연구 결과가 ACM SIGGRAPH를 비롯한 주요 학회 및 학술지에 소개 되고 있다.

본 논문에서는 주로 컴퓨터 그래픽스 학계에 소개된 사운드 관련 논문을 바탕으로, 사운드 관련 연구의 주된 주제들에 대하여 소개하며, 컴퓨터 그래픽스의 연구와의 연관성을 살펴보도록 한다. 특히 이번 논문에는 컴퓨터 애니메이션과 가상 환경에서 특히 중요성을 갖는 ‘사운드 합성’과 ‘사운드 공간화’에 대한 주제에 대하여 소개하도록 한다.

최근까지도 대부분의 가상환경에서 사운드를 생성하는 방법은 특정 사건이 발생할 때 녹음된 사운드를 이용하는 것이었다. 즉 예를 들어 미리 총알이 총에서 발사되는 사운드를 녹음한 후, 권총의 방아쇠를 당기는 사건이 일어났을 때 이 소리를 재생하는 것이다. 이는 가상세계에 대한 초창기 연구에서, 특정 사건이 일어났을 때 미리 정의되어있는 애니메이션 시퀀스를 이용하는 것과 유사하다고 볼 수 있다.

이러한 방법은 간단하고 빠르다는 명백한 장점이 있으나, 다음과 같은 두 가지 단점이 있다. 첫번째로 녹음된 소리는 계속 반복되어 재생되게 된다. 따라서 사건이 오랫동안 지속될 때나 혹은 비슷한 사건이 계속적으로 발생할 때에는 소리가 계속적으로 반복되게 되고, 이는 현실감을 떨어뜨리게 된다. 또한 단지 재생하는 소리를 이용하는 방법으로는 표현할 수 없는 많은 물리적 효과들이 있다. 예를 들어, 두 물체 사이에 충돌이 일어나면 사운드의 크기와 음색은 충돌력(impact force)의 크기와 방향에 따라 결정되게 된다. 만약 충돌에 대한 시나리오가 조금이라도 변화한다면, 사운드도 그 변화를 나타내야 한다. 이러한 미묘한 변화는 실제감을 높이는 데 큰 역할을 한다. 녹음된 소리를 이용하는 것의 두 번째 단점은 모든 사운드 현상에 대하여 원본 사운드를 녹음하는 것은 노력이 많이 요구되는 지루한 작업이라

는 것이다.

반면에 사운드를 물리적으로 합성하여 생성할 수 있다면, 충돌의 위치, 재질의 속성 및 물체의 지오메트리와 같은 요인으로 발생하는 미묘한 톤(tone)과 음색(timbre)의 변화를 표현할 수 있게 된다. 다만 사운드의 특성상, 사운드를 합성하는데 요구되는 비용은 상당히 크기 때문에 이를 인터랙티브하게 사용하는 것에 어려움이 있었으나, 최근 컴퓨터의 성능이 높아지고 사운드 합성의 속도를 높이는 다양한 연구 결과들이 나오에 따라 이를 실시간으로 수행하는 다양한 연구가 컴퓨터 그래픽스 학계에 소개 되고 있다.

사운드의 공간화, 즉 공간적인 사운드 렌더링은 컴퓨터 게임, 가상 환경, 자동차 시뮬레이션 등 인터랙티브한 가상 환경에서 현실감을 위한 매우 중요한 요소이다. 특히 이러한 애플리케이션에서 사용되는 장면의 복잡성은 최근 크게 증가하였다. 예를 들어 거리 장면에서 자동차들이나 스포츠 게임의 경기장 안에 관중들은 모두 잡음이나 사운드를 생성하는 물체들인데, 복잡한 장면에서 이러한 물체의 수는 매우 많다.

따라서 이러한 복잡한 장면에서 여러 사운드를 효율적으로 렌더링하는 것은 사운드 연구의 또다른 중요한 주제 중에 하나이다. 이 역시 컴퓨터의 성능 향상과, 사운드 렌더링 비용을 줄이는 효율적인 알고리즘의 개발로 인하여 최근 크게 발전하고 있다.

본 논문에서는 위에서 소개한 두 가지 주제에 대하여 소개한다. 2장에서는 사운드 합성에 대한 다양한 연구 결과들을 소개한다. 유한요소법(FEM) 기반 사운드 합성(2.1장) 및 모달 사운드 합성(2.2장), CFD 기반 사운드 합성(2.3장)에 대한 연구들이 소개된다. 3장에서는 사운드 공간화에 대한 연구 결과를 소개한다. 3.1장에서는 공간화 사운드 렌더링에서 소개하며, 3.2장에서는 사운드 렌더링 파이프라인에 대하여 설명한다. 4장에서 마지막으로 결론을 맺는다.

2. 사운드 합성

서론에서 언급했듯이 초창기 그래픽스 연구에서 사용하는 사운드는 주로 녹음된 사운드를 사용하여 생성하였다. Terzopoulos[3]는 찢어지는 천을 시뮬레이션을 할 때 발생하는 사운드를 생성하기 위하여, 미리 천이 찢어질 때 생성되는 사운드를 녹음한 후, 실제 시뮬레이션에서 천이 다 찢길 때까지 녹음한 사운드를 반복하여 재생하였다.

어떠한 물체의 표면에 탄성이 있을 때, 이 물체에 외부 충격이 가해지면 이 표면에 진동이 일어나게 된다. 이러한 진동은 주변 공기에 영향을 미쳐 압력에 대한 파동(pressure wave)을 발생시키게 된다. 이 파동이 만일 20 Hz에서 22,000 Hz 사이의 주파수를 갖는다면, 사람의 귀에서는 이를 감지하게 되고, 사람들은 사

운드에 대하여 인식을 하게 된다. 따라서 사운드 합성에 대한 연구들은 물체의 표면의 진동을 어떠한 방법을 사용하여 모델링하는지에 따라 나뉘어질 수 있다.

2.1 유한요소법 기반 사운드 합성

사운드를 물리적으로 합성하려는 기존 연구들은 컴퓨터 그래픽스 분야보다는 음향학 분야나 컴퓨터 음악 분야에서 더욱 활발히 진행되었다. 특히 이들은 주로 악기에서 생성되는 사운드를 정확히 모델링하는 것에 관심을 가지고 있었다[4]. 일반적으로 현악기나 피아노, 타악기 같은 악기 소리를 물리적으로 시뮬레이션하려는 다양한 시도가 주를 이루고 있었는데, 타악기의 시뮬레이션에서 사용되는 바(bar)나 막(membrane)에 대한 시뮬레이션을 유한요소법(finite element method)를 사용하여 해결하는 몇가지 연구결과가 있었다[5, 6, 7].

O'Brien[8]은 이와 유사한 개념을 이용하여 단단한 물체의 움직임에서 발생하는 사운드를 모델링하는 연구를 컴퓨터 그래픽스 학계에 소개하였고 이는 SIGGRAPH 2001에 소개되었다. 이 연구에서는 비선형적인 유한요소법을 이용한다. 외부 힘에 대한 물체의 반응을 모델링 하기 위하여 물체의 표면의 행동을 분석하고, 생성되는 사운드 주파수 중에서 귀에 들릴 수 있는 범위의 사운드 주파수를 뽑아 내기 위하여 몇가지 필터를 적용한다. 비선형 유한요소법을 사용하기 때문에 버클링(buckling)같은 비선형적인 행동을 모델링 할 수 있고 다양한 시뮬레이션 파라미터를 적용할 수 있다는 장점이 있지만, 이 방법은 매우 큰 계산량을 요구하며 공식화(formulation)와 구현이 복잡하다는 단점이 있다.

이 연구에서는 유한요소법과 같이 컴퓨터 그래픽스에서 널리 사용되는 기술을 사운드에 적용하는 예를 보이고 있지만, 컴퓨터 그래픽스와 사운드 간의 연구에는 고려해야 할 한가지 차이가 있는데, 그것은 둘 간의 타임스텝의 차이가 크다는 것이다. 컴퓨터 그래픽스 시뮬레이션은 30~60 프레임율을 가지면 높은 수준의 애니메이션을 생성할 수 있지만, 높은 수준의 사운드는 보통 44,100 Hz 이상의 샘플링율을 가진다. 거의 1000배 가까운 타임스텝의 차이가 있기 때문에, 유한요소법을 이용하여 물체의 사운드를 추출하는 것에는 큰 비용이 들게 된다. 이 연구에 따르면, 1500개가 안되는 표면 요소로 이루어진 물체에서 6분 가량의 사운드를 합성하는데 최대 26시간 이상이 소모되었다고 보고하고 있다.

최근에 SIGGRAPH ASIA 2009에 소개된 An의 연구[9]에서는 효율적인 디포메이션을 위한 큐버처(cubature) 최적화 방법을 제안하였다. 이 연구에서는 사운드 앰플리케이션에 대한 결과도 보고하고 있는데 이 최적화에 의해 사운드 생성 시뮬레이션 시간도 크게 단축되었음을 알 수 있다.

2.2 모달 사운드 합성

인터랙티브한 사운드 합성을 위하여 최근까지 가장 널리 사용되는 방법은 모달 사운드 합성(modal sound synthesis)이다. 컴퓨터 그래픽스 학계에 처음으로 진동 모드(vibrational mode)를 사용하여 물체의 충동을 모델링하는 방법은 van den Doel[10, 11]의 연구에서 찾아볼 수 있다.

모달 사운드 합성에서는 외부 힘에 대한 물체의 반응에 대한 해석적인 해(analytical solution)를 구한 후 이를 이용한다. 모달 합성에서 충동에 의한 물체의 반응(response)은 다음과 같다:

$$s(t) = \sum_k a_k e^{-\alpha_k t} \sin(\omega_k t), \quad (1)$$

여기에서 $s(t)$ 는 시간에 따른 신호이며, ω_k 와 α_k 는 각각 모드(mode) k 에 대한 주파수(frequency)와 감쇠율(decay rate)이다. 즉, 물체에 충동이 일어나면, 이 물체에서는 다양한 주파수를 갖는 모드가 발생하게 되고, 이들을 모두 더함으로써 사운드를 합성해내는 방법이다.

식 1에서 ω_k 와 α_k 는 실시간 시뮬레이션 이전에 전처리로 물체의 기하학적 정보를 이용하여 구할 수 있는 값들이다. a_k 값은 각 모드들의 크기를 나타내는 것으로, 실시간 시뮬레이션에서는 다양한 물리 법칙에 의하여 변화하는 이 값들만을 구하여 바로 사운드를 합성해 낼 수 있다. 또한 매 사운드 프레임마다 값을 계산하는 것이 아니라, 비디오 프레임 혹은 이보다 약간만 빠른 프레임율로 계산하여도 원하는 샘플링율의 사운드 샘플들을 생성할 수 있기 때문에 인터랙티브한 사운드 생성이 가능한 것이다.

van den Doel은 앞서 소개한 연구들에서 해석적인 해가 알려져 있는 접시같은 간단한 시스템에서 생성되는 사운드만을 연구하였지만, 그 이후에 그는 이러한 시스템을 비롯하여 더욱 복잡한 시스템에 대한 모달 사운드 합성 방법을 컴퓨터 그래픽스 학계에 소개하였다[12]. 이 논문도 SIGGRAPH 2001에서 발표되었다. 이 연구에서는 구르는(rolling) 사운드 등을 효율적으로 생성해내는 방법도 소개했다.

모달 사운드 합성에서 중요한 것 중에 하나는 주어진 물체의 모달 성분, 즉 각 모드의 주파수와 감쇠율을 구하는 것이다. van den Doel의 논문에서는 해석 해가 알려진 물체를 이용하거나, 그렇지 않은 물체인 경우에는 물리적인 측정을 통하여 이 값들을 구하였다.

같은 해에 SIGGRAPH에서 발표된 Pai[13]의 논문에서는 로보틱스 시스템을 이용하여 이 값들을 구하는 방법을 제안하였다 이는 이전 연구[14]를 확장한 것으로 정교한 마이크로폰을 이용하여 모드들의 파라미터를 추출하는 방법을 소개하고 있다. van den Doel은 이 방법을 이용하여 모달 파라미터를 구하였다.

SCA 2002에서 O'Brien[15]이 소개한 연구는 사운드를 모달

합성 방법으로 합성하지만, 더욱 복잡하거나 측정하기 힘든 물체들의 사운드도 생성해낸다. 이러한 물체들의 모달 파라미터를 구하기 위하여, 이 모델에 유한요소법 시뮬레이션을 적용한다. 이 시뮬레이션은 물체의 모양이 변하지 않는다면 처음에 한번만 적용한 후 파라미터들을 추출한 후 이를 계속 이용할 수 있다. 실시간에는 물리 시뮬레이션을 통하여 각 모드들의 크기만 구하여 사운드를 합성하게 된다.

그 이후에 Raghuvanshi[16]는 물체의 모달 파라미터를 추출하기 위하여 매스-스프링 모델을 이용한 연구를 발표하였다. 이 방법은 이전에 유한요소법 보다는 정밀하지 않은 방법이지만, 더욱 간단한 공식화와 구현으로도 일반적인 사운드를 생성하는 작은 스케일의 표면 진동을 적절히 모델링 할 수 있으며 해석적 해를 이용하여 풀릴 수 있음을 보였다.

컴퓨터 그래픽스의 다양한 연구에서 사람의 시각에 대한 한계성을 이용하는 것처럼, 사운드에서도 사람의 청각의 한계성을 이용하여 계산의 효율성을 높이고자 하는 연구가 진행되었다. 사람은 서로 가까운 주파수의 사운드를 구분하는데 제한적인 능력이 있다[17]. 즉 충분히 가까운 주파수에 있는 두 사운드가 연속적으로 들린다면, 평균적인 사람은 이 두 사운드가 다른 주파수의 사운드인지, 같은 주파수의 사운드가 두 번 반복되는지 구분할 수 없다는 것이다. 충분히 가까운 거리에 있는 여러 주파수가 있을 때, 이를 대표하는 하나의 주파수의 사운드로 대체해도 인식에는 큰 차이가 없다. 이 방법을 이용하여 보통 하나의 사운드를 표현하기 위하여 몇 천개의 모드가 요구되는 사운드를 몇 백개 범위의 모드를 이용하여서 표현할 수도 있다[16, 18].

그 외에도 여러가지 음향학적인 성질을 이용하여 사운드 합성 시 사용되는 모드의 개수를 줄이는 연구들이 수행되었는데, van den Doel[19, 20]에서는 주파수 마스킹 방법을 사용한다. 실시간에 방출되는 사운드 중에서, 이웃한 주파수의 크기가 클 경우 이에 가려져서 인식이 안되는 주파수의 사운드를 찾고 이를 합성하지 않는 방법이다. 이러한 마스킹 기법은 PAC(Perceptual audio coding)등에서도 사용하는 방법으로 오디오 압축 등에 사용되는 기술이다. PAC는 소위 mp3 라고 흔히 불리는 MPEG Layer 3 표준[21]을 포함한 MPEG 오디오 코딩에 일반적으로 사용되는 기법이다.

일반적인 물체의 모달 사운드를 시간 영역이 아닌 주파수 영역에서 나타내면 조밀하지 않은(sparse) 형태로 나타난다. Bonneel[22]은 이러한 성질을 이용하여 주파수 영역에서 적은 수의 푸리에 상수를 더함으로써 빠르게 모달 합성을 하는 방법을 SIGGRAPH 2008에 소개하였다. 이 연구에서는 모드에 대한 효율적인 STFT(Short-time Fourier Transform)을 유도하여 사용한다. 이 방법을 사용하여 일반적인 시간 영역에서의 모달 합성[23]에 비하여 5~8 배의 속도 향상을 얻을 수 있었다고 보고한다.

최근까지도 사운드 합성 방법은 모달 합성을 이용하는 방식이 주로 사용되고 있다. 이전 유한요소법 방식등과 비교하여 선형적인 움직임만을 모델링할 수 있기 때문에 사운드의 다양한 효과를 표현하기가 힘들다는 단점이 있지만, 이 방법은 실시간으로 사운드 합성이 가능하기 때문에 인터랙티브한 사운드를 생성할 수 있다는 큰 장점이 있기 때문이다.

2.3 CFD 기반 사운드 생성

대부분의 사운드 합성 연구에서는 진동하는 단단한 물체에 의하여 방출되는 사운드에 대한 연구가 주를 이루고 있지만, SIGGRAPH 2003에 소개된 Dobashi[24]의 연구는 공기 역학(aerodynamic)에 의해 발생하는 사운드를 생성하는 방법을 제시한다. 이 방법은 CFD(Computational Fluid Dynamics)의 분야에서 도출된 식을 바탕으로 사운드를 합성하는 방법으로, 칼을 휘두를 때의 사운드나, 바람 소리등을 시뮬레이션 할 수 있는 방법이다. 이 방법 역시 물체의 움직임이나 바람의 속도 등을 입력 받아 실시간에 공기 역학적인 사운드를 생성할 수 있다.

3. 사운드 공간화

사운드의 공간화는 컴퓨터 게임, 가상 환경, 자동차 및 비행기 시뮬레이션 등 인터랙티브 가상 환경에서 사용자에게 현실감을 부여하기 위한 매우 중요한 요소이다[25]. 이전 장에서는 사운드가 물리 법칙을 이용하여 어떻게 생성될 수 있는가에 대한 연구들을 소개했다면, 이번 장에서는 이렇게 만들어진 사운드가 우리가 사는 공간 안에서 어떻게 퍼지고 우리의 귀에 들리는지에 대한 연구들을 소개한다.

3.1 공간적 사운드 렌더링

사운드는 여러모로 빛과 비슷하다. 사운드와 빛은 모두 파동으로 전해지며, 호이겐스의 원리(Huygens' principle)에 따라 모든 방향으로 퍼져나간다. 그 둘은 매질이 연속적인 곳에서는 반사(reflect)하고 굴절(refract)한다. 또한 장애물을 만나면 회절(diffraction)한다.

하지만 이 둘은 다른 전파 속도와 파장을 가지고 있기 때문에 차이점도 존재한다. 빛은 거의 지연(delay)이 없이 즉시 전달되지만, 사운드의 상대적으로 느린 속도는 쉽게 지연을 발생시켜, 에코(echo)나 반향(reverberation) 효과를 만들어낸다. 또한 사운드는 빛과 달리 매우 쉽게 회절한다.

사운드의 이러한 효과는 이전부터 널리 이해되어 왔고, 콘서트홀 디자인 등에 실제적으로 중요하게 사용되기 때문에 음향학 분야에서 다양한 연구 결과들이 존재한다[26]. 하지만 사운드의 이러한 다양한 특성을 계산하는데는 엄청난 계산량이 요구되기

때문에, 인터랙티브한 가상 환경에서 이러한 요소들은 거의 무시되어 왔었다. 그러나 이러한 효과가 없다면 현실감은 크게 감소할 수 밖에 없다.

1990년대 후반에 이르러 비로서 실제적으로 이러한 사운드 현상들을 실시간으로 시뮬레이션하고자 하는 연구들이 발표되었다. Funkhouser와 그의 동료들은 컴퓨터 그래픽스 분야에서 이러한 음향학 모델링 연구의 선구자적인 역할을 하였는데, SIGGRAPH 1998년도에 소개된 논문[27]에서는 먼저 고정된 음원에서 높은 정확도를 갖는 반향과 사운드가 전해지는 경로를 계산하는 빔 트레이싱(beam tracing) 방법을 소개하고, 미리 모든 방향에서 가능한 반향과 전해지는 경로를 계산하여 데이터 구조에 저장한 후, 이 데이터 구조를 이용하여 실시간에 움직이는 청취자에 따라 반향 효과를 계산하는 방법을 제안하였다.

SIGGRAPH 1999년에 소개된 논문[28]에서는 이 연구를 확장하여 분산 가상 환경(Distributed virtual environment)에 적합한 알고리즘을 소개했다. 이러한 환경에서는 네트워크 등으로 연결되어있는 여러 사용자가 있기 때문에, 이들간의 인터랙션으로 발생하는 사운드 효과까지 고려해야 한다. 따라서 이 연구에서는 기존 방법의 속도를 크게 향상시킬 수 있는 음향 모델링을 위한 세 가지의 새로운 빔 트레이싱 방법을 제시한다. 그 이후에 Min[29]은 이 트레이싱 방법 중 하나인 프라이어리티 기반(priority-driven) 빔 트레이싱 알고리즘에 대한 더욱 발전된 연구를 발표하였다.

3D 가상 세계에서 실제 사운드 모델링에서는, 앞서 연구된 반향 및 직접음 외에도 주변 물체에 의하여 반사되는(reverberant) 사운드가 음원의 위치 파악과 환경에 대한 공간적 특성을 이해하는데 중요한 역할을 한다. 이전 연구들에서 회절은 무시되거나, 단지 통계적인 근사를 통하여 모델링 되는 수준이었다.

회절 현상은 사운드가 자신과 같은 주파수의 오더의 크기를 갖는 장애물에 의하여 분산(scattering)되는 효과를 의미하는 것으로, 빌딩의 인테리어나 도시등과 같이 음원과 청취자 사이가 가려져 있는 경우가 많을 때 더욱 중요시 여겨지는 현상이다. 이러한 환경에서는 음원이 보이지 않아도 사운드는 들려야 하는 상황이 많기 때문이다. 만일 이러한 회절 현상이 없다면, 가상환경에서 사용자가 코너를 돌거나 어떠한 물체가 눈앞에서 사라져버리면 들리는 사운드의 공간감에 급격한 변화를 느낄 수 있다. SIGGRAPH 2001년도에 소개된 Tsingos[30]의 논문은 UTD(Uniform Theory of Diffraction)[31]을 이용하여 이러한 회절 효과를 효율적으로 계산하는 방법을 제안하여 더욱 실제적인 음향학적인 모델링을 제안한다.

또 다른 관점으로, 실제적으로 진동하는 물체에서 생성되는 사운드는, 물체 내부에서 생성되는 복잡한 회절과 상호반사(interreflection)효과를 포함한다. 그 이유는 사운드는 빛과 달리 물체의 크기와 비교될만한 크기의 파장을 가지고 있기 때문

이다. 사람의 청각은 10kHz (3.4cm의 파장)에서 가장 민감하고, 20 kHz (2cm보다 작은 파장)까지 감지하기 때문에, 물체의 크기와 관련되는 주파수를 가지는 모든 물체의 지오메트리와 복잡하게 반응하게 된다.

이러한 효과는 컴퓨터 그래픽스에서 셀프-쉐도우잉(self-shadowing)이나 상호반사같은 글로벌 조명(global illumination)과 유사하다. 컴퓨터 그래픽스에서도 이러한 효과를 계산하는 것은 매우 비용이 많이 드는 일이었지만, Sloan[32]은 3D 물체의 선형적인 반응을 미리 계산함으로써 인터랙티브한 조명 렌더링 기법을 소개하였다.

James[33]은 SIGGRAPH 2006에서 소개된 논문에서, 이와 유사한 개념으로 복잡한 회절과 상호반사 효과까지 고려한 사운드를 생성하는 알고리즘을 제안하였다. 우선 진동하는 물체에서 모드를 계산한 후, 이를 각각 사운드 압력 필드(acoustic transfer function)와 연관시키고 이들을 다중극(multipole) 소스로 근사한다. 실시간에 이러한 다중극 소스들을 더함으로써 효율적으로 상호반사같은 효과를 시뮬레이션 할 수 있었다.

3.2 사운드 렌더링 파이프라인

컴퓨터 그래픽스 학계에서 사운드 렌더링 파이프 라인에 대한 초창기 연구는 Takala의 논문[34]에서 찾아볼 수 있다. 이 연구에서는 그래픽스 파이프라인과 동시에 진행되는 사운드 파이프 라인을 제안하였으며, 이를 이용하여 애니메이션과 사운드트랙과의 동기화를 이룰 수 있는 프레임워크를 제시하였다.

사운드 렌더링 파이프라인에 들어가는 비용을 줄이고자 하는 연구는 최근까지 그리 활발하게 연구되지 않았다. 초창기의 대부분 연구는 사운드 렌더링에 들어가는 필터 연산에 대한 연구들이었다. Marten과 Chen의 논문[35, 36]에서는 시그널 프로세싱 연산의 속도를 높이기 위하여 HRTF(Head Related Transfer Functions)의 PCA분석을 이용하였다.

사운드 파이프라인에서 렌더링되는 요소들의 수를 조절하는 개념은 Fouad[37]에서 처음 소개되었다. 사용가능한 자원의 양에 따라서 인식에 기반한 메트릭을 이용하여 중요한 사운드에 더욱 많은 사운드 샘플을 생성하게 된다. 이는 LOD(level-of-detail)방식으로 사운드를 프로세싱 하는 것으로도 볼 수 있다.

제한된 소프트웨어 및 하드웨어 상황에서 클러스터 방식을 이용하여 여러 음원을 렌더링하는 연구도 소개되었다[38]. 이는 컴퓨터 그래픽스에서 여러 조명에 있는 장면을 오프라인으로 렌더링할 때 사용되는 기술과도 유사하다[39]. 여기에는 사람의 청각의 위치 측정 능력에는 한계가 있다는 정신-음향학적 증거가 사용된다[40].

그 후에 몇가지 인터랙티브한 음향적 환경을 나타내려는 몇가지 시도가 있었지만[41, 42, 43], 수백개의 움직이는 음원이 있는

복잡한 가상 장면을 위한 실시간 3D 사운드 렌더링 파이프라인은 SIGGRAPH 2004년에 소개된 Tsingos[44]의 논문에서 본격적으로 제안되었다. 실시간으로 인식적으로 들리지 않는 음원들을 제거한 후, 남은 사운드 소스들을 작은 수의 클러스터로 그룹화한다. 이 때는 이전의 연구[38]보다 더욱 개선된 클러스터링 기술을 사용한다. 각 클러스터는 하나의 임포스터(imposter) 음원으로 표현된다. 사운드 프로세싱을 이 임포스터 음원에만 적용함으로써 계산 비용을 크게 줄이게 된다. 이 방법은 컴퓨터 그래픽스에서 사용하는 어클루전 컬링(occlusion culling)과 LOD방법과 유사한 측면이 있다[45].

위 논문에서는 사람의 청각의 한계를 이용한 결과를 사운드 렌더링에서도 사용하고 있다. 즉 모달 사운드 합성에서 마스킹 기법이 사용된 것처럼(2.2장 참고) 여러 사운드가 동시에 존재할 때 이들간의 마스킹 효과를 고려하여 렌더링 되는 사운드의 수를 줄이는 방법을 사용한다.

클러스터링을 이용한 사운드 렌더링의 또다른 연구[46]에서는 컴퓨터 그래픽스에서 사용하는 포인트 기반 렌더링을 사용하여 클러스터링으로 더욱 효율적으로 하는 방법을 제시한다.

이러한 기술들로 사운드 렌더링의 효율성이 크게 높아지긴 하였지만, 이 기술들은 모두 정적인 장면에만 적용되는, 즉 미리 모든 음원의 움직임 정보를 아는 상황에만 적용될 수 있는 기술이었다. Moeck[47]은 동적이며 더욱 많은 수의 음원들을 다루고, 더욱 효율적인 클러스터를 계산할 수 있는 계층적 클러스터링 알고리즘을 제안하였다.

이 연구에서는 유저 스테디를 이용하여 사운드 클러스터링을 할 때 시각적 요소의 영향에 대하여도 연구를 하였다. 뷰 프러스텀(view frustum) 안에 있는 음원이 더욱 중요성을 가질 것이라는 가정을 기반으로 새로운 매트릭을 제안하며 이 방법은 기존에 제안된 오디오 saliency 맵(audio saliency map)[48]과 같은 기술보다 더욱 뛰어난 결과를 가짐을 보였다.

3D 사운드 렌더링의 기술은 컴퓨터 그래픽스 기술과 함께 적용되어 가상 현실이나 애니메이션, 시뮬레이션에 이용되는 것이 주된 목적이다. 하지만 최근까지도 사운드 렌더링 과정에서 동시에 진행되는 시각적 요소가 사운드 렌더링에 미치는 영향에 대하여서 고려하지 않았다. 최근에 와서 위에서 소개한 Moeck[47]의 연구를 비롯하여 이러한 크로스모달(cross-modal)효과를 이용한 연구들이 나오고 있다.

SIGGRAPH 2008에 소개된 Bonneel[22]의 연구에서는 합성된 사운드와 녹음된 사운드를 함께 효율적으로 렌더링하는 방법을 제안한다. 사실 많은 애플리케이션에서 이 두 사운드는 함께 사용될 필요가 있다. 총소리같은 사운드는 충돌력등의 물리 정보에 따라 변화해야 하지만, 헬리콥터 소리나 주변에 잡음 소리는 녹음된 소리만으로도 충분히 시뮬레이션이 가능하기 때문이다.

만일 어떠한 장면에서 갑자기 많은 수의 합성된 파편음들을 발생시켜야하는 경우에는 짧은 순간에 많은 계산량이 요구되며, 이는 시스템의 자원을 한순간에 많이 써야되는 경우가 발생한다. Bonneel의 논문에서는 이러한 상황에서 자원을 효율적으로 스케줄링하기 위하여 일종의 크로스모달 정보를 이용한다. 여러 크로스모달 연구에서 시각적 사건과 음향적 사건의 시간적 불일치(asynchrony)에 대해서 논하는데, Guski[49]는 실험을 통하여 이러한 시간적 불일치의 허용값을 200 msec라고 보고하였다. Bonneel의 논문에서도 이 연구 결과를 이용하여 200 msec 이내의 사운드 지연을 허용하여 자원을 효율적으로 분배하는 방식을 제안하고 있다.

가장 최근에 발표된 Grelaud[50] 논문에서는 LOD를 정할 때 크로스모달 정보를 이용한다. 이 논문은 사용자가 재질(material)을 인지할 때 그래픽스와 사운드의 상호간의 영향을 연구한 결과[51]를 이용하여 BRDF로 렌더링할 때 사용하는 하모닉스의 밴드의 수와, 모달 사운드 합성을 할 때 사용하는 모드의 수를 효율적으로 줄임으로써 사운드 렌더링 파이프라인의 효율성을 높인다.

4. 결론

본 논문에서는 컴퓨터 그래픽스 분야의 학술지와 학회에 소개된 사운드 관련 논문을 중심으로 이들의 주된 주제와 내용을 간략하게 소개하였다. 특히 본 논문에서는 가상현실 및 애니메이션에서 특히 중요한 '사운드 합성'과 '사운드 공간화'에 대한 주제에 대하여 알아보았다.

사운드 합성은 최근 모달 사운드 합성 기법의 큰 발전으로 인하여 실시간으로 물리적으로 변화하는 사운드를 빠르게 합성할 수 있게 되었다. 합성되는 사운드의 수준과 합성에서 사용되는 자원은 트레이드-오프(trade-off)가 있으므로, 더욱 효율적으로 합성이 이루어질 수록 사운드의 수준도 높아지게 된다. 따라서 기술이 발전할 수록 더욱 현실성 있는 사운드가 생성될 수 있리라 기대한다. 그리고 현재는 충돌 사운드 등의 사운드에 대한 연구가 중심을 이루고 있는데, 앞으로는 더욱 다양한 사운드를 물리적으로 생성하는 알고리즘도 개발될 것이다.

사운드 공간화는 다양한 음향적 효과를 효율적으로 계산하는 알고리즘이 개발되고, 이를 더욱 효과적으로 처리할 수 있는 사운드 렌더링 기법이 개발됨에 따라 매우 많은 음원이 존재하는 장면에서도 현실감 있는 사운드를 렌더링할 수 있게 되었다. 그리고 최근 크로스모달 효과를 이용하여 효율성을 높이는 연구 결과가 발표되고 있다. 사운드와 그래픽스의 크로스모달 효과는 최근까지도 컴퓨터 그래픽스와 사운드 분야에서 크게 연구되지 않고 있는 주제이긴 하지만, 신경과학(neuroscience)의 많은 연구[52, 53]에서는 최근까지도 이러한 다양한 효과를 제시하고

있다. 따라서 이러한 연구들을 바탕으로 하여, 공통점과 차이점이 동시에 많은 컴퓨터 그래픽스와 사운드 분야는 서로에게 도움을 주면서 앞으로 계속 발전해 나갈 것이다.

감사의 글

본 연구는 문화체육관광부 및 정보통신연구진흥원의 IT원천기술 개발사업(2008-F-031-01)의 연구결과로 수행되었음

참고 문헌

- [1] N. I. Durlach and A. S. Mavor, *Virtual reality scientific and technological challenges*. National Academy Press, 1995.
- [2] R. L. Storms, "Auditory-visual cross-modal perception phenomena," Ph.D. dissertation, Naval Postgraduate School, Monterey, California, 1998.
- [3] D. Terzopoulos and K. Fleischer, "Deformable models," *The Visual Computer*, vol. 4, no. 6, pp. 306–331, 1988.
- [4] P. R. Cook, *Real sound synthesis for interactive applications*. Natick, MA: A.K.Peters. Ltd., 2002.
- [5] I. Bork, A. Chaigne, L.-C. Trebuchet, M. Kosfelder, and D. Pillot, "Comparison between modal analysis and finite element modeling of a marimba bar," *Acoustica united with Acta Acustica*, vol. 85, no. 2, pp. 258–266, 1999.
- [6] J. Bretos, C. Santamaria, and J. A. Morál, "Finite element analysis and experimental measurements of natural eigenmodes and random responses of wooden bars used in musical instruments," *Applied Acoustics*, vol. 56, pp. 141–156, 1999.
- [7] F. Orduña-Bustamante, "Nonuniform beams with harmonically related overtones for use in percussion instruments," *Journal of the Acoustical Society of America*, vol. 90, no. 6, pp. 3582–3583, 1991.
- [8] J. F. O'Brien, P. R. Cook, and G. Essl, "Synthesizing sounds from physically based motion," in *Proc. ACM SIGGRAPH*, 2001, pp. 529–536.
- [9] S. S. An, T. Kim, and D. L. James, "Optimizing cubature for efficient integration of subspace deformations," in *Proc. ACM SIGGRAPH Asia*, 2009.
- [10] K. van denDoel and D. K. Pai, "Synthesis of shape dependent sounds with physical modeling," in *Proc. the International Conference on Auditory Displays*, 1996.
- [11] K. van denDoel and D. K. Pai, "The sound of physical shapes," *Presence*, vol. 7, no. 4, pp. 382–395, 1998.
- [12] K. van denDoel and D. K. Pai, "Foley automatic: physically-based sound effects for interactive simulation and animation," in *Proc. ACM SIGGRAPH*, 2001, pp. 537–544.
- [13] D. K. Pai, K. van denDoel, K. James, J. Lang, J. E. Lloyd, J. L. Richmond, and S. H. Yau, "Scanning physical interaction behavior of 3d objects," in *Proc. ACM SIGGRAPH*, 2001, pp. 87–96.
- [14] J. L. Richmond and D. K. Pai, "D.k. robotic measurement and modeling of contact sounds," in *Proc. the International Conference on Auditory Displays*, 2000.
- [15] J. F. O'Brien, C. Shen, and C. M. Gatchalian, "Synthesizing sounds from rigid-body simulations," in *Proc. ACM SIGGRAPH Symposium on Computer Animation*, 2002, pp. 175–181.
- [16] N. Raghuvanshi and M. C. Lin, "Interactive sound synthesis for large scale environments," in *Proc. ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 2006, pp. 101–108.
- [17] A. Sek and B. C. Moore, "Frequency discrimination as a function of frequency, measured in several ways," in *Audio Anecdotes*, 2003.
- [18] N. Raghuvanshi and M. C. Lin, "Physically based sound synthesis for large-scale virtual environments," *IEEE Computer Graphics and Application*, vol. 27, no. 1, pp. 14–18, 2007.
- [19] K. van denDoel, D. Knott, and D. K. Pai, "Measurements of perceptual quality of contact sound models," in *Proc. the International Conference on Auditory Displays*, 2002, pp. 345–349.
- [20] K. van denDoel, D. Knott, and D. K. Pai, "Interactive simulation of complex audiovisual scenes," *Presence: Teleoperators and Virtual Environments*, vol. 13, no. 1, pp. 99–111, 2004.
- [21] E. M. Painter and A. S. Spanias, "A review of algorithms for perceptual coding of digital audio signals," in *Proc. Digital Signal Processing*, 1997, pp. 179–208.
- [22] N. Bonneel, D. Drettakis, N. Tsingos, I. Viaud-Delmon, and D. James, "Fast modal sounds with scalable frequency-domain synthesis," in *Proc. ACM SIGGRAPH*, 2008.

- [23] K. van denDoel and D. K. Pai, "Modal synthesis for vibrating objects," in *Audio Anecdotes*, 2003.
- [24] Y. Dobashi, T. Yamamoto, and T. Nishita, "Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics," in *Proc. ACM SIGGRAPH*, 2003, pp. 732–740.
- [25] P. Larsson, D. västfjäll, and M. Kleiner, "Better presence and performance in virtual environments by improved binaural sound rendering," in *Proc. AES 22nd International Conference on virtual, synthetic and entertainment audio*, 2002, pp. 31–38.
- [26] H. Kuttruff, *Room Acoustics, 3rd Edition*. London: Elsevier Science, 1991.
- [27] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. West, "A beam tracking approach to acoustic modeling for interactive virtual environments," in *Proc. ACM SIGGRAPH*, 1998, pp. 21–32.
- [28] T. Funkhouser, P. Min, and I. Carlbom, "Real-time acoustic modeling for distributed virtual environments," in *Proc. ACM SIGGRAPH*, 1999, pp. 365–374.
- [29] P. Min and T. Funkhouser, "Priority-driven acoustic modeling for virtual environments," *Computer Graphics Forum*, vol. 19, no. 3, 2000.
- [30] N. Tsingos, T. Funkhouser, T. Ngan, and I. Carlbom, "Modeling acoustics in virtual environments using the uniform theory of diffraction," in *Proc. ACM SIGGRAPH*, 2001, pp. 545–552.
- [31] D. A. McNamara, C. W. I. Pistorius, and J. A. G. Malherbe, *Introduction to the uniform geometrical theory of diffraction*. Artech Houser, 1990.
- [32] P. P. Sloan, J. Kautz, and J. Snyder, "Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments," in *Proc. ACM SIGGRAPH*, 2002, pp. 527–536.
- [33] D. L. James, J. Barbic, and D. K. Pai, "Precomputed acoustic transfer: Output-sensitive, accurate sound generation for geometrically complex vibration sources," in *Proc. ACM SIGGRAPH*, 2006, pp. 987–995.
- [34] T. Takala and J. Hahn, "Sound rendering," in *Proc. ACM SIGGRAPH*, 1992, pp. 111–220.
- [35] W. Martens, "Principal components analysis and resynthesis of spectral cues to perceived direction," in *Proc. International Computer Music Conference*, 1987, pp. 274–281.
- [36] J. Chen, B. V. Veen, and K. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *Journal of the Acoustical Society of America*, vol. 97, pp. 439–452, 1995.
- [37] H. Fouad, J. Hahn, and J. Ballas, "Perceptually based scheduling algorithms for real-time synthesis of complex sonic environments," in *Proc. the International Conference on Auditory Displays*, 1997.
- [38] J. Herder, "Optimization of sound spatialization resource management through clustering," *Journal of Three Dimensional Images, 3D-Forum Society*, vol. 13, no. 3, pp. 59–65, 1999.
- [39] E. Paquette, P. Poulin, and G. Drettakis, "A light hierarchy for fast rendering of scenes with many lights," *Computer Graphics Forum*, vol. 17, pp. 63–74, 1998.
- [40] B. C. Moore, *An introduction to the psychology of hearing*. Academic Press, 4th edition, 1997.
- [41] J. Fouad, J. Ballas, and D. Brock, "An extensible toolkit for creating virtual sonic environments," in *Proc. the International Conference on Auditory Displays*, 2000.
- [42] E. Wenzel, J. Miller, and J. Abel, "A software-based system for interactive spatial sound synthesis," in *Proc. the International Conference on Auditory Displays*, 2000.
- [43] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *Journal of the Audio Engineering Society*, vol. 47, no. 8.
- [44] N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual audio rendering of complex virtual environments," in *Proc. ACM SIGGRAPH*, 2004, pp. 249–258.
- [45] T. Funkhouser and C. Sequin, "Adaptive display algorithms for interactive frame rates during visualization of complex virtual environments," in *Proc. ACM SIGGRAPH*, 1993, pp. 247–254.

- [46] M. Wand and W. Strasser, "Multi-resolution sound rendering," in *Proc. Symposium on Point-Based Graphics*, 2004.
- [47] T. Moeck, N. Bonneel, N. Tsingos, G. Drettakis, I. Viaud-Delmon, and D. Aloza, "Progressive perceptual audio rendering of complex scenes," in *Proc. ACM SIGGRAPH Symposium on interactive 3D Graphics and Games*, 2007, pp. 189–196.
- [48] C. Kayser, C. Petkov, M. Lippert, and N. Logothetis, "Mechanisms for allocating auditory attention: An auditory saliency map," *Current Biology*, vol. 15, pp. 1943–1947, 2005.
- [49] R. Guski and N. Troje, "Audiovisual phenomenal causality," *Perception and Psychophysics*, vol. 65, no. 5, pp. 789–800, 2003.
- [50] D. Grelaud, N. Bonneel, M. Wimmer, M. Asselot, and G. Drettakis, "Efficient and practical audio-visual rendering for games using crossmodal perception," in *Proc. ACM SIGGRAPH Symposium on interactive 3D Graphics and Games*, 2009.
- [51] N. Bonneel, C. Suied, I. Viaud-Delmon, and G. Drettakis, "Bimodal perception of audio-visual material properties for virtual environments," *ACM Transactions on Applied Perception (Accepted with minor revisions)*, 2009.
- [52] W. Hairston, M. Wallace, J. V. Stein, J. Norris, and J. Schirillo, "Visual localization ability influences cross-modal bias," *Journal of Cognitive Neuroscience*, vol. 15, pp. 20–29, 2003.
- [53] D. Alais and D. Burr, "The ventriloquism effect results from near-optimal bimodal integration," *Current Biology*, vol. 14, pp. 257–262, 2004.

〈저자 소개〉



유 민 준

2004년 2월 아주대학교 정보및컴퓨터공학부 (학사)
2006년 2월 연세대학교 컴퓨터과학과 (석사)
2006년 3월 ~ 현재 연세대학교 컴퓨터과학과 박사과정
〈관심분야〉 컴퓨터 음악과 사운드 프로세싱, 멀티미디어 프로세싱등



이 인 권

1989년 2월 연세대학교 전산과학과 (학사)
1992년 2월 포항공과대학교 컴퓨터공학과 (석사)
1997년 2월 포항공과대학교 컴퓨터공학과 (박사)
1997년 9월 ~ 1999년 2월 오스트리아 Vienna University of Technology 박사후연구원
1999년 3월 ~ 2001년 2월 포항공과대학교 정보통신연구소 선임연구원
2001년 3월 ~ 2003년 8월 아주대학교 미디어학부 교수
2003년 8월 ~ 현재 연세대학교 컴퓨터과학과 교수
〈관심분야〉 컴퓨터 애니메이션, 기하 모델링, 컴퓨터 음악등