

청취환경 차이에 따른 3차원 오디오 기술 개발 동향

Status of 3D Audio Technology Development for the difference of Listening Environments

□ 서정일, 이용주, 장인선, 유재현, 강경옥 / 한국전자통신연구원 방송통신융합연구부 방통미디어연구부

요약

3차원 오디오 기술은 3차원 오디오를 녹음하는 기술에서 부호화, 전송, 재현에 이르기까지 모든 오디오 신호처리 과정을 포함한다. 하지만 오디오 신호를 재생하는 단말의 환경이 궁극적으로 3차원 오디오를 청취자에게 제공하는 마지막 단계가 되므로 재생환경에 따라서 적용되는 기술에 차이가 있다. 본 논문에서는 음악 콘텐츠를 소비하는 다양한 재생 단말에서 현실감을 증대시키기 위해 적용되는 3차원 오디오 기술의 연구동향을 살펴본다.

Abstract

3D Audio Technologies include whole signal processing steps from acquisition to reproduction through encoding and transmitting technologies. However, there is a certain difference on adapted technologies according to audio presentation environments, because the presentation environment is the last step to provide 3D audio to

listeners. In this paper, we describe variable 3D audio technologies to adapt variable audio presentation environments for consuming music contents.

1. 서론

20세기초부터 사용되던 스테레오 재생 시스템은 양이효과(binaural effect)를 이용하여 좌우 두 라우드스피커를 이용하여 전방에 한정된 음장감을 제공하는 기술이다. 스테레오 재생 시스템은 CD(Compact Disc)에 적용되어 아직까지도 대부분의 음악 콘텐츠들이 스테레오 사운드로 제공되고 있으며 오디오 신호를 녹음하고 재생하는 가장 기본적인 오디오 시스템이라 할 수 있다. 전방에 한정된 음

* 본 연구는 정보통신부 및 정보통신연구진흥원의 IT신성장동력핵심기술개발사업의 일환으로 수행하였음. [2007-S-004-01, 무안경 개인형 3D 방송기술]

장감을 청취자를 둘러싸는 평면공간으로 확장하기 위하여 5.1채널로 대표되는 멀티채널 시스템이 제안되어 HDTV(High Definition TeleVision), DVD(Digital Versatile Disc), Blu-ray 등에 적용되고 있다.

앞서 설명한 재생 시스템을 이용하여 오디오 장면을 녹음하거나 재생할 때 청취자에게 어떠한 음장감을 제공할 수 있는가에 따라서 2차원 오디오와 3차원 오디오가 결정된다. 멀티채널 오디오가 제공하고 있는 2차원 평면상의 서라운드 음장을 3차원 오디오로 잘못 해석하여 관련 제품 홍보나 논문 등에서 설명하고 있는데, 이는 잘못된 표현이며 서라운드 음장을 제공하는 기존의 멀티채널 오디오 콘텐츠는 2차원 음장을 제공하고 있으므로 엄밀하게 말해서 3차원 오디오라 할 수 없다. 3차원 오디오란 재생 시스템에 관계없이 3차원 공간상에서 형성되는 음장을 녹음하고 재현하는 기술을 뜻하며 관련한 녹음, 신호처리 및 재생 기술을 통칭하여 3차원 오디오 기술이라고 한다.

3차원 오디오 기술은 녹음현장을 그대로 복원하는 음장 복원 기술과 가상 공간상에 음원들을 배치하여 편집자가 원하는 음장을 구성하는 가상음장 재현 기술로 크게 나눌 수 있으며 응용 분야에 따라 달리 적용되고 있다.

3차원 오디오가 청취자에게 제공되는 최종 단계인 오디오 출력 장치(라우드스피커 또는 헤드폰)의 종류나 배치형태에 따라 3차원 오디오를 재현하기 위한 신호처리 과정은 확연한 차이를 가지게 된다. 다시 말하면, 헤드폰이나 스테레오 라우드스피커 환경에서 적용되는 3차원 오디오 기술과 5.1채널과 같은 멀티채널 환경에서 적용되는 3차원 오디오 기술 사이에는 근본적인 차이가 있음을 의미한다. 따라서 3차원 오디오를 효과적으로 재현하기 위해서

는 재생환경에 적절한 알고리즘을 이용하여야 하며, 녹음단계에서도 의도한 3차원 음장감이 손실되지 않도록 재생환경을 고려하여야만 한다.

최근에 급속하게 보급된 MP3 플레이어와 같은 모바일 디지털 오디오 플레이어에서는 머리전달함수(HRTF, Head Related Transfer Function)에 기반한 3차원 오디오 기술들이 적용되고 있으나 헤드폰 재생시 두뇌정의(IHL; In-Head Localization) 문제와 HRTF 필터링에 의한 음질왜곡으로 인해 청취자들이 피로감을 느끼는 문제점이 제기되고 있으며, 차량이나 홈씨어터 시스템과 같은 멀티채널 재생환경에서 3차원 음장을 제공하는데 적용되는 기술들은 제한된 라우드스피커 개수 또는 알고리즘 자체의 불안정성으로 인하여 완벽한 3차원 음장을 제공하는 sweet-spot의 한계를 가지게 되어 다수의 청취자에게 동일한 음장을 제공할 수 없다는 한계점을 지니고 있다. 본 논문에서는 휴대용 오디오 플레이어와 멀티채널 음장 재현시스템에서 양질의 3차원 오디오를 제공하기 위한 기술들의 연구 동향을 소개하고 향후 연구 방향과 응용 방향을 제시하고자 한다.

본 논문은 다음과 같이 구성된다. 2장에서는 휴대 단말을 위한 3차원 오디오 기술들의 연구동향을 소개하고, 3장에서는 다채널 재생환경을 위한 음장재현 기술들에 대해서 소개하며, 4장에서는 향후 연구 방향과 3차원 오디오의 응용분야를 예측하는 것으로 결론을 맺는다.

II. 휴대단말을 위한 3차원 오디오 기술

최근 멀티미디어 기술의 발전으로 인해, mp3 player, PMP, DMB 등과 같은 휴대 단말을 통해 멀

티미디어를 소비하는 경우가 많아지고 있다. 특히, 휴대 전화기에 mp3 player가 내장이 되고, DMB 수신 기능이 추가됨에 따라, 이러한 경향은 더욱 증가하고 있는 추세이다.

한편, 이러한 휴대 단말은 크기가 제한적일 수 밖에 없기 때문에, 소형 스피커를 이용하거나, 헤드폰 또는 이어폰을 이용하여 오디오를 청취하는 것이 일반적이라 할 수 있다. 휴대 단말 환경에서 오디오 신호를 재생하는 경우 단말에 장착된 스피커의 제한적인 요소와 헤드폰 또는 이어폰 청취시의 특성으로 인해 공간감과 현실감이 떨어지는 요소가 존재한다. 가장 대표적인 것이, 헤드폰 또는 이어폰을 이용하여 오디오를 청취하였을 때 음상이 머리 내부에 맺히는 현상과, 소형 스피커로 오디오를 재생하였을 때 저대역 신호가 재생되지 않는다는 문제이다.

헤드폰 또는 이어폰을 이용하여 오디오를 청취하는 경우, 그 음상이 머리 내부에 맺히게 되는 현상이 발생하는데, 이러한 현상은 청취자로 하여금 공간감이나 현실감을 저하시키는 요인이 된다. 이러한 이유로 SRS(Sound Retrieval System), BBE(Base Booster Effect), DNSe(Digital Natural Sound Engine) 등과 같이 3차원 효과를 주는 기능을 휴대 단말에 추가하여 청취자로 하여금 좀 더 현실감 있는 오디오를 제공하기 위한 노력이 이루어지고 있다.

휴대 단말에 포함된 소형 스피커로 오디오 신호를 재생하는 경우, 스피커의 물리적인 특성으로 인해 저대역 신호가 재생이 되지 않는 현상이 발생한다. 이는 스피커를 제작하는데 사용되는 소자 및 스피커의 크기와 관련이 있는 것으로서, 물리적으로 이를 해결하는 것에는 어려움이 따르게 되므로, 오디오 신호처리 기술을 이용하여 심리음향학적으로

저대역 신호가 있는 것처럼 처리를 해주는 기술을 통해 이러한 문제점을 해결하고 있다.

이와 같이, 휴대 단말에서 오디오를 재생하였을 때의 제한적인 특성을 해결하기 위한 연구가 지속적으로 이루어지고 있는데, 본 절에서는 앞서 언급하였던 가장 대표적인 두 가지 특성에 대해 좀더 자세하게 기술한다.

1. 오디오 신호 외재화 기술

현재 휴대용 오디오 기기나 개인용 오디오 기기에 있어서 스피커보다는 헤드폰에 의한 오디오 재생 방법을 주로 사용하고 있는데, 이와 같이 귀에 근접하거나 삽입된 음향발생 장치에 의해 오디오 신호를 청취하는 경우, 양 귀 사이의 1차원 구간에만 음이 생성되는 특징을 가지게 된다. 이러한 현상은 음상의 내재화 현상(IHL) 이라고 하는데, 내재화 현상은 공간감이나 입체감을 떨어지게 하여 음향의 현실감을 저하시키는 요인이 되며, 또한 청취자로 하여금 피로를 쉽게 느끼게 하는 요인이 된다[1].

음상 내재화 현상의 원인은 여러 가지가 있으나, 주요한 원인으로는 아래와 같은 것들이 있다.



- 개인의 머리 및 컷바퀴 등 인체에 의한 음향 특성의 변화 부재
- 공간의 반사 및 잔향에 의한 공간 음향 특성 미반영
- 머리의 움직임에 의한 음향 전달특성의 일관적인 변화 미반영
- 시각 및 기억에 의한 음원의 위치 정보 부재
- 친밀한 음원 및 공간의 전달특성 예측 능력 미적용

이러한 주요한 요인 외에도 음상 내재화 현상을 발생할 수 있는 많은 요인들이 있는 것으로 알려져

있는데, 아래와 같은 요인들이 그 예이다.



- 이어폰 착용에 의해 귀에 주어지는 음압이 지나치게 안정됨.
- 두 귀에 전달되는 경로가 독립적이며, 미세하게 다름.
- 전달 시스템 및 전기음향 변환기의 비선형 왜곡
- 중이(middle ear)에 비정상적인 음향 공급·골도(뼈의 진동에 의한 음향 전달)

음상 내재화 문제를 해결하여 헤드폰/이어폰을 통한 청취 시 음상이 머리의 외부에 맺히도록(OHL; Out of the Head Localization)하는 기술을 음상 외재화(Sound Externalization) 기술이라 하는데, 주로 음상 내재화의 원인이 되는 요소들을 보상해주는 방법을 사용하고 있다. 음상을 완벽하게 외재화 시키기 위해서는 이러한 모든 음상 내재화 요인을 반영하여야 하지만, 여러 가지 현실적인 이유로 실현이 불가능하다고 할 수 있다. 최근에는 공간의 반사 및 잔향에 의한 공간 음향 특성, 개인의 머리 및 귓바퀴 등 인체에 의한 음향 전달 특성, 머리 움직임에 의한 음향 전달 특성 변화 등과 같이 음상 내재화의 주요한 요인에 대한 보상을 통해 외재화를 실현하려는 접근들이 주로 연구되어 오고 있다.

본 절의 다음에서는 최근 이루어지고 있는 주요한 외재화 기술에 대해 구체적으로 살펴보고 기술한다.

1) 개인의 머리 및 귓바퀴 등 인체에 의한 음향 특성의 변화 보상

일상생활에서 우리가 듣는 소리는 음원으로부터 나오는 소리가 아니라, 머리, 어깨, 귓바퀴 등에 의해 회절, 반사를 거친 소리이다. HRTF는 일정한 거

리로부터 청취자의 귀로 소리가 입력되는 전달함수로서 청취자의 몸통과 머리, 귀의 영향을 반영한 일종의 충격응답이다. 이어폰으로 사운드를 재생할 때, 외부에서 소리를 듣는 것과 같은 효과를 내기 위해서는 HRTF를 적용할 필요가 있다.

대부분의 HRTF 데이터는 더미헤드를 이용하여 측정이 이루어지고 있으며, 음상정위에 많이 사용되고 있다. 그러나 각 개인마다 머리 크기, 귀의 모양 등이 다르기 때문에 더미헤드를 이용하여 측정된 HRTF의 경우 다수의 청취자에게 동일한 효과를 나타내지는 않는다. 즉, 개인화되지 않은 일반적인 HRTF는 정확한 각도 인식뿐 아니라, 높이 인식, 정/후면 인식에도 혼란을 가져올 수 있다는 단점을 가지게 된다[2]. 이러한 이유로 최근에는 개인화된 HRTF를 측정하는 방법에 대한 연구가 이루어지고 있다.

개인화된 HRTF를 획득하는 방법의 대표적인 것으로는 소형 마이크로폰을 귀안에 장치하여 직접 개인의 HRTF를 측정하는 방법, 머리의 크기나 귓바퀴의 형상에 의한 개인 HRTF 추정하는 방법, 체형에 따른 HRTF 데이터 DB를 구축하고 개인에 맞는 HRTF를 선택하도록 하는 방법, 소형 마이크로폰이 장착된 이어폰과 외부 스피커를 이용하여, 실시간으로 스피커와 귀 사이의 오디오신호 전달특성을 획득하는 방법 등이 있다. 그러나, 개인의 HRTF를 직접 측정하는 방법과 소형 마이크로폰이 장착된 이어폰을 이용하는 방법의 경우는 복잡한 과정과 장치로 인해 상업적으로 적용하는데에 어려움이 있다는 문제점이 있으며, 머리의 크기나 귓바퀴의 형상에 의한 개인 HRTF 추정하는 방법은 머리의 크기나 귓바퀴의 정확한 데이터를 얻는 방법이 아직은 기술적으로 성숙하지 않아 정확하지 않다는 문제점이 있다. HRTF 데이터 DB로부터 개인에 맞

는 HRTF를 선택하도록 하는 방법의 경우는 정확성이 떨어져서 효과의 유효성이 얼마나 있는지가 명확하지 않다는 점 등의 문제점이 있다.

이와 같이 개인화된 HRTF를 획득하는 것은 아직 여러 가지 문제점을 가지고 있어서 상업적으로 이용되고 있지는 못한 상태이나, 멀지 않은 미래에는 이러한 문제점을 해결하여, 쉽게 개인화된 HRTF를 얻을 수 있을 것으로 생각되며, 외재화 성능도 높일 수 있을 것으로 생각된다.

2) 공간의 반사 및 잔향에 의한 공간 음향 특성 보상

소리가 발생하는 음원과 청취자 사이에는 벽, 책상 등과 같이 많은 물체들이 존재한다. 이러한 물체들로 인해 음원에서 발생하는 소리는 반사와 회절을 거치게 되며, 최종적으로 청취자가 듣는 소리에 도 영향을 미치게 된다. 즉, 청취자는 음원에서부터 직접 청취자에게 전달되는 소리 뿐 아니라, 벽, 책상 등에 의해 반사된 소리를 함께 듣게 되는 것이다. 따라서, 이어폰으로 사운드를 재생할 때, 외부에서 소리를 듣는 것과 같은 효과를 내기 위해서는 이와 같은 반사음들을 고려할 필요가 있다.

일반적으로 스튜디오 등에서 녹음된 음성이나 음악 등은 반사음이 충분히 반영되어 있지 않으므로, 외재화 효과를 주기위해 인공적으로 반사음 및 잔향을 추가하는 방법을 사용한다. 이러한 반사음 및 잔향을 획득하는 방법은 여러 가지가 있을 수 있는데, 직접 룸임펄스 응답을 획득하여 이용하는 방법, 인공적으로 룸 환경을 구축하여, 그 룸에서의 반사음 및 잔향을 시뮬레이션(simulation)하여 획득하는 방법 등이 대표적이다.

일반적으로 사용하는 인공잔향 방법의 경우 고정된 반사음 및 잔향만을 사용하고 있어, 실제 공간과

의 특성이 일치하지 않게 되는데, 이로 인해 사람에 따라 느끼는 공간감의 편차가 심하며, 외재화 효과도 제한된다는 문제점이 있다. 이러한 문제점에 대한 연구는 지속적으로 이루어지고 있는데, 휴대용 장비를 사용하는 환경에서 머리의 움직임에 따라 상대적으로 반사음 및 잔향을 변화시켜주는 방법이 이러한 연구 중 하나이다.

3) 머리의 움직임에 의한 음향 전달특성의 일관적인 변화 보상

일상생활에서 우리는 항상 움직이고 있는데, 이러한 움직임에 따라 동일한 음원에서 나오는 소리도 변화하게 된다. 이는 사람의 움직임에 따라 앞서 언급하였던 음원과 귀간의 전달함수가 변화하게 되고, 반사음 및 잔향도 변화하기 때문이다. 이어폰으로 소리를 청취하는 환경에서는 사람이 움직이더라도, 소리의 전달 경로가 바뀌지 않기 때문에, 소리가 변화하지 않는다. 이는 이어폰을 사용하지 않는 일상적인 청취환경과는 다른 현상으로서, 음상이 머리에 맺히게 되는 내재화 현상의 원인으로 작용한다.

최근에는 이러한 머리 움직임의 변화를 보상해주는 방법에 대한 연구가 이루어지고 있는데, 머리의 움직임에 의한 음향 전달특성의 일관적인 변화를 재현하기 위해, 머리의 움직임을 감지하고 머리의 움직임 양에 따른 보상처리를 해 주는 방법이다. 이러한 처리가 많은 지연을 동반하게 되면 적절한 효과를 얻을 수 없으므로, 모든 처리가 실시간으로 이루어 질 수 있도록 하여야 한다. 이러한 이유로 수평 방향의 움직임만을 보상하는 방법도 있고, 상하 움직임이나 각도에 대한 움직임까지도 보상하는 방법이 있지만, 자유도가 높을수록 그 처리가 매우 복잡해지고 필요한 하드웨어적인 수단도 복잡해지게 된

다. 수평 방향의 움직임만을 보상하는 헤드폰이 상용화되어 출시되고 있는데, 머리 움직임에 대해 안정된 음상을 제공하여 음상정위 및 외재화에도 효과가 있는 것으로 알려져 있다.

이와 같이 머리 움직임을 보상해 주는 방법은 머리의 움직임을 감지하는 장치를 추가해야 한다는 점과 움직임에 대응하는 음상의 제어 처리에 상당한 계산량이 필요하다는 것이다. 상용화된 제품의 경우 헤드폰에 센서를 장착하여 비교적 장착이 용이하였다고 볼 수 있으나, 일반적으로 사용되는 이어폰에 이러한 센서를 장착하는 것에는 가격적인 면에서나, 크기 면에서 어려움이 있을 수 있다. 또한, 상용 제품의 경우 수평 방향의 머리움직임에 의한 소리의 보상을 간단한 형태로 구현하였으나, 정확한 보상을 위해서는 더욱 복잡한 연산이 필요하며, 이를 상용화하는 데에는 아직도 어려움이 있는 것으로 생각된다. 또한 휴대용 장비에 있어서는 손에 들고 사용할 때 그 위치가 수시로 변할 수 있기 때문에, 이에 대한 고려가 없다면 머리 움직임 보상 효과가 감소되는 문제점도 있다. 머리움직임을 보상하는 기술은 이와 같은 문제점을 아직까지는 가지고 있으나, 센서 기술의 발달과 신호처리 기술의 발전으로 인해 멀지 않은 미래에는 이러한 문제점이 해결될 것으로 예상된다.

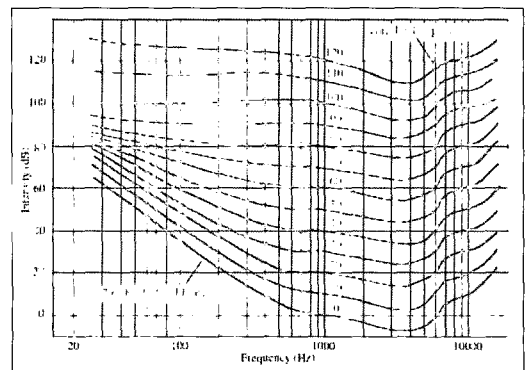
본 절에서는 음상 내재화 현상 및 이를 극복하기 위한 기술에 대해 간략하게 살펴보았다. 음상의 완벽한 외재화를 위해서는 음상 내재화를 일으키는 모든 요인들을 고려하여 보상해 주는 것이 좋겠지만, 현실적으로는 많은 어려움이 있다. 최근에 연구되고 있는 개인화된 HRTF, 반사음 및 잔향음, 머리 움직임의 보상 등과 같은 연구는 가장 외재화 효과가 크며, 다른 요인들의 효과를 대체할 수 있는 요인이라고 생각되는데, 아직 상업적으로 이용될 만큼 기술

이 성숙되지는 않으나, 그 발전 속도를 고려하면 멀지 않은 미래에는 헤드폰/이어폰 환경에서도 스피커로 소리를 듣는 듯한 소리를 들을 수 있을 것으로 기대된다.

2. 오디오 신호 재생대역 확장 기술

도입부에서 설명한대로, 모바일 기기에서 사용되고 있는 소형 라우드스피커로 3차원 오디오를 재생할 경우, 약 400Hz 이하의 저주파수 대역을 원음 그대로 재생시킬 수 없다. 이 저주파 대역을 재생시키기 위한 기존의 대역 확장 기술들은 저주파 대역 성분의 신호를 임의적으로 강조하거나 증폭하여 보강하는 방식과 조화음(harmonics) 성분을 이용하여 심리 음향학적으로 저주파 대역을 생성하는 방식으로 나뉜다.

그림 2.2.1의 등라우드니스 곡선에서 보다시피 효과적인 저대역 재생을 위해서는 큰 에너지 소모를 필요로 한다. 따라서 저주파 대역 성분의 신호를 임의로 강조하거나 증폭하는 기술은 라우드스피커에 엄청난 양의 에너지를 전달해야 하기 때문에 더 큰 힘으로 라우드 스피커들을 드라이브하게 된다. 이는



〈그림 2.2.1〉 Equal loudness-level curves derived from Fletcher-Munson data

라우드스피커들을 오버드라이빙하는 결과를 초래하기 때문에 백그라운드 잡음의 증가, 왜곡 발생 및 라우드스피커 손상 등의 문제점이 발생하게 된다.

이와 같이, 물리적인 저주파 대역 확장 기술은 라우드스피커 출력에 무리를 주고 출력 신호의 음질 열화 등을 일으킬 뿐 아니라 에너지 소모 또한 크기 때문에 소형화/경량화를 요하는 휴대단말에서 사용하기에는 적절치 못하다.

이에 대한 대안으로 고려되는 기술이 바로 심리음향기반 저주파 재생대역 확장기술이다. 이 기술은 누락 기본음 이론(missing fundamental theory)에 기반을 두고 있으며, 이는 기본 주파수(fundamental frequency)가 없는 소리가 그의 조화음들에 의해 마치 존재하는 것으로 인식되게 하는 기술이다. 이는 사람의 뇌가 기본 주파수 만의 해석이 아니라 조화음들의 비율에 의해서도 음(tone)의 피치(pitch)를 인식하기 때문에 발생하는 현상이며, 이 결과로써 임의의 음에서 기본 주파수가 누락되었다 해도 우리는 같은 피치를 인지하게 된다. 누락 기본음 효과에 기반한 상용 저역대역 확장 기술은 미국 Waves사의 BaxxBass[3,4], SRS의 TruBass[5] 등이 있으며 이외에도 Phillips[6] 등 여러 기관에서 관련 솔루션을 보유하고 있다.

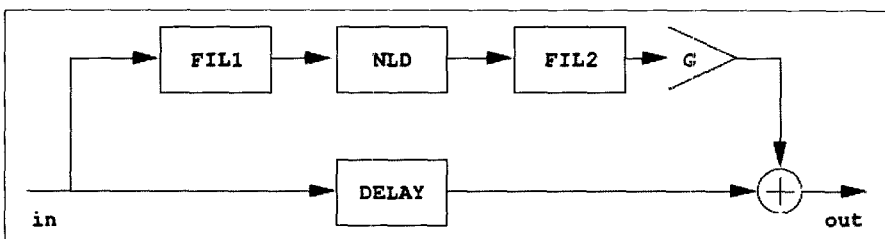
누락 기본음 이론에 기반한 저주파 재생대역 확

장 기술의 기본 구조 및 각 모듈 별 기능은 그림 2.2.2와 같다[6].

입력된 오디오 신호에 대해 FIL1에서 스피커의 저역 컷오프(cutoff) 주파수 이하의 대역에 존재하는 주파수 요소를 선택한다. FIL1을 통과한 주파수 요소의 기본 주파수를 추출하고 이에 대한 조화음(harmonics)을 생성(NLD) 한 후, FIL2에서는 NLD에 의해 생성된 조화음의 주파수 스펙트럼 모양을 적합하게 한다. 이후 수정된 조화음을 입력된 신호에 추가하여 재생함으로써 심리음향적으로 저주파 재생 대역을 재생할 수 있다.

III. 다채널 재생 환경을 위한 음장 재현 기술

대화면에서 3차원 비디오나 초고화질 비디오를 제공하는 3DTV나 UDTV 환경에서 현실감을 제공하기 위해서는 현장과 동일한 음장을 제공하는 것도 중요한 문제이다. 헤드폰이나 5.1채널 이하의 개수를 가지는 라우드스피커 재생환경에서 원음장과 동일한 음장을 다수의 청취자에게 제공하기는 매우 어려운 문제이며, 수십개 또는 수백개의 라우드스피커를 이용하여 원음장과 동일한 음장을 재현하는

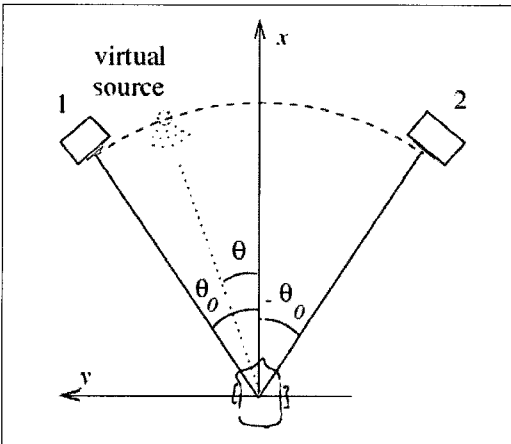


(그림 2.2.2) General BWE framework

여러 가지 기술들이 제안되어 있다. 본 절에서는 다 채널 라우드스피커를 이용하여 음장을 재현하는 기술과 연구현황에 대해서 소개한다.

1) 패닝기반 멀티채널 오디오 재생기술

멀티채널 라우드스피커 배치를 이용하여 음장을 재현하는 기술중에 가장 고전적이며 널리 이용되는 패닝(panning) 기술은 그림 3.1.1과 같은 표준 스테레오 오디오 재생환경에서 두 라우드스피커 사이의 원하는 위치에 가상음원을 위치시키기 위해 각 라우드스피커에 인가하는 신호크기를 제어하는 기술이다[7].



〈그림 3.1.1〉 표준 스테레오 청취환경($\theta_0 = 30^\circ$)

대표적인 패닝기법인 크기패닝 기법(사인 법칙이라고 부르기도 함)에서는 가상음원을 원하는 각도(θ_s)에 위치시키기 위하여 아래 식을 만족하도록 라우드스피커 신호의 출력크기(g_1 과 g_2)를 조절한다. 단, 식 (3.1-1)은 $\omega\mu \approx 1$ (ω 는 음의 각속도, μ 는 발생한 시간지연) 일 경우로 근사화한 수식이므로 패닝 기법은 600Hz 이하의 저주파수 대역에서만 유효하다.

$$\frac{\sin\theta_s}{\sin\theta_0} = \frac{g_1 - g_2}{g_1 + g_2} \quad (3.1-1)$$

5.1채널 셋업과 같이 청취자와 같은 평면상에 라우드스피커를 배열하여 음장을 조절하는 2D 라우드스피커 셋업에서는 가상음원을 위치시키기 위한 방법으로 인접하는 두 라우드스피커 사이에 위상을 정위시키는 pair-wise amplitude panning 기법을 일반적으로 적용한다. 이 방법은 가상음원의 위치가 라우드스피커의 위치와 동일하면 최적의 효과를 가져온다는 장점은 있지만, 라우드스피커들 사이에 가상음원을 정위시키는 경우 음상이 뚜렷하지 않고 모호하게 된다. 즉, 정위하고자 하는 위치에 따라 음상의 resolution이 변화한다는 단점(different directional spreading)을 가지고 있다. 이러한 현상은 가상음원을 정위시키기 위해 인가하는 음향신호가 두 개의 라우드스피커에 제한되기 때문에 일어난다.

모든 라우드스피커들이 청취자와 같은 평면상에 위치하지 않는 셋업을 3D 라우드스피커 셋업이라고 지칭하며, 일반적으로 2D 라우드스피커 셋업에 고저감을 주기위한 라우드스피커들을 추가로 배치한다. 3D 라우드스피커 셋업에서 음상을 정위시키는 패닝기법을 주로 Triplet-wise 기법이 사용된다. Triplet-wise 기법에서는 음상이 정위되는 시점에서 최대 3개의 라우드스피커가 이용된다. 따라서, 3D 라우드스피커 셋업에서의 라우드스피커들 배치 구조는 삼각형으로 이루어진 격자들로 구분할 수 있으며 가상음원을 위치시키기 위하여 격자들 중 하나가 이용된다. Ville Pulkki는 임의의 3D 라우드스피커 셋업에 적용할 수 있도록 3차원 벡터의 투영을 통한 패닝 기법을 Vector Based Amplitude Panning (VBAP)이란 이름으로 제안하였다[8]. 그러나, 3D 라우드스피커 셋업은 2D 라우드스피커

셋업과 동일한 이론에 근거한 패닝 기법을 이용하므로 이로 인해서 발생하는 장단점은 동일하게 나타나게 된다.

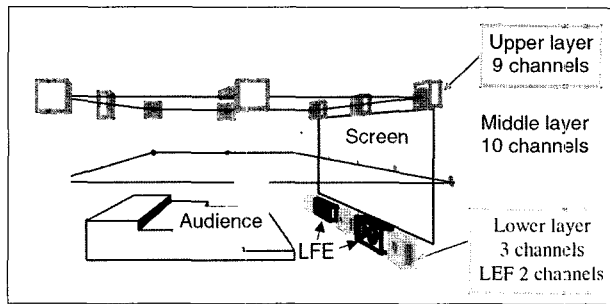
패닝기법을 이용하는 음장재현 기술들 가운데 최근들어 각광을 받고 있는 엠비소닉스는 기본적으로 마이크로폰 제작 기술이지만 인접한 라우드스피커 사이의 거리가 동일하도록 라우드스피커들을 배치하고 진폭 패닝(amplitude panning) 기법을 모든 라우드스피커를 이용하여 적용함으로써 음상을 정위시키는 방법이다[9]. 단, 모든 라우드스피커에 신호를 인가함으로써 최적 청취영역을 벗어난 영역에서 음상이 깨지는 단점은 여전히 존재한다. 이를 극복하기 위하여 제안된 것이 2차 엠비소닉스(Second-order Ambisonics) 시스템이다. 2차 엠비소닉스 시스템에서는 이득값이 식 (3.1-2)와 같이 결정되므로, 패닝방향의 반대방향의 이득값이 두드

러지게 작아진다. 식 (3.1-2)에서 N은 라우드스피커의 개수이며 α_i 는 라우드스피커의 위치를 각도로 표현한 것이다. 따라서 모든 라우드스피커에 신호를 인가함으로써 발생하는 역효과는 줄일 수 있다. 엠비소닉스 시스템은 라우드스피커의 셋업이 대칭인 구조를 가져야 하며, 라우드스피커의 개수를 어느 정도까지 증가시킨 이후에는 더 이상의 성능 향상을 기대할 수 없다는 한계를 가지고 있다.

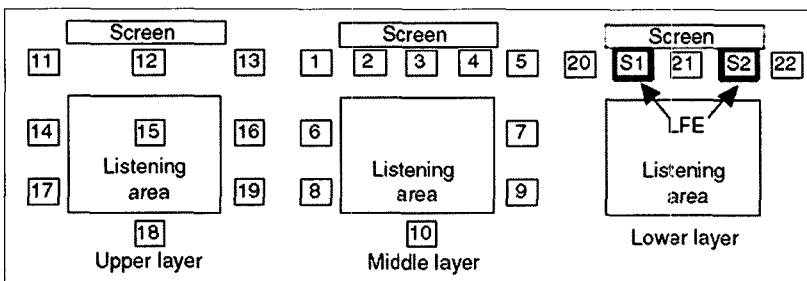
$$g_i = \frac{1}{N}(1 + 2\cos \alpha_i + 2\cos 2\alpha_i) \quad (3.1-2)$$

2. NHK 22.2채널 오디오 재생기술

NHK의 22.2채널 시스템은, 그림 3.2.1과 3.2.2에서 보는 바와 같이, 세 개의 수평 층(layer)와 두



(그림 3.2.1) 22.2 멀티채널 오디오 시스템의 라우드스피커 배치 (side view)



(그림 3.2.2) 22.2 멀티채널 오디오 시스템의 라우드스피커 배치 (upper view)

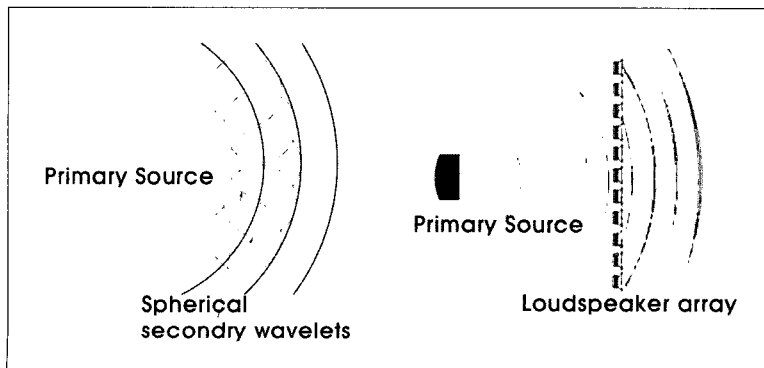
개의 LFE(Low Frequency Effect)로 이루어져 있다. 아래 층은 3개의 라우드스피커로, 중간 층은 10개의 라우드스피커로, 그리고 위 층은 9개의 라우드스피커로 구성된다[10,11].

5.1채널 오디오 시스템은 영화관에서 뿐만 아니라, DVD-Video, DVD-Audio, SACD(Super Audio CD) 그리고 디지털 방송에 이르기까지 다양한 분야에서 널리 사용되고 있다. 그리고 비록 2채널 스테레오 시스템에 비하여 향상된 공간감을 제공하기는 하지만, 보다 넓은 청취공간을 형성하는데 있어서 5.1채널 오디오 시스템에는 여러 제약 사항이 뒤따르게 된다. 특히, 최적 청취 영역(sweet spot)을 좁게 형성하기 때문에 극장과 같이 넓은 청취공간에서는 이상적이지 않다. 차세대 방송 시스템을 개발하는 데에 있어서 주안점은 실재감과 현장감을 어떻게 향상시키는가이다. 실재감과 현실감은 전면 2차원 상에서 안정적인 음상 정위, 청취자 주위의 모든 방향에 걸쳐 음상의 재생 그리고 청취자의 적절한 포위감과 더불어 3차원적 공간감 등과 같은 요소들을 필요로 한다. 22.2 멀티채널 오디오 시스템은 보다 넓은 청취공간에서 향상된 실재감과

현장감을 제공한다[12].

3. 파면 음장 합성기술

3차원 오디오 기술은 영화나 레코딩 산업에 필수적이었는데 50년대 초반에는 3차원 오디오 시스템이 2개 혹은 그 이상의 채널을 이용했다. 하지만 2개 이상의 채널을 추가한 데에 따른 효과는, 추가적인 기술이나 경제적 효과를 정의하기에 충분한 결과를 나타내지 못했다. 당시 3차원 오디오 재생에는 두 가지 중요한 문제점이 있었는데 가상 음원이 두 개의 라우드스피커 사이에만 놓일 수 있다는 점과 청취영역의 제한이 발생한다는 점이다. 70년대를 지나 4개의 채널을 이용해 가상 음원의 위치를 확장시킨 quadraphony 연구가 이루어졌으며, 최근에는 5.1채널 시스템을 비롯한 다수의 새로운 서라운드 표준이 극장이나 가정용 극장 시스템에 적용되었다. 그러나 이러한 새로운 3차원 오디오 기술들은 가상음원이 제한된 영역에만 놓일 수 있다는 단점은 해결해주었지만, 청취영역을 확장시키지는 못했다.



〈그림 3.3.1〉 호이겐스 원리(좌)와 라우드스피커 어레이를 통한 음장의 합성으로 원 음원의 파형 재생(우)

80년대 후반 Berkhout에 의해 그 개념이 소개된 WFS(Wave Field Synthesis)는, 그림 3.3.1에 나타난 바와 같이, 파면(wave front)상의 모든 점은 구형 2차 파면의 점 음원(point source)으로서 작용한다는, 호이겐스 이론(Huygens's principle)에 기반을 두고 있으며 키르히호프-헬름홀츠 적분(Kirchhoff-Helmholtz integral)과 레일레이 정리(The Rayleigh's representation theorem)에 기반한 알고리즘에 의해, 작고 가깝게 배치된 다수의 라우드스피커로 구현한 어레이를 통해 음장의 중첩으로 음장에 해당하는 파장(wave field)을 합성해내는 기술이다[13,14].

이 기술은 재생하고자 하는 음원의 원래의 파형을 정확하게 재생할 수 있으며, 가상 음원을 라우드스피커 어레이 뒤나 청취 공간 내부 등 공간상의 어느 곳이나 정확하게 정위시킬 수가 있고, 진행 방향에 따라서 음압 레벨이 거의 감소하지 않는 특성을 가지는 평면파를 합성해 낼 수도 있다. 이러한 특성들로 인해 청취 위치에 무관하게 모든 청취자에게 동일한 음상을 제공할 수가 있어 기존의 3차원 오디오

기술이 가지는 청취 영역이 제한되는 문제점을 해결할 수가 있다. WFS는 이러한 재생 특성에도 불구하고, 높은 연산량으로 인해 현재까지도 넓은 애플리케이션을 확보하지 못하고 있는데 2003년 독일 프라운호퍼 IDMT에서는 IOSONO 라는 WFS 편집 및 재생 애플리케이션을 이용하여, 100개의 좌석이 설치된 일메나우 극장에서 그림 3.3.2 와 같이 198개의 스피커를 사용한 WFS 어레이를 설치하기도 하였다[15].

물론 WFS도 단점을 가지고 있는데, 시스템을 구현하는데 있어서, 무한한 연속적인(continuous) 파면을 유한한 불연속적인(discrete) 라우드스피커 배열을 통하여 재생을 하기 때문에 발생하는 spatial alias 및 truncation effect 등과 같은 문제점이 그것이다. 이는 이론적인 WFS를 실제 구현할 경우 필연적으로 발생할 수 밖에 없는 문제점으로, spatial alias는 라우드스피커가 일정한 간격을 가지고 배치되어 있기 때문에 이를 통해 재생되는 주파수 재생 범위에 제한을 받는 것이며, truncation effect는 유한한 개수를 가지는 라우드스피커 배열의 양 끝



〈그림 3.3.2〉 일메나우 극장에 설치된 WFS 애플리케이션

단에서 발생하는 파장이 전체 배열에서 발생하는 파장의 왜곡을 일으키게 된다는 것으로 이들을 보완하는 방안에 대해서도 많은 연구가 이루어지고 있다. 또한 시스템 구현에 많은 수의 라우드스피커가 필요함으로 인해 발생하는 공간 상의 제약이 또 다른 문제점으로 남아 있어 DML(Distributed Mode Loudspeaker) 등과 라우드스피커를 대체할 수 있는 방안에 대해서도 연구가 이루어지고 있다.

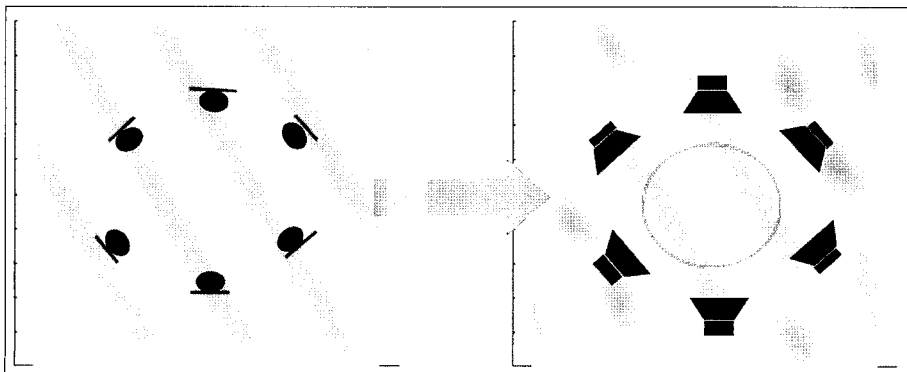
4. 공간 음장 복원기술

멀티채널 음장재현 시스템의 목적은 실제 음장신호를 재현(recreate a real sound field)하는 것과 가상의 음장신호를 합성(synthesize a virtual sound field)하는 데 있다. 가상음장신호를 합성하는 과정은 비디오 게임이나 노래 및 영화 제작과정에 해당한다. 음원 신호들은 합성하거나 제어된 공간(예를 들어 무향실이나 스튜디오)에서 근접 마이크로폰(close microphone) 기술을 이용하여 녹음하며, 사운드 엔지니어에 의해서 가상공간에 임의로

배치하고 잔향기 같은 효과를 걸어줌으로서 원하는 음장신호를 합성한다. 이때는 가상 음원들에 의해 합성되는 원하는 음장에 대한 정보를 미리 정의할 수 있다는 장점이 있다. 멀티채널 오디오 재생 시스템의 목적이 실제 음장신호를 재현하는데 있다면, 실제 음장 신호를 녹음할 수 있어야 하며, 음장을 표현하는 모든 정보들은 규정된 포맷으로 인코딩 되어야 한다.

그림 3.4.1과 같이 공간음장 복원기술은 마이크로폰 어레이를 이용하여 실제 음장신호를 녹음한 후 인코딩하여 저장하고, 저장된 음장정보를 디코딩한 후 라우드스피커 어레이를 이용하여 공간음향을 재생하는데 관련된 일련의 과정을 지칭한다.

대표적인 공간음장 복원기술로써 HOA, LSM이 있다. HOA(High Order Ambisonics)는 first-order Ambisonics의 확장된 형태이며, 음장신호를 공간 조화음(spherical harmonics) 성분으로 분리하는 이론에 기반을 두고 있다[16]. 즉, 음장은 homogeneous Helmholtz equation에 기반하여 Bessel 함수와 spherical harmonics의 Laplace series로 식 (3.4-1)과 같이 표현할 수 있다.



〈그림 3.4.1〉 공간음장 복원과정 개념도

$$p(r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} (A_{m,n}(k)j_n(kr) + B_{m,n}(k)y_n(kr))Y_{m,n}(\theta, \phi) \quad (3.4-1)$$

$$J = \sum_{k=1}^K (p_k - \hat{p}_k)^2 \quad (3.4-2)$$

위 식에서 k 는 wave number를, $Y_{m,n}$ 은 spherical harmonics를, j_n 과 y_n 은 1차와 2차 Bessel function 을, $A_{m,n}$ 과 $B_{m,n}$ 은 series coefficients 또는 spherical harmonic components를 나타낸다.

ISVR에서 제안된 LSM(Least Squares Method)은 원하는 음장(target sound field)과 복원되는 음장(reconstructed sound field)사이의 오차를 줄이는 방향으로 음장재현 시스템을 적용시키는 기술이다[17]. 재현하고자 하는 음장 신호가 마이크로폰 어레이를 이용하여 녹음되고, 마이크로폰 어레이가 라우드스피커 어레이 안쪽으로 배치되며, 마이크로폰 어레이가 배치되는 지역이 원하는 음장을 복원하는 영역(sweet spot)이며, 라우드스피커 어레이와 마이크로폰 어레이가 동일한 좌표계를 사용한다고 가정하면, 최적의 음장복원 조건은 원음장을 녹음한 마이크로폰 신호(p)가 라우드스피커에서 복원된 음장을 녹음한 마이크로폰 신호(\hat{p})와 동일할 경우이다. 이러한 조건에서 두 마이크로폰 신호사이의 전체 자승오차는 식(3.4-2)와 같이 정의할 수 있으며 이를 최소화 하도록 라우드스피커 입력신호를 찾아가는 과정이 LSM이라 할 수 있다.

IV. 결론

본 논문에서는 3차원 오디오 기술의 연구동향을 오디오 단말의 재생환경에 따라 구분하여 살펴보았다. 5.1채널 음장 재생 시스템의 급속한 보급으로 인하여 소비자들의 고품질 음향에 대한 욕구가 증대하고 있으며, 차세대 디지털 방송 형태로 제시되고 있는 3DTV나 UDTV(Ultra Definition TV)에서도 현장감 있는 음향을 제공하는 것이 기본적인 요구 사항이므로 고품질 3차원 오디오 기술을 개발하기 위한 연구는 국내외에서 지속적으로 진행 중이다. 특히 휴대단말에서는 3차원 비디오와 마찬가지로 3차원 오디오를 장시간 청취하였을 때 발생하는 피로도를 줄이기 위한 노력도 3차원 오디오 기술 개발과 병행하여 이루어지고 있다.

3차원 오디오 기술의 상용화는 스테레오 재생 시스템을 이용하는 휴대단말을 중심으로 적용되고 있으므로 3차원 오디오 기술의 성능 뿐만 아니라 휴대단말에 탑재하기 위하여 알고리즘의 복잡도를 낮추는 기술도 중요한 연구 분야의 하나라고 할 수 있다.

참고 문헌

- [1] N. Sakamoto, T. Gotoh and Y. Kimura, "On 'Out-of-Head Localization' in Headphone Listening," JAES, vol. 24, pp. 710-716, Nov. 1976.
- [2] Chong-jin Tan and Woon-seng Gan, "Direct Concha Excitation for the Introduction of Individualized Hearing Cues," J. Audio Eng. Soc., Vol.48, No 7/8. 2000 July/August
- [3] D. Ben-Tzur, M. Colloms, "The effect of the MaxxBass Psychoacoustic Bass Enhancement system on loudspeaker design," 106th AES Convention, 1999.

- [4] D. S. Waves Ltd, "Method and system for enhancing quality of sound signal," US patent 5,930,373, 1999.
- [5] SRS Labs, INC., "Low-frequency audio enhancement system," US patent 6,285,767 B1, 2001.
- [6] R. M. Arts, E. Larsen and O. Ouweltjex, "A unified approach to low- and high- frequency bandwidth extension," 115th AES Convention, 2003.
- [7] V. Pulkki, and M. Karjalainen, "Localization of amplitude-panned virtual sources I: stereophonic panning," J. Audio Eng. Soc., vol. 49, no.9, pp. 739-752, Sep. 2001.
- [8] Ville Pulkki, "Spatial Sound Generation and Perception by Amplitude Panning Techniques," Ph.D Thesis, Helsinki University, 2001.
- [9] M.A.Gerzon, "Panpot laws for multispeaker stereo," in The 92nd Convention 1992 March 24-27 Vienna, Audio Engineering Society, Preprint No. 3309, 1992.
- [10] Hamasaki kimio, et al., "Wide Listening Area with Exceptional Spatial Sound Quality of a 22.2 Multichannel Sound System," 122nd AES Convention, 2007, Preprint No. 7037.
- [11] Hamasaki kimio, et al., "The 22.2 Multichannel Sound System and Its Application," 118th AES Convention, 2005, Preprint No. 6406.
- [12] Hamasaki kimio, et al., "5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System," 117th AES Convention, 2004, Preprint No. 6226.
- [13] A. J. Berkhout, et al., "Acoustic control by wave field synthesis," J. Acoust. Soc. Am, 93(5), May 1993
- [14] Diemer de Vries, et al., "Experience with a Sound Enhancement System Based on Wavefront Synthesis," 95th AES Convention, 1993, Preprint No. 3748
- [15] Karlheinz Brandenburg, et al., "Wave Field Synthesis: From Reseach to Applications," Fraunhofer IDMT & Ilmenau Technical University
- [16] J.Daniel, J.B.Rault, and J.D. Polack, "Ambisonics encoding of other audio formats for multiple listening conditions," in Proc. 105th AES Convention, 1998, Preprint No. 4795.
- [17] O.Kirkeby and P.A.Nelson. Journal of the Acoustical Society of America 94, 2992-3000. Reproduction of plane wave sound fields, 1993.

필자 소개

서정일



- 1994년 : 경북대학교 전자공학과(공학사)
- 1996년 : 경북대학교 대학원 전자공학과(공학석사)
- 2005년 : 경북대학교 대학원 전자공학과(공학박사)
- 1998년 ~ 1999년 : LG 반도체 근무
- 1999년 ~ 2000년 : 현대 전자 근무
- 2000년 ~ 현재 : 한국전자통신연구원 방송통신융합연구부문 방통미디어연구부 선임연구원
- 주관심분야 : 음향신호처리, 음성신호처리, 객체기반 3차원 오디오 신호처리, MPEG, DMB 등

장인선



- 2001년 : 충북대학교 전기전자공학부(공학사)
- 2004년 : 포항공과대학교 대학원 컴퓨터공학과(공학석사)
- 2004년 ~ 현재 : 한국전자통신연구원 방송통신융합연구부문 방통미디어연구부 연구원
- 주관심분야 : 오디오 신호처리, 음원분리, 객체기반 3차원 오디오 신호처리, MPEG, DMB 등

필자소개



이용주

- 1999년 : 경북대학교 전자공학과(공학사)
- 2001년 : 경북대학교 대학원 전자공학과(공학석사)
- 2001년 ~ 현재 : 한국전자통신연구원 방송통신융합연구부문 방통미디어연구부 선임연구원
- 주관심분야 : 대화형 방송 시스템, 음향신호처리, 객체기반 3차원 오디오 신호 처리 등



유재현

- 2003년 : 홍익대학교 전자전기공학부(공학사)
- 2005년 : 서울대학교 대학원 전기컴퓨터공학부(공학석사)
- 2005년 ~ 현재 : 한국전자통신연구원 방송통신융합연구부문 방통미디어연구부 연구원
- 주관심분야 : 오디오 신호 처리, 가상 음성 정위 기술, 객체기반 3차원 오디오 신호 처리 등



강경욱

- 1985년 : 부산대학교 물리학과(이학사)
- 1988년 : 부산대학교 대학원 물리학과(이학석사)
- 2004년 : 한국항공대학교 대학원 항공전자공학과(공학박사)
- 1991년 ~ 현재 : 한국전자통신연구원 방송통신융합연구부문 방통미디어연구부 책임연구원
- 주관심분야 : 음향 신호처리, 3차원 오디오, 맞춤형 방송 등