

MDCT 를 이용한 HRTF 신호처리 계산량 감소방법

박만호, 최송인

한국전자통신연구원 이동통신연구소
이동서비스연구부 이동 A/V 연구팀

Complexity Reduction of HRTF Processing using MDCT

Manho Park, Songin Choi

Mobile A/V Research Team

Mobile Telecommunication Research Laboratory

Electronics and Telecommunications Research Institute

E-mail : mano@etri.re.kr

Abstract

최근 게임, 멀티미디어 콘텐츠, 가상현실 등을 제공하는 멀티미디어 장치에서 2 개의 스피커나 헤드폰을 이용하여 3 차원 입체 음향효과를 내고자 하는 가상 입체음향 기법에 관한 연구가 많이 이루어지고 있다. 가상 입체음향 기법 중 다채널 입체음향을 이용하여 입체 음향효과를 가진 2 채널 출력을 구성하는 방식의 대표적인 것으로는 입체음향 정위효과를 내는 HRTF(Head Related Transfer Function)이 있다. 그러나 이 방법은 전달 함수와 입력 신호의 컨볼루션(Convolution)방법을 이용함으로써 계산량이 복잡하여 실시간 구현에 많은 제약을 주고 있다.

본 논문에서는 다채널 입력으로 현재의 DVD, HDTV 등에서 사용되는 Dolby-Digital 복호화기를 사용하여 입체음향 시스템을 구현할 경우 가상입체음향 신호처리에 사용되는 HRTF 계수를 MDCT 계수로 구성하여 이를 통한 가상입체음향 신호처리방식의 계산량 감소방법에 대하여 설명하고 있다.

I. 서론

이동용 혹은 무선용으로 1 인용 멀티미디어 기기의 보급이 늘어남에 따라 사용자가 선택할 수 있는 서비스 콘텐츠의 경우에서도 게임과 영상을 중심으로 3 차원

입체음향을 제공하는 경우가 늘고 있으며, 이러한 1 인용 이동 혹은 무선 멀티미디어 기기의 특성상 음향신호의 재생방법으로 헤드폰과 같은 재생기기의 이용이 그 중요성을 더해가고 있다.

일반적으로 입체 음장감을 구성하기 위하여 사용되는 방법으로는 채널의 확장이 있는데 널리 사용되고 있는 예로는 DVD(Digital Video Disc)와 HDTV(High Definition TV)등에서 사용되는 AC-3 5.1 채널 시스템이 있다. 그러나 무선 혹은 이동용 기기의 경우 단순할 채널의 확장이라는 방법을 통한 입체음향의 제공이 어려우며 이에 따라 가상입체음향 신호처리방법을 이용한 방법이 널리 사용된다[1]. 다채널 음향신호, 특히 5.1 채널 음향신호를 2 채널 신호로 변환하는 방법으로는 Dolby-Digital 오디오 코덱에서 제공되는 Down-Mix 방식과 같이 정해진 규칙에 따라 각 채널에 일정한 가중치를 가한 후 가중치가 곱해진 각각의 채널 신호들을 더하여 구성하는 매트릭스 방식이 있으나 이러한 매트릭스 방식은 실제 입력이 가지는 위치정보에 대한 특성은 무시되며 채널간 간섭이 존재한다는 단점이 있다. 이에 비하여 HRTF 방법은 원음의 위치정보를 보존하여 입체음장감에서 매트릭스 방식에 비해 우수한 성능을 보이지만 가상입체음향 신호처리에 필요한 HRTF DB를 위한 메모리가 필요하고, 계산량이 많다는 점이 실시간 구현에 큰 제약으로 지적되고 있다.

II. HRTF 알고리즘

HRTF는 자유음장(Free Field)에서 일정한 입사각으로 사람의 외이도(Ear Canal)에 이르는 소리를 표현하는 전달함수를 말하며, 가상 현실 응용시스템이나 가청(Auralization) 시스템의 바이노럴(Binaural)신호 합성에 이용된다. 소리가 사람의 두 귀에 전파되는 과정에서 특정한 방향에 대응하는 필터링을 거치게 되는데 이 필터링을 지각함으로써 청취자는 음원의 방향을 지각할 수 있게 된다. 이 필터링은 사람의 몸통, 머리 그리고 귀바퀴로부터의 반사와 회절에 기인하며 HRTF에 의해 포괄적으로 설명된다[1][2]. 바이노럴 신호의 합성에 있어 필터링은 보통 Convolution의 형태로 수행되므로 이를 위해서는 HRTF가 시간영역의 데이터, HRIR(Head-Related Impulse Response)의 형태로 주어진다. 임펄스 응답의 측정은 보통 더미헤드(Dummy head)를 사용하여 무반향실(Anechoic Chamber)에서 수행된다. HRTF 알고리즘은 일반적으로 잘 알려진 선형 필터링의 원리를 이용한다. $x_1(t)$ 가 자유 음장 하에서 스피커로 표현되는 전기 신호이고, $y_1(t)$ 가 청취자의 고막에 위치한 탐침 마이크로 수집되는 $x_1(t)$ 의 결과 신호라고 할 때, 이와 유사하게 $x_2(t)$ 는 헤드폰에 의해서 표현되는 신호이고 이에 대응하는 탐침 마이크로 수집되는 신호를 $y_2(t)$ 라 할 수 있다. 주어진 $x_1(t)$ 의 신호에 대하여 HRTF의 목적은 $x_2(t)$ 에 적절한 신호처리기법을 적용하여 $y_2(t)$ 가 $y_1(t)$ 와 같도록 만드는 것이다. $x_2(t)$ 신호를 처리하기 위해 필요한 적절한 필터의 구성은 주파수 영역에서 설명될 수 있다. $x_1(t)$ 신호의 주파수 영역의 표시를 X_1 혹은 $X_1(j\omega)$ 라고 하고 $y_1(t)$ 신호의 주파수 영역의 표시를 Y_1 혹은 $Y_1(j\omega)$ 이라고 할 때 탐침 마이크로에서 응답은 수식 1과 같이 표현될 수 있다.

$$Y_1 = X_1 LFM \quad (1)$$

where L: Loud-speaker Transfer Function
 F: Free-field-to-eardrum Transfer Function
 M:

$x_1(t)$ 와 같은 방식으로 $x_2(t)$ 신호의 탐침 마이크로에서 응답을 표현하면 수식 2와 같다.

$$Y_2 = X_2 HM \quad (2)$$

where H: Headphone-to-eardrum Transfer Function

위에서 언급한 바와 같이 $y_2(t)$ 가 $y_1(t)$ 와 같다고 하면, 즉 $Y_1 = Y_2$ 라고 가정하면 X_2 는 식 3과 같이 표현할 수 있다.

$$X_2 = X_1 LF / H \quad (3)$$

위 식에서 가상입체 음향 신호처리에 필요한 전달함수 T 를 식 4와 같이 정의 할 수 있다.

$$T = LF / H \quad (4)$$

그러므로, 만일 $x_1(t)$ 를 식 4의 전달함수로 통과시킨 결과신호 $x_2(t)$ 를 헤드폰을 재생한다면 탐침 마이크에서 녹음되는 신호 $y_1(t)$ 는 동일한 신호를 자유 음장 하에서 재생한 것과 같은 신호가 된다[3][4].

III. AC-3를 이용한 가상입체음향 신호처리

본 절에서는 AC-3 음향 부호화기를 이용한 다채널 오디오 서비스에서 가상입체음향 신호처리 방법에 대하여 설명하고자 한다. 먼저 AC-3 음향 부호화 방식에 대하여 설명하고, 이어서 MDCT를 이용한 HRTF 가상입체음향 신호처리의 적용에 대하여 설명하도록 한다.

3.1. AC-3(Dolby-Digital) Audio Codec

AC-3(Audio Coding-3)는 돌비 연구소가 영화의 사운드 트랙용으로 개발한 디지털 오디오 코딩을 나타내며 Dolby-Digital이라는 표현으로 사용된다. 돌비 디지털의 특징은 5.1 채널의 다채널 음향을 지원하는 것으로 프론트와 센터, 리어 및 저역의 신호를 담당하는 LFE(Low Frequency Effect) 채널로 구성된다. 돌비 디지털 오디오 코딩은 5.1 채널을 처리할 때 384Kbps에서 448Kbps의 전송량을 가지도록 압축하게 되는데, 그 압축비는 약 12:1 정도 된다. 또한 출력의 샘플링 주파수도 32, 44.1, 48kHz에 모두 대응할 수 있다. 현재 돌비 디지털이 채택되어 사용되고 있는 대표적인 예로는 DVD 플레이어와 HTDV가 있는데 NTSC 방식의 영화 DVD에서는 돌비 디지털이 표준 포맷으로 되어 있다.

돌비 디지털 복호화기는 그림 1과 같이 부호화된 bit stream을 동기화 하는 부분, 에러를 검출하는 부분, 그리고 부호화된 포락선 및 양자화된 가수와 같은 여러 형태의 정보를 재구성 하는 부분으로 구성된다. 이때, 재구성은 bit allocation 부분을 이용하여 압축된 가수를 되돌리고 양자화된 가수를 복원하게 된다. 이렇게 복원한 지수와 가수를 시간 영역으로 전환한 뒤 이를 처리하여 PCM 형식으로 복호화하는 것이다[5].

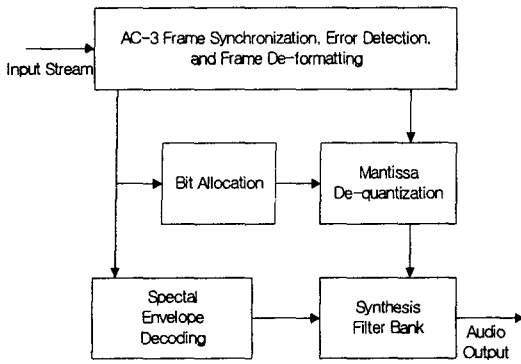


그림 1. AC-3 Decoder

3.2. MDCT 를 이용한 가상입체 음향신호처리

일반적인 HRTF 신호처리는 모노 음원과 청취자 상이의 음원전달경로를 모델링한 HRTF 계수를 이용하여 해당 음원에 대하여 시간영역에서 컨볼루션을 행하거나 주파수 영역으로 변환 후 곱셈기를 이용하는 방식을 사용한다. 그러나 시간영역의 컨볼루션 방법은 음원과 HRTF 계수간의 컨볼루션에 많은 연산량이 필요하며, 주파수 영역에서 처리 또한 모노음원을 주파수 대역으로 변환하여야 하는 전처리 과정을 필요로 함으로써 계산량에 부담을 주게 된다. Dolby-Digital 오디오 부호화기의 경우 복호화된 지수와 가수를 이용하여 시간영역의 계수로 전환하는 과정에서 iMDCT(inverse Modified Discrete Cosine Transform)를 사용하게 된다. MDCT는 선형 직교 중첩 변환으로써 계수의 50%이상이 중첩되는 특징이 있다. MDCT는 식 5와 같이 표현되는데 N개의 오디오 샘플을 가지는 입력 $x(n)$ 으로부터 $N/2$ 개의 MDCT 주파수 계수 X_i 를 얻는데 이용되는 표현식이다.

$$X_i = \sum_{k=0}^{N-1} x_k \cos\left(\frac{\pi}{2N}\left(2k+1+\frac{N}{2}\right)(2i+1)\right) \quad i=0,1,\dots,\frac{N}{2}-1 \quad (5)$$

MDCT는 상대적으로 좋은 스펙트럼 분해능(Spectrum Resolution)을 갖고 있으며, FFT와 같은 기능을 하지만 MDCT를 사용할 경우에는 N개의 오디오 샘플을 $N/2$ 개의 MDCT 주파수 계수로 변환할 수 있는 장점이 있다. 또한 FFT와 같이 시간축의 컨볼루션은 주파수축의 곱셈과 대응되는 성질이 있다.

그러므로 N개의 데이터로 이루어진 HRTF 계수를 MDCT 처리하게 되면 $N/2$ 개의 MDCT 주파수 계수가 생성되며, 해당 HRTF 계수를 데이터베이스로 구축한 뒤

사용자의 요구에 따라 원하는 음원 위치에 대한 MDCT 계수로 변환된 HRTF 계수를 입력신호의 MDCT 계수에 곱함으로써 가상입체음향 신호처리에서 계산량을 감소시킬 수 있다. 이때 필요한 곱셈의 연산은 좌측 HRTF 신호처리에 $N/2$ 개, 우측 HRTF 신호처리에 $N/2$ 개이므로 총 N개가 된다.

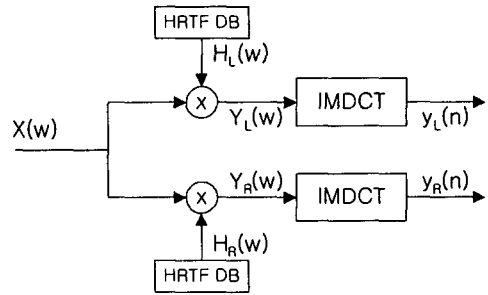


그림 2. iMDCT 방식을 이용한 HRTF 신호처리

IV. AC-3를 이용한 가상입체음향 시스템

4.1. 가상입체음향 시스템의 구성

가상 입체음향 시스템의 입력으로는 5.1 채널 형식의 Dolby Digital 패킷을 사용하게 되며, 이를 이용하여 구성되는 출력의 표본화 주파수는 48kHz이다. 시스템의 구성은 그림 3에서 보는 것과 같이 주변 채널에 가상입체음향 신호처리를 함으로써 가상으로 주변채널의 위치를 변경하는 방식을 사용하였다. 또한 출력신호 값의 포화를 막기 위해 각 채널을 일정비율로 감쇠시켰다.

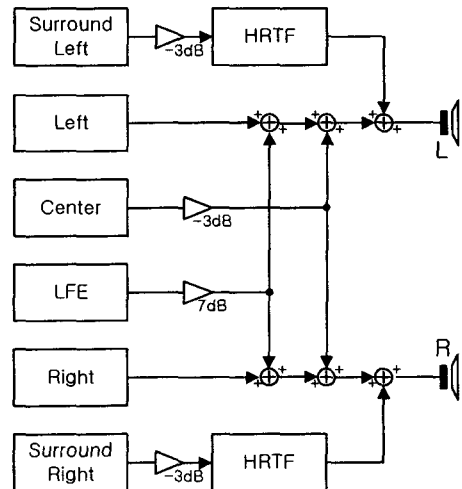


그림 3. 가상 입체음향 시스템의 구성도

4.2. 가상입체음향 시스템의 구현

본 논문에서 설명한 가상입체음향 시스템의 구현에 사용된 AC-3 디지털 오디오 디코더는 일반적으로 사용되는 S/W 디코더의 소스를 이용하였으며, 실시간 구현을 위한 시험 개발용 플랫폼으로는 TI DSP Solution 중 하나인 TMS320C6711 DSK 를 이용하였다. 이 DSK 는 150Mhz 로 동작하는 부동소수점 DSP 와 16M SDRAM, 부트 Flash 롬, JTAG 그리고 기타 주변기기 로 구성된 개발자용 Base board 이다.

동일한 입력에 대하여 PC 상의 AC-3 복호화 결과와 비교 검증을 통해 DSP 구현의 정확성을 확인하였으며, 가상입체음향 신호처리의 경우에도 일반적인 주파수 영역에서 FFT 를 통한 방법과 제안된 MDCT 방법 간의 결과를 비교하여 결과가 같다는 것을 검증하였다.

가상입체음향 신호처리에 사용된 HRTF 는 1994 년 MIT Media Lab 에서 제공한 MIT HRTF 데이터 베이스로 3 차원 사운드 응용과 음상정위 연구에서 가장 널리 사용되는 DB 이며 44.1kHz 표본화주파수에서 16384point 의 길이를 갖는 충격응답에 대한 반응 형태로 구성되어 있다.

최종적인 구현의 결과는 6 채널 AC-3 패킷 데이터를 복호화 하는데 필요한 상대적인 계산량은 평균 0.47CPU Time 이 소요되었으며, FFT 를 이용한 HRTF 신호처리에는 ' Numerical Recipes in C ' 에 제공되는 FFT 함수를 이용한 경우 그림 3 에서 나타난 것과 같이 주변 채널의 신호처리 하는데 약 0.7CPU Time 이 소요됨으로써 실시간 구현에 제약이 있음을 확인할 수 있었다. 하지만 제안한 MDCT 를 이용한 HRTF 신호처리방식의 경우에는 주변채널의 가상입체음향 신호처리에 0.06CPU Time 이 소요됨으로써 실시간 구현을 가능하게 하였고 주변채널뿐만 아니라 다른 채널의 가상입체음향 신호처리의 실시간 구현 또한 가능하게 되었다.

V. 결론

본 논문에서는 Dolby-Digital 오디오 부호화 알고리즘을 이용한 다채널 가상 입체음향 시스템 구현에서 가상입체음향 신호처리의 계산량 감소방법에 대하여 살펴보았다. AC-3 복호화 과정에 한 부분인 iMDCT 변환을 이용하여 가상입체음향 신호처리에 사용되는 HRTF

계수를 MDCT 계수로 변환한 후 사용하는 방식은 일반적인 시간축상의 컨볼루션 방법이나 주파수축의 곱셈을 통한 방법에 비하여 가상입체음향 신호처리에서 현격히 계산량을 감소 시킬 수 있으며 이에 따라 실시간 구현에 대한 부담을 감소시킬 수 있다는 것을 알 수 있었다.

향후 추가적인 입체음향 효과에 대한 연구와 실험이 요구되며 제안된 알고리즘을 이용한 실시간 입출력 제어를 위한 부분을 완전히 구현하여 독립적인 플랫폼 형식으로 구현되어야 할 것이다.

참고문헌

- [1] Begault, D.R., "3-D Sound for Virtual Reality and Multimedia" AP Professional, 1994
- [2] Gardner, W.G., "HRTF Measurement of a KEMAR Dummy-Head Microphone, MIT Media Lab Preceptual Computing Technical Report #280, MIT 1994
- [3] Wightman, F.L., "Headphone simulation of free-field listening. I: Stimulus synthesis" J. Acoust. Soc. Am. 85(2), Feb. 1989
- [4] Gardner, W.G., "Transaural 3-D audio", MIT Media Lab Preceptual Computing Technical Report #342, MIT 1995
- [5] Advanced Television System Committee(ATSC) Standard Doc. A/52, "Digital Audio Compression Standard(AC-3)", Nov. 1994