

사이버음향(CyberSound) - 가상세계의 음향

김형교*, 이의택**

한신대학교 정보통신학과 교수*, 한국전자통신연구원 영상정보처리연구실 실장·책임연구원**

요약

컴퓨터 음향발생에 관한 연구는 컴퓨터 음악, 인간-컴퓨터 상호작용, 데이터 청각화 등의 분야에서 오랫동안 진행되어 왔지만, 최근 들어 컴퓨터 애니메이션이나 가상세계등에서 시각적 효과와 함께 보다 입체감 있고 현실감 있는 가상환경을 제공하기 위해 더욱더 중요한 문제로 떠오르고 있다. 지금까지 음향발생을 위해 음향의 모델링이나 합성등 음향 자체에 대한 요소 기술들에 관해서는 많은 연구가 진행되었으나 컴퓨터 애니메이션, 가상세계등과 같이 영상내 동작 내지 사건과 음향이 서로 밀접하게 연관된 분야에서 필수적인 음향을 영상내 동작과 통합 처리할 수 있는 기술에 대한 연구는 초보적인 단계에 머무르고 있다. 최근 들어, 음향의 입체감과 임장감을 강화하기 위하여 3차원 음향이라는 개념이 도입되고 있고 이의 구현에 대한 연구가 활발하게 진행되고 있다. 여기서는 컴퓨터 애니메이션이나 가상현실등에서 영상내 물체의 움직임이나 사건 그와 동기된 음향의 자동생성 및 이의 3차원 음향효과 발생 기술 원리를 사이버음향(CyberSound)이라는 개념으로 묶어서 소개하면서, 이의 전망을 기술하고자 한다.

1. 서론

실세계에서 음향은 주어진 환경 속에서 물체의 동작에 따라 발생되며 그 물체의 성질이나 동작을 일으키는 사건과 환경조건에 따라 그 음향특성이 결정된다. 컴퓨터 그래픽스, 컴퓨터 애니메이션, 가상현실 등의 가상세계에서도 이런 특성을 고려하여 음향을 영상내 물체의 동작과 통합하여 처리해야만 더욱더 실감나게 묘사할 수 있게된다. 또한 영화, 게임, 광고, 만화 및 방송같이 영상과 음향의 통합처리가 필수적인 분야에서도 제작과 수

정이 용이하게 될 수 있을 것이다.

지금까지는 가상세계 생성기술에서 음향은 영상내 물체의 움직임과는 독립적으로 생성된다. 따라서 우리는 서로 관련이 없는 두 세계로부터 만들어진 영상내 물체의 움직임과 음향을 보고 듣는 셈이 되며 그 결과 몰입감은 반감된다. 또한 수작업 애니메이션이나 컴퓨터 애니메이션에서 영상부분을 먼저 제작한 후 나중에 음향을 영상에 동기시켜 부가하는 방식을 채택해 오고 있으나 이 방법은 상당한 시간과 고도의 숙련을 요하는 수작업으로 제작을 어렵게 하는 요인이 되고 있다.

가상현실이나 컴퓨터 애니메이션에서 영상내 물체의 동작과 관련된 음향을 효과적으로 생성하기 위해서는 음향의 모델링, 동기화, 그리고 렌더링이라는 3단계의 과정을 필요로 하며 최종적으로 사용자로 하여금 몰입감을 주기 위해서는, 생성된 음향에 3차원 효과를 줄 수 있어야 한다. 3차원 효과를 더한 음향이란 좌우의 파노라마 효과를 주는 기존의 스테레오와는 달리 음원의 위치를 전후 좌우 상하등 입체적으로 구별할 수 있도록 처리된 음향을 뜻한다. 음향의 모델링은 컴퓨터 음악 분야에서 (1-5) 오랫동안 연구가 되었지만 가상현실과 컴퓨터 애니메이션 분야에서 중요한 것은 영상내 물체의 동작으로 인하여 생성된 매개변수가 음향의 매개변수와 적절하게 대응될 수 있도록 음향의 모델링을 효과적으로 매개변수화 시키는 것이다. 음향의 동기화는 사용자 인터페이스 (6,7) 및 데이터 청각화 (data sonification) (5) 분야에서 많은 연구가 행해져 왔다. 음향의 렌더링(8)이란 각각의 음향모델을 이용하여 복합적인 음향을 생성하는 과정을 뜻하는데 이것은 마치 컴퓨터 그래픽에서 기하학적인 모델로부터 영상내 움직임을 생성하는 과정과 유사하다. 음향의 렌더링과 관련된 문제들은 음향학이나 신호처리 분야에서 많이 연구되었다. 3차원 음향은 주로 사람의 두 귀에 음원으로부터 음향이 도달할 때 각 전달경로를 디지털 필터로 표시한 머리전달함수 (Head-Related

Transfer Function, HRTF) (14)를 이용하여 구현하는 방향으로 연구가 진행 중이다.

이상에서 살펴본 바와 같이 음향의 모델링, 동기화, 그리고 렌더링이라는 각 단계에 대하여 독자적인 연구는 상당히 진행되어 왔으나 컴퓨터 애니메이션이나 가상현실등 가상세계 구축을 위한 기술에서 필수적인 영상내 물체의 동작과 음향을 통합적으로 처리 할 수 있는 방법에 대한 연구는 이제 시작에 불과하다. 본고는 컴퓨터 애니메이션이나 가상환경에서 응용될 수 있는 영상내 물체의 움직임과 그와 동기된 음향의 자동생성 기술 및 3차원 음향발생 기술의 원리와 그 요소 기술들, 그리고 앞으로 전망 등을 사이버 음향(CyberSound)이라는 개념을 통하여 소개하고 기술한다.

II. 사이버음향

음향의 생성은 컴퓨터 음악분야에서 많은 연구가 진행되어 왔다. 대부분의 컴퓨터 음악 시스템은 여러 음향 신호를 더하고 빼고 곱할 수 있는 소프트웨어 모듈을 서로 연결하여 음향을 발생시킨다. 이러한 시스템은 상당한 융통성이 있으나 가상환경이나 컴퓨터 애니메이션에 응용하기에는 두 가지 중요한 결점이 있다. 첫째, 컴퓨터 음악 시스템에서 모델링하는 음향은 모두가 음악적이기 때문에 가상환경에서 일어나는 모든 사건에 대한 음향은 모델링 할 수가 없으며 둘째, 여기서 생성된 악기의 매개변수는 음악적인 연주를 지향하고 있으므로 임의의 가상 환경에서 물체나 환경특성에 관련된 음향을 모델링 하기 어렵다.

현재 가상환경에서 음향은 대부분 실제음향을 표본화하거나 미디 (MIDI) 장치를 이용하여 합성한다. 이런 접근방식은 음향이 영상의 동작과 상관 관계가 있도록 매개변수화 시키기가 용이하지 않다. 물론 가상환경에서 실제음향의 이용이 차지하는 역할은 있지만 아주 제한된다. 여기서 음향은 강약이나 고저 등에 의하여 매개변수화 되는데 이것을 물체의 동작과 대응시키기는 것은 힘들다. 또한 미디 장치는 특정한 하드웨어를 이용하여 제한된 종류의 음향만을 합성 할 수 있으므로 일반적인 임의의 음향발생이 요구되는 환경에는 응용할 수 없다.

컴퓨터 음악에서는 MUSIC V, Csound, cmusic, Fugue(1-4)와 같은 기능적 구성방법이나 푸리에 합성, 신호의 곱셈, 필터링 그리고 주파수 변조와 같은 방법이 이용되고 있으나 일반적인 음향을 그것의 매개변수가 영

상의 동작에 대응되도록 하는 문제는 고려하지 않고 있다. 이것은 동작과 음향을 동기 시키는데는 필수적인 문제이다.

사이버음향은 VR(virtual Reality)에서 가상세계 내에서 발생하는 음향으로 임의의 음향이 자동적으로 발생 되는 것이 요구되는 환경에 맞게 만들어진 음향자동발생 시스템에서 만들어지는 음향으로 다양한 음향이 매개변수 조절에 의해 생성되고 합성될 수 있는 음향 합성 모델링 기술과 영상내 물체의 움직임이나 사건과 동기되어 자동적으로 발생하는 기술 그리고 이를 보다 현실감 있게 표현하도록 하는 3차원 음향효과 기술들이 통합되어야 만들어 질 수 있다. 우선 음향 자동 생성 모델에 관한 연구동향을 보면, 이때까지의 연구 중에서 영상내 물체의 동작과 그와 관련된 음향을 통합적으로 고려하여 매개변수화 시키는 시도로서 대표적인 것으로 미국 조지 워싱턴대의 팀버 트리(9)를 들 수 있다. 이것은 일반적인 음향을 기능적 구성방법에 의하여 매개변수화 시킨다.

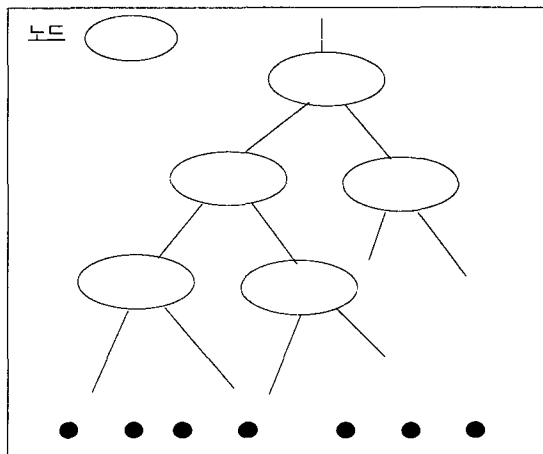


그림 1. 팀버트리 구조

1. 팀버트리

팀버트리는 영상합성에서 음영트리(shade tree) (10)과 유사하다. 음영 트리는 영상을 합성하는데 다양한 음영과 무늬의 통합을 허용하는 기능적 구성방법으로 트리 구조의 장점은 모듈성과 다양한 기술들을 간단하게 통합하는데 있다. 트리의 노드는 벡터의 내적이나 정규화와 같은 연산을 나타낸다. 각 노드는 음영을 조절하는 변수에 대한 연산을 행하여 새로운 변수를 생성한다. 이와

같이 하여 트리전체에 대한 연산이 끝나면 최종 색깔이 구해진다. 팀버트리도 이와 유사하다. 팀버트리의 노드는 다른 팀버트리에 대한 연산을 행하는데 각 팀버트리는 음향이나 매개변수를 나타낸다. 하나의 팀버트리 전체에 대하여 계산을 마치면 특정한 음색을 가진 음향이 생성된다. 객체지향 용어로 기술하면 매개변수가 주어진 팀버트리는 음향 클래스(class)의 추상화로 볼 수 있다. 따라서 트리 클래스의 라이브러리와 기본노드를 이용하여 새로운 클래스를 만들 수 있다. 주어진 트리의 매개변수를 시간에 따라 변화 시키면 팀버트리는 시간에 따라 변하는 음향을 생성한다. 그림 1에 팀버트리의 구조 예를 보인다.

2. 팀버트리의 구조

팀버트리의 각 노드는 다음중 하나를 나타낸다.

- 상수
- 동작제어 시스템으로부터 주어진 매개변수 (예: 시간, 기본주파수, 각 모멘텀 등)
- 디지털로 표본화된 음향
- 다변수 수학함수
- 일련의 상수로 구성된 벡터
- 라이브러리상의 다른 팀버트리

더하기, 빼기, 곱하기, 나누기, 지수함수, 로그함수와 같은 기본적인 수학 함수뿐만 아니라 여러 가지 기본과

1. 기본과의 생성

2. 잡음 톤 생성

3. 벌의 동작에 따라 평균 주파수를 정의하는 서브트리

4. 초기화

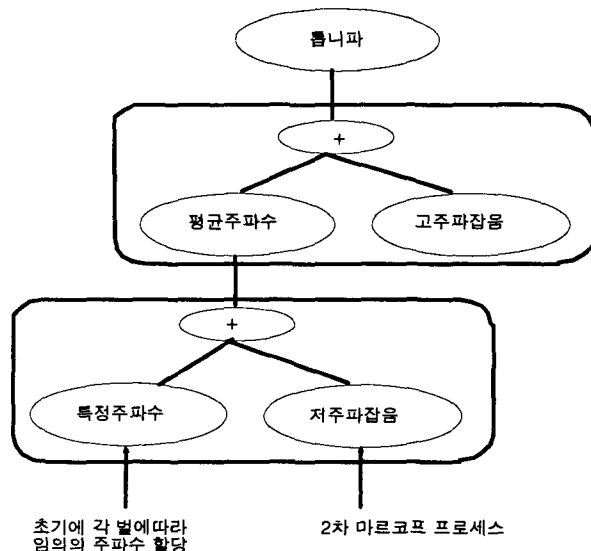


그림 2. 벌의 날개음향에 대한 팀버트리 구성과정

형(톱니파, 삼각파 등), 잡음(백색 잡음, Perlin 잡음 [11] 등), 신호처리함수(필터링, 컨볼루션 등)와 같은 특수함수들도 구현가능 하다.

시간영역에서 팀버트리의 계산은 공간영역에서 음영 트리의 계산과 유사하다. 사운드 트랙의 각 표본에서 주어진 매개변수들로 팀버트리를 계산하면 트리의뿌리값은 그 표본 점에서의 음향이 된다. 만약 팀버트리가 음향 표본 노드를 갖지 않으면 재표본 과정은 할 필요가 없다.

3. 팀버트리의 생성

컴퓨터 그래픽에서 모든 물체와 무늬를 모델링 하는데는 여러 방법이 있듯이 음향도 마찬가지이다. 물체의 진동음향을 나타내는데는 물리적 법칙이 이용되며 그 외에도 푸리에 합성법, genetic algorithm등 이 있는데, 아직까지는 설계자의 경험과 창의력을 바탕으로 한 시행 착오방법에 많이 의존한다.

그림 2는 행동양식적으로 제어되는 (behaviorally-controlled) 벌의 날개소리를 팀버트리를 이용하여 합성하는 과정을 보인다. 우선 벌이 어떤 방식으로 날개소리를 발생시키는지를 보자. 벌은 날개를 아래로 천천히 그리고 위로 빠른 속도로 움직여서 소리를 내므로 어떤 특정한 주파수의 톱니 파를 가정 할 수 있다. 이렇게 발생된 소리는 기계적인 단조로운 소리가 되는데 이를 개선하기 위해 주파수를 편이 시킬 필요가 있다. 이 과정은 어떤

평균주파수에 고주파 잡음을 더한 것으로 모델링 할 수 있다. 더욱이 벌의 날개소리는 시간에 따라 계속적으로 변하는데 위에서 설정한 평균주파수가 변하지 않는다면 이를 충분히 모델링 할 수가 없다. 따라서 우리는 평균 주파수를 특정한 주파수와 2차 마르코프 프로세스에 의한 저주파 잡음의 합으로 모델링 한다. 이렇게 생성된 특정한 매개변수를 가진 트리는 한가지의 음향을 나타낼 수 있지만 트리구조 자체는 음향의 전체 클래스를 대표 할 수 있다. 예컨대 앞에서 모델링한 트리구조는 벌의 날개소리, 여러 종류의 곤충의 날개소리를 비롯하여 전 기톱 소리등에 응용 될 수 있으며 매개변수를 어떻게 정하느냐에 따라 특정한 벌의 특정한 날개소리를 나타낼 수 있다. 이 예는 음향의 생성과 표현 시스템의 목적을 제시한다. 즉 음향의 팀버트리는 애니메이터에 있어서 음향이 어떻게 생성되는가 하는 기본적인 개념으로부터 구축되기 때문에 음향의 매개변수와 그 음향을 발생하는 영상의 동작 매개변수를 용이하게 대응시킬 수가 있다. 중요한 것은 단순히 음향을 생성하는 것이 아니라 영상의 동작과 대응이 용이하도록 음향을 매개변수화 하는 것이다.

팀버트리 구성에 있어 간단하지만 일반적인 것 중의 하나가 푸리에 합성법이다. 이것은 원하는 음향을 얻기 위하여 여러 주파수 성분을 더하는 것인데 특히 기타의 현과 같은 진동음향을 생성하는데 효과적이다. 더 복잡한 진동음향을 생성하기 위해서는 물리적 법칙들이 이용

되며 이 방법의 장점은 음향의 매개변수와 물리적인 속성을 쉽게 대응시킬 수 있다는 점이다. 이 것은 바람에 움직이는 차임 벨 소리의 모델링 등에 이용될 수 있다.

영상 렌더링에서 Phong shading과 같은 발견적 해결방법에 의한 음영 계산법은 물리적으로 정확하지 않지만 실제 경우에 무리 없이 적용 할 수 있다. 음향 렌더링에서도 마찬가지로 물리법칙을 엄격하게 적용하지 않고 발견적 해결방법을 이용하여도 듣는 사람은 차이를 별로 느끼지 못한다. 따라서 대부분의 팀버트리를 구성할 때 기본적인 물리법칙을 바탕으로 하되 매개변수는 발견적 해결방법을 이용하여 구한다.

충돌, 미끄러짐, 굴림, 마찰과 같은 물체간의 물리적인 접촉에 의하여 발생하는 음향도 팀버트리 구조로 모델링 할 수 있다. (12) 충돌음향은 푸리에 합성법에 의하여 구할 수 있는데 진동형태가 아주 복잡하기 때문에 다 음과 같은 통계분포함수로 표시되어야 한다.

$$signal(t) = \sum_{i=1}^n e^{-c\omega_i t} \sin(\omega_i t)$$

여기서 ω_i 는 랜덤 주파수를, 그리고 c는 댐핑항을 나타내며 이 식에 해당하는 팀버트리를 그림 3에 보인다.

여기서 주파수 성분의 개수 (n_freq), 랜덤 주파수의 범위 (base 및 max), 댐핑 계수 (damp_ratio) 등은 매개변수로 주어진다. rvector는 주어진 범위 내에서의 임의의 값을 가진 벡터를 발생하고 각 주파수 성분의 감

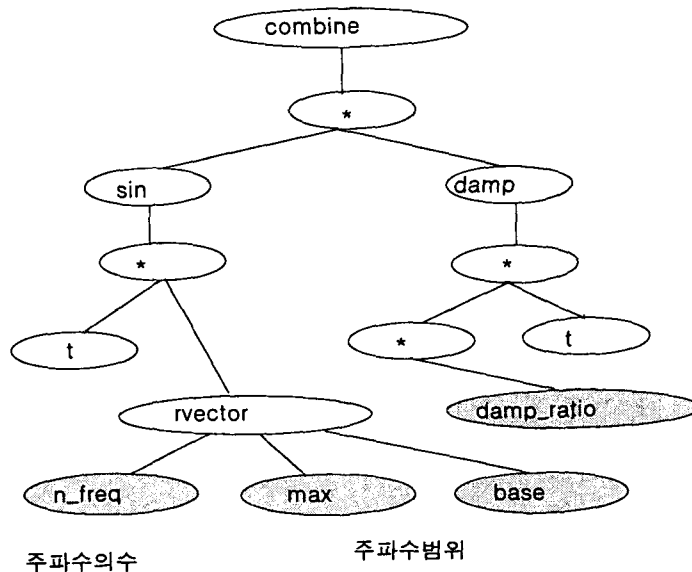


그림 3. 충돌음향의 팀버트리

쇠 비율은 damp 노드의 값에 따라 정해진다. combine 노드는 모든 벡터를 더하여 스칼라값을 계산한다. 각 매개변수는 음향을 야기하는 사건과 물리적 속성에 대응하므로 서로 다른 매개변수의 조합은 주어진 클래스 내(이 예에서는 충돌음)에서 목재, 금속음향등 다양한 종류의 음향을 발생할 수 있다.

이러한 발견적 해결 방법은 어느 정도의 숙련을 요구한다. 사용자가 일단 몇 가지 종류의 음향을 발생하는 방법을 개발한 후 어떤 특정한 음향을 발생하기 위해 무슨 종류의 함수들을 어떻게 결합시켜야 하나를 이해한다면 새로운 종류의 음향발생을 위한 팀버트리의 설계는 훨씬 용이할 것이다.

많은 경우 발견적 해결방법에 의하여 팀버트리를 구성할 때 설계자의 의도와는 다른 음향을 얻는 경우가 있는데, 예를 들어 벌의 날갯짓 음향을 원할 때 실제로는 모기의 날갯짓 음향이 발생될 수 있다. 또, 올바른 정량화를 행하지 않고도 원하는 음향을 발생시킬 수 있는 팀버트리를 구할 수도 있으며 원하는 음향을 발생한 후 그 클래스에 속하는 다른 종류의 음향을 조사해 볼 수도 있다. 이런 경우 genetic programming(GP)(13)이라고 하는 유전자 알고리즘이 이용된다. GP는 컴퓨터 보조 탐색으로 유용한 컴퓨터 프로그램이나 공식을 구하기 위해 재생, 변이, 선택 등을 이용하는 진화론적 개념을 도입한 최적화 방법이다. GP를 이용하여 주어진 팀버트리의 매개변수들을 변화시켜 한 종류의 클래스 내에서 여러 가지의 음향을 발생시킬 수도 있고, 같은 매개변수들로 서로 다른 클래스의 음향을 나타내는 완전히 다른 팀버트리를 구성할 수도 있다.

Ⅲ. 사이버음향과 동작의 동기

영상내 물체의 동작과 음향의 연결은 '동기'에 의하여 이루어진다. 동기는 언제 음향이 발생되어야 하는 타이밍 문제뿐만 아니라 동작의 매개변수와 음향의 매개변수를 어떻게 사상시키느냐 하는 문제를 모두 해결해야 한다. 일단 음향 클래스가 정의되고 그것의 매개변수가 팀버트리에 의하여 결정되면 동기는 음향 클래스의 매개변수와 동작제어 시스템이 생성한 동작 매개변수를 결합하여 각각의 음향을 발생시킬 때 행해진다. 여기서 동작제어 시스템이란 컴퓨터 애니메이션이나 가상환경에서 물체 및 사용자의 위치를 제어하고 갱신하는 시스템을 뜻하는데 일반적으로 동작은 사용자의 몸짓이나 운동,

physically-based 모델링에 의한 시뮬레이션, behavioral 모델링, 키프레임 같은 동역학 등에 의하여 일어난다. 동작에 따라 음향이 연속적으로 변하는 경우 시간 종속 변수를 도입하며 이 경우 시간에 따라 변하는 신호를 정의하는 값은 각 애니메이션 프레임에서 동작 제어 시스템에 의하여 계산된다. 음향의 표본 주파수(8000-44100Hz)는 영상(30Hz)보다 훨씬 높으므로 애니메이션 프레임 사이의 시간 종속 변수 값을 계산하기 위해서 보간법이 이용된다.

애니메이션이나 동작제어 시스템이 음향 영역에서 현재 어떤 음향이 발생하는가를 기술하기 위해서 Sonic Scene Description Language (SSDL) (9)이 이용된다. 이것은 RenderMan과 같은 장면 묘사 언어와 유사한 것으로 일단 작성되면 최종 사운드 트랙을 얻기 위해 렌더링 시스템에 공급된다. 이렇게 하여 SSDL로 표시된 동기과정은 렌더링 과정과 독립적으로 행해진다.

SSDL은 팀버트리를 이용한 음향 클래스의 정의, 각 음향의 발생, 발생된 음향의 시작 및 종료, 시간종속변수의 정의 및 그 변수들의 키프레임값 등에 대하여 기술한다. 뿐만 아니라 SSDL은 마이크의 위치, 방향, 속성과 음원, 그리고 3-D에서 다른 물체등 음향 렌더링에서 필요한 물리적인 물체의 데이터도 표현한다. 여러 시각 장면 묘사언어와 같이 SSDL은 구문 및 기능 인터페이스를 모두 포함하기 때문에 음향 장면은 C++을 이용하여 프로그램으로 나타낼 수도 있고 기능 인터페이스를 이용하여 동작제어 시스템으로부터 바로 출력 받을 수 있다.

음향을 SSDL로 나타내면, 음향의 종류를 지정할 수 있고 동작 제어를 행하는 장치와 직접 결합 가능하다. 그림 3은 SSDL로 표현된 음향들이 physically based 애니메이션 시스템에서 어떻게 이용되는가를 보여준다. 충돌음향의 클래스를 정의하기 위하여 그림 2에서 보인 팀버트리가 사용되는데 두 입방체가 서로 다른 물체라 할 지라도 (하나는 금속 다른 쪽은 목재) 음향 발생을 위하여 동일한 팀버트리가 이용된다. 동작 제어 장치가 충돌을 감지하면 충돌음향 발생이 시작된다. 충돌하는 물체의 재질에 따라 적절한 팀버트리의 매개변수들이 선택되며 동작 제어 장치가 계산한 충돌력에 따라 충돌음향의 크기가 결정된다. 이때 시간종속변수들도 독립변수로 공급된다.

SSDL은 음향발생과정에서 일련의 팀버트리를 생성한다. 이 팀버트리들은 각 음향에 따른 매개변수들을 나타내는 일련의 노드를 제외하고는 구조가 동일하다. 시간종속변수는 키값의 보간을 행하는 키프레임 노드로 표시되

며 팀버트리의 뿌리에는 언제 음향이 발생되고 중지되어야 하는가를 나타내는 스위치 역할을 하는 타이머 노드가 첨가된다. 이 트리들은 나중에 렌더링과정에서 서로 결합된다.

IV. 사이버음향 렌더링

음향 렌더링이란 각 음향물체로 부터 복합적인 사운드트랙을 형성하는 과정을 말하는 것으로 영상 렌더링이나 텍스춰(texture) 매핑과 방법적으로 유사하다. 즉, 최종적인 음향을 발생시키기 위하여 음향에 일련의 변환을 행하는데 여기에는 물체공간으로부터 마이크공간으로의 변환, 청각화, 주의환경 효과 등이 있다. 물체공간으로부터 마이크공간으로의 음향 변환은 물체와 마이크의 거리 때문에 발생하는 음향의 감쇠와 지연에 의한 것으로 팀버트리에서 감쇠/지연 노드로 표시할 수 있다.

가상환경에서 청각화 분야에 있어서는 상대적으로 많은 연구가 행해졌는데 그 중에서 Finite Impulse Response (FIR) 필터를 이용한 HRTF의 시뮬레이션(14)이나 단순한 심리음향학적 원리(15)를 이용한 몇몇

방법은 이미 성공적으로 실현되었다. 이 모든 방법의 공통점은 가상환경에서 음원의 위치에 따라 음향신호를 필터링 하기 때문에 청음장치의 지향성 효과는 그것이 스테레오 지향성 마이크이든 HRTF필터로 나타낸 귀든 단순히 팀버트리의 노드로 취급 할 수 있다는 것이다.

주위환경 효과도 팀버트리의 노드로 나타낼 수 있다. 여기에는 발견적 해결방법이나 복잡한 음향이론등이 이용되며 한 방법은 음향 에너지를 추적하는 것이다. 이것은 광선 추적법과는 약간 다른데 그 이유는 음향과 빛의 성질이 서로 다르기 때문이다. 각 물체는 음향 추적을 위하여 특별히 설계된다. 예를 들어 각 물체의 반사, 회절, 투과는 3차원 양방향 반사율 분포 함수(Bidirectional Reflectance Distribution Function, BRDF)로 있으며 이것은 물체의 모양과 재질에 의하여 고유하게 결정된다.

이상과 같은 노드들을 각 음향을 나타내는 팀버트리에 부가하여 그 트리들을 결합하면 전체적인 음향은 하나의 팀버트리로 표시할 수 있다. 즉 이 트리는 음향 리시버에 의하여 감지되는 음향을 나타내며 그 트리가 주어진 매개변수 값으로 계산되고 표본화되면 최종적인 사운드트랙을 발생한다. 앞서서도 설명한 바와 같이 시간

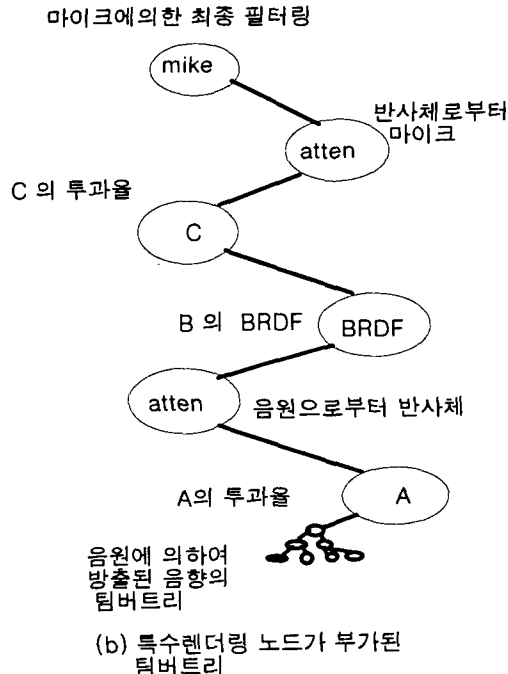
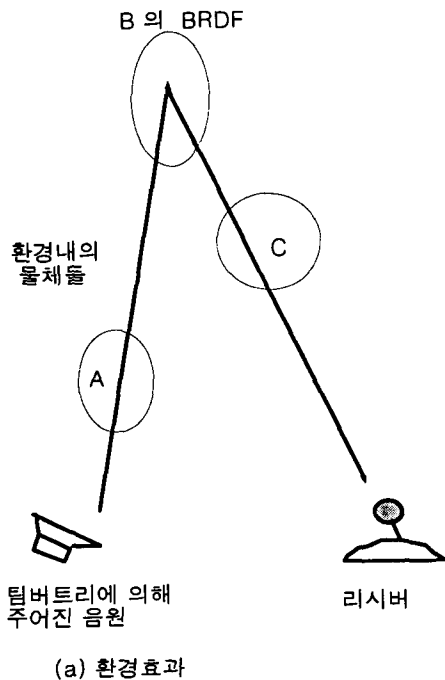


그림 4. 팀버트리에 부가된 환경노드

영역에서 팀버트리를 계산하는 것은 공간영역에서 세이드 트리를 계산하는 것과 아주 유사하다. 팀버트리의 뿌리의 출력은 주어진 표본 점에서의 계산된 음향이 된다. 그림 4는 렌더링 하고자하는 음향을 기술할 때 위에서 기술한 특수 노드가 어떻게 사용되는가를 보인다.

V. 3차원 음향 효과 - 사이버 음향 생성의 마지막 단계

1. HRTF

일반적으로 실세계에서 음향은 스테레오가 아닌 모노이다. 그러나 사람은 두 개의 귀를 가지고 있기 때문에 2채널 스테레오처럼 들린다. 음향이 사람의 귀에 도달하면 뇌와 귀가 음원을 구별한다. 즉, 귀의 모양과 위치, 어깨, 그리고 몸통이 음향과 상호작용을 하게되는데 이를 바탕으로 사람의 귀와 뇌는 서로 결합하여 귀에 도달하는 음향의 음원 위치를 정확히 알아내는 것이다. 이런 과정을 자세히 살펴보도록 하자.

· 좌우의 구별

사람이 긴 복도의 중앙에 서있을 때 오른쪽에 있는 문이 닫힌다고 하자. 이때 발생하는 음향은 왼쪽 귀보다 오른쪽 귀에 먼저 도달한다(이것을 Interaural Time Delay, ITD 라고 한다). 또한 그 음향의 세기는 오른쪽 귀쪽이 약간 더 강할 것이다 (이것을 Interaural Intensity Difference, IID 라고 한다). 이런 두 가지 정보를 바탕으로 뇌는 왼쪽이 아니라 오른쪽 문이 닫혔다고 판단하게 된다. 사실 사람의 뇌와 귀는 이런 판단에 아주 능숙하여 눈을 감고도 쉽게 한다.

· 전후의 구별

사람의 귓바퀴(pinna)는 귀에 도달하는 음향에 대한 필터역할을 하여 뇌가 음원의 전후 여부를 구별하도록 한다.

· 상하의 구별

귓바퀴에 의하여 변형된 음향은 음원의 높이를 구별하는 데도 역할을 한다. 음원의 높이는 고막에 도달하는 음향에 나타난 주파수 패턴에 따라 구별된다. 귓바퀴의 비대칭적인 결의 반사에 의하여 고유의 높이에 의존하는 간섭 패턴이 생성되며 음원의 높낮이에 따라

간섭 패턴이 달라진다. 사람의 뇌는 이 간섭 패턴을 해석하여 음원의 높이를 구별한다.

이상과 같이 사람의 몸체, 귀바퀴, 어깨, 머리등에 의하여 음원으로부터 발생된 음향이 여러형태의 반사와 회절을 함에 따라 고막에 도달하는 음향의 주파수 성분은 음원의 공간적인 위치에 따라 본래의 그것과는 다른 형태를 가지게 된다. 즉 동일한 음원으로부터 음향이 발생될 때라도 그 음원의 위치에 따라 사람의 고막에 도달하는 음향은 고유의 방향지향성 필터를 통과했다고 볼 수 있으며 이것을 바탕으로 음원의 위치를 구별하는 것이다.

음원과 각 귀사이의 방향지향성 필터는 주파수 응답 특성으로 나타낼 수 있으며 이것을 HRTF라고 한다. HRTF의 모델링에 대하여 많은 연구가 행해졌으며 특히 MIT Media Lab.에서는 더미 헤드(dummy head)를 이용하여 그 주위의 463곳에서의 HRTF를 구하였다 [16].

2. 3차원 음향 발생

3차원 음향이란 좌우의 파노라마 효과를 주는 기존의 스테레오와는 달리 음원의 위치를 전후 좌우 상하등 입체적으로 구별할 수 있도록 처리된 음향을 뜻한다. 이 3차원 음향을 발생시키기 위해서 이때까지는 Dolby Surround AC-3과 같이 주로 여러개의 스피커와 다중 채널방식으로 구현하였으나 이런 방식은 시스템의 공간 구성 및 경제적 부담이 큰 단점이 된다. 따라서 HRTF를 이용하면 헤드폰이나 두 개의 스피커 만으로 3차원 음향을 재생할 수 있게 되어 3차원 게임, 가상현실, 영상의 시스템등에서 시각적 효과와 함께 보다 입체감 있고 현실감 있는 환경을 저렴하고도 간단하게 구현해 줄 수 있게 된다. HRTF를 이용하여 3차원 음향을 재생하기 위해서는 많은 양의 디지털 필터 계산을 필요로 하는데 현재 계산량을 줄이기 위한 여러 알고리즘이 개발되고 있으며, 또한 프로세서의 속도향상에 따라 곧 소프트웨어로 구현 가능할 것이다.

VI. 결론

컴퓨터 음향발생에 관한 연구는 컴퓨터 음악, 인간-컴퓨터 상호작용, 데이터 청각화 등의 분야에서 오랫동안

안 진행되어 왔지만 최근 들어 컴퓨터 애니메이션이나 가상환경등에서 시각적 효과와 함께 보다 입체감 있고 현실감있는 가상환경을 제공하기 위해 더욱더 중요한 문제로 떠오르고 있다. 지금까지 음향발생을 위해 음향의 모델링이나 합성등 음향 자체에 대한 요소 기술들에 관해서는 많은 연구가 진행되었으나 컴퓨터 애니메이션, 가상세계등과 같이 영상내 동작내지 사건과 음향이 서로 밀접하게 연관된 분야에서 필수적인 음향을 영상내 동작과 통합 처리할 수 있는 기술에 대한 연구는 초보적인 단계에 머무르고 있다. 최근 들어, 음향의 입체감과 입장감을 강화하기 위하여 3차원 음향이라는 개념이 도입되고 있고 이의 구현에 대한 연구가 활발하게 진행되고 있다. 따라서 본 고에서는 컴퓨터 애니메이션이나 가상현실등에서 영상내 물체의 움직임이나 사건 그와 동기된 음향의 자동생성 및 이의 3차원 음향효과 발생 기술 원리를 사이버음향이라는 개념으로 묶어서 소개하면서, 이의 전망을 기술하였다.

영상합성에서 이용되고 있는 음영트리와 유사한 개념인 팀버트리는 음향을 매개변수 화하여 표현 합성하기 때문에 동작의 매개변수와 동기 시킨다면 음향과 동작의 통합처리에는 상당히 유리하다. 실제로 몇몇 상황들은 구현되었기 때문에 이런 분야에서의 응용가능성을 제시해 주지만 해결되어야 문제들도 많다. 첫째 앞에서도 기술 한바와 같이 팀버트리를 구성함에 있어 여러 물리적인 법칙들을 이용하는데 실제로 발생된 음향은 사용자가 의도했던 바와는 다를 수가 있다. 뿐만 아니라 올바른 정량화를 행하지 않고도 원하는 음향을 발생시킬 수도 있기 때문에 팀버트리의 구성은 설계자의 능력과 시행착오법에 의존하는 경우가 많다. 둘째 구성된 팀버트리의 연산은 상당한 계산을 요하기 때문에 실시간에서 처리하기 위해서는 다중 프로세서에 의한 새로운 연산방식의 개발이 필수적이다.

한편 좌우의 파노라마 효과를 주는 기존의 스테레오와는 달리 음원의 위치를 전후 좌우 상하등 입체적으로 구별할 수 있도록 처리된 3차원 음향이라는 개념이 도입되어 3차원 게임, 가상현실, 영상회의 시스템등에서 시각적 효과와 함께 보다 입체감 있고 현실감 있는 환경을 저렴하고도 간단하게 구현 해줄 수 있는 길을 열어주었다. 특히 HRTF를 이용하면 헤드폰이나 두 개의 스피커만으로도 3차원 음향효과를 효율적으로 발생시킬 수 있는데, 여기에는 많은 양의 디지털 필터 계산을 필요로 하기 때문에 이를 줄이기 위해 여러 알고리즘이 제안되고 있다. 이와 함께 프로세서의 속도향상에 따라 머지않

아 소프트웨어로 구현 가능할 것이다.

사이버음향이라는 개념은 컴퓨터 애니메이션이나 가상현실등에서 영상내 물체의 움직임이나 사건 그와 동기된 음향을 자동생성하여 여기에 3차원 음향효과까지 줄 수 있는 전과정을 하나로 통합처리하기 위한 첫 시도이다. 앞으로의 연구방향은 동작과 음향의 더 완전한 결합을 보장 해 줄 수 있는 음향의 모델링 및 매개변수화와 3차원음향을 실시간에서 소프트웨어로 용이하게 구현할 수 있도록 진행되어야 할 것이다.

참 고 문 헌

- [1] M. Mathews, *The Technology of Computer Music*, MIT Press, MA, 1969.
- [2] B. Vercoe, *Csound: A manual for the Audio Processing System and Supporting Programs*, MIT Media Lab, MIT, MA, 1986.
- [3] F. Moor, *Element of Computer Music*, Prentice Hall, Englewood Cliffs, NJ., 1990.
- [4] R. Dannenberg, C. Fraley, and P. Velikonj, "Fugue: A Functional Language for Sound Synthesis," *IEEE Computer*, vol. 24, no. 7, pp. 36-42, Jul. 1991.
- [5] C. Scaletti, "The Kyma/Platypus Computer Music Workstation" in *The Well-Tempered Object: Musical Applications of Object Oriented Software Technology*, Stephen Travis Pope, ed. MIT press, 1991.
- [6] M. Blattner, D. Smikawa, and R. Greenburg, "Earcons and Icons: Their Structure and Common Design Principles," *Human-Computer Interaction*, vol. 4, no 1, pp. 11-44, 1989.
- [7] W. Gaver, "Synthesizing Auditory Icons," *Proc. of INTERCHI*, 1993.
- [8] T. Takala and H. Hahn, "Sound Rendering," *Proc. of SIGGRAPH'92, ACM Computer Graphics*, vol. 26, no. 2, pp. 211-220
- [9] J. Hahn, J. Geigel, J. W. Lee, L. Gritz, T. Takala, and S. Mishra, "An Integrated Approach to Sound and Motion," *J. of Visualization and Computer Animation*, vol. 6, no. 2, pp. 109-123
- [10] R. Cook, "Shade Trees," *Proc. of SIGGRAPH'84, ACM Computer Graphics*, vol. 18, no. 3, pp. 195-206
- [11] K. Perlin, "An Image Synthesizer," *Proc. of SIGGRAPH'85, ACM Computer Graphics*, vol. 19, no. 3, pp. 287-296
- [12] J. Hahn, "Realistic Animation of Rigid Bodies," *Proc. of SIGGRAPH'88, ACM Computer Graphics*, vol. 22, no.3, pp. 299-308
- [13] J. Koza, *Genetic Programming*, MIT Press, Cambridge, MA, 1992.
- [14] E. M. Wenzel, "Localization in Virtual Acoustic Displays," *Presence: Teleoperators and Virtual Environments*, vol. 1, pp. 80-107, 1992.

- [15] S. T. Pope and L. E. Fehlen, "The Use of 3-D Audio in a Synthetic Environment: An Aural Renderer for a Distributed Virtual Reality," *Proc. IEEE VRAIS'93*, pp. 176-182
- [16] W. G. Gardner, "Transaural 3-D audio", *MIT Media Lab.*

Technical Report no. 342, 1995.

- [17] 김형교, 이의택, "3차원 애니메이션을 위한 음향자동생성기술동향", *전자통신 동향분석*, 제11권 제4호, 1996.12. 한국전자통신연구원

필자소개



김형교

- 1956년 4월 16일생
- 1978년 2월 서울대학교 전기공학과 졸업, 공학사
- 1980년 2월 서울대학교 전자공학과 졸업, 공학석사
- 1983년 3월 ~ 1985년 8월 인덕전문대학교 전자과 교수
- 1993년 7월 ~ 1995년 2월 한국전자통신연구원 멀티미디어 통신연구실 선임연구원
- 1993년 3월 Georgia Institute of Technology, School of Electrical Eng., Ph. D.
- 1995년 3월 ~ 1997년 2월 상명대학교 정보과학과 교수
- 현재 한신대학교 정보통신학과 교수
- 주관심 분야 : DSP, 가상현실



이의택

- 1978년 서울대학교 공과대학 공업교육학과 전자전공(학사)
- 1982년 서울대학교 대학원 전자공학과(석사)
- 1996년 KAIST 전기 및 전자(박사)
- 1980년 2월 ~ 현재 한국전자통신 연구원 책임연구원(영상정보처리 연구실)
- 주관심 분야 : 디지털 스튜디오, 디지털 라이브러리, 3차원 애니메이션, 실감통신