



I. 서론

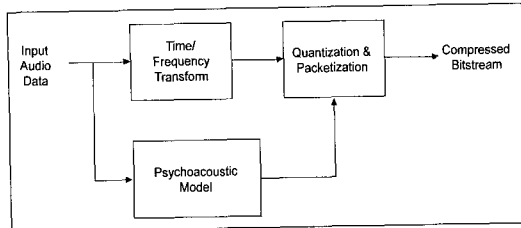
전화나 오디오 방송뿐만 아니라, TV방송, 화상전화, 영상회의, 대화형 멀티미디어 서비스(정보, 교육, 오락 등)와 같은 시각통신도 기본적으로 오디오를 필요로 하고, 미래의 문자, 그래픽, 팩스, 정지영상, 전자메일 등도 음성이나 오디오를 하나의 수단으로 제공하게 될 것이다. 또한, 전송 미디어로서뿐만 아니라 패키지 미디어로서 오디오가 갖는 장점도 매우 다양하다. 이러한 상황을 인지하여 보면 오디오 신호를 압축하기 위한 연구는 매우 중요하고 필요한 기술이 된다.

인터넷 환경의 비약적인 발전에 따른 MP3 오디오 콘텐츠의 급속한 보급과, 디지털 방송을 위한 오디오 압축 방식의 선정에 따른 논란으로 최근에 오디오 압축 방식에 대한 관심이 국내외에서 높아져 있다.

본 논문에서는 최근 수 년간 국제표준으로 채택되었거나 활발히 연구 중인 최신 오디오 압축 방식을 분석하고 연구 동향을 소개하고자 한다.

II. 오디오 압축 방식

오디오 신호를 압축하는 방법은 신호 파형 특성을 이용하는 무손실 압축(Lossless Coding) 방식과 사람의 청각 특성을 이용하는 손실 압축(Lossy Coding) 방식으로 구분할 수 있다. 무손실 압축 방법은 원본 오디오 신호와 압축 과정을 거친 오디오 신호가 파형상으로 일치한다는 특성을 가진다. 그러나 손실 압축 방식은 원본 오디오 신호와 청각적으로 차이가 없는 한계까지 오디오 신호를 압축하는 방식이므로 압축된 결과의 파형은 원본 오디오 신호와 많은 차이를 보이게 된다. 시간축상이나 주



〈그림 1〉 심리음향모델을 이용한 오디오 압축방식의 개념도

과수축상에 인접한 오디오 신호들 간에 에너지가 큰 신호에 의해서 상대적으로 작은 에너지를 가지는 신호가 들리지 않는 현상을 마스킹 효과(masking effect)라고 한다. 이러한 마스킹 효과를 모델링한 것을 심리음향모델(Psycho-acoustic Model)이라고 하며, 심리음향모델은 오디오 신호와 마스킹 문턱치(Masking Threshold)간의 에너지비(Signal to Masking Ratio, SMR)를 이용하여 양자화 비트수를 결정하는 정보를 제공하게 된다. 〈그림 1〉은 심리음향모델을 이용하여 오디오 신호를 압축하는 손실 압축방식의 개념도를 도시한 것이다. 대표적인 손실 압축방식으로는 MPEG-1, MPEG-2 BC, MPEG-2/4 AAC, AC-3, WMA 등이 있다.

3장에서는 MPEG을 중심으로 표준화가 진행중이거나, 새로운 오디오 압축 방식으로 논의되고 있는 알고리즘들에 대해서 간략하게 설명한다.

Ⅲ. 최신 오디오 압축 방식

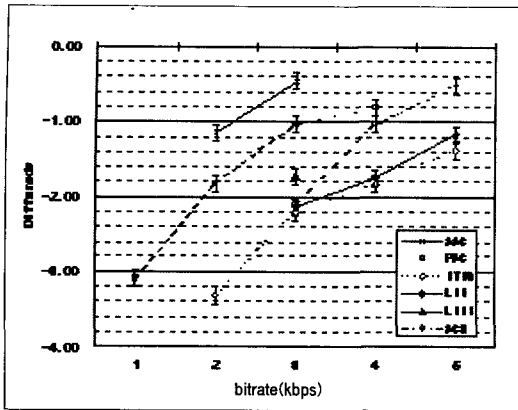
1. AAC

MPEG에서는 1994년부터 MPEG-1/2 포맷과의 호환이 필요 없는 응용 분야에의 활용을 위한

새로운 표준화 작업에 들어갔는데, 표준화 과정에서 붙여진 이름은 MPEG-2 NBC (Non-Backward Compatible)였다. MPEG-2 NBC 목적은 384kbps 이하의 비트율로 5채널 전대역폭 오디오 신호가 ITU-R에서 정의하는 “청각적으로 구분할 수 없는(indistinguishable)” 오디오 음질에 도달하는 것이었으며, 1996년 가을 BBC와 NHK에서 실시한 테스트에서 MPEG-2 NBC는 ITU-R에서 요구하는 음질을 만족함을 보여 주었다. 1997년 4월에는 새로운 멀티채널 오디오 압축 방식의 국제 표준으로 발표되었는데 이때의 이름이 MPEG-2 AAC (Advanced Audio Coding) 이다. MPEG-2 AAC의 기본 알고리즘에 PNS와 LTP 압축 도구를 추가하여 MPEG-4 AAC로 확장되어 표준화 되었다.

AAC는 표준화 단계부터 MPEG-1/2 오디오와의 호환성을 고려하지 않았으므로, 전 세계의 진보된 오디오 압축 기술을 가진 기관들이 협력하여 음질과 압축율면에서 최고의 성능을 가지도록 한 결과이다. AAC는 매우 낮은 비트율에서 방송 음질 수준의 오디오를 제공하기 위해 고해상도 필터뱅크(high-resolution filter bank), 예측 기법(prediction), 시간 영역 양자화 잡음 제거기법(temporal noise shaping), 허프만 부호화(Huffman coding) 등을 결합하여 사용한다.

〈그림 2〉는 CRC에서 5.1채널 오디오 데이터를 AAC와 MP3, AC-3, PAC 등으로 압축하여 주관적인 테스트를 한 결과를 도표로 나타낸 것이다. 〈그림 2〉에서 나타나는 것과 같이 MPEG-2 AAC를 이용하여 5.1채널 오디오 신호를 320kbps 로 압축 하였을 때의 음질이 480kbps 로 압축한 AC-3, MPEG-2 layer-II 보다도 우수



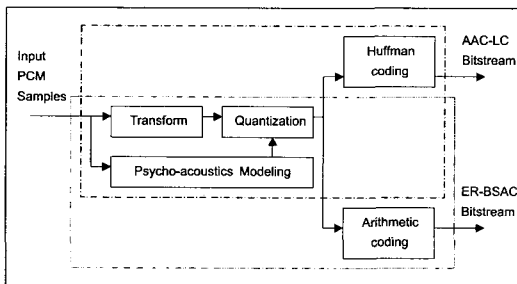
〈그림 2〉 MPEG-2 AAC 음질 테스트 결과

한 결과를 나타내었다.

2. BSAC

삼성전자가 제안하여 MPEG-4 오디오 압축 방식의 표준 기술 중의 하나로 채택된 BSAC (Bit Sliced Arithmetic Coding)은 〈그림 3〉과 같이 AAC의 압축 방식에서 허프만 부호화 방식의 양자화 부분을 산술 부호화(Arithmetic Coding) 방식으로 대체하는 방법이다.

〈그림 4〉와 같이, 산술 부호화 방식을 사용함으로써 비트 슬라이스 단위로 비트율의 조절이 가능



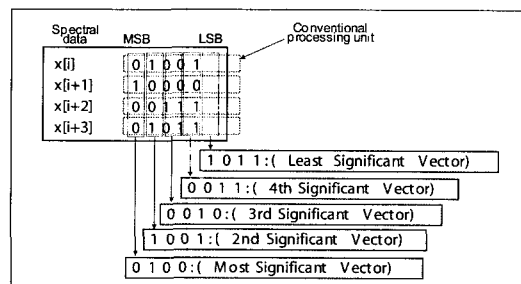
〈그림 3〉 AAC 와 BSAC의 부호화기 개념도

하므로, 전송 선로의 환경에 따라 오디오 신호의 비트율을 능동적으로 변화시키는 기능을 지원할 수 있으며, 기본적으로 AAC와 동일한 압축 알고리즘을 사용하므로 같은 압축률에서 AAC와 유사한 음질을 제공할 수 있다.

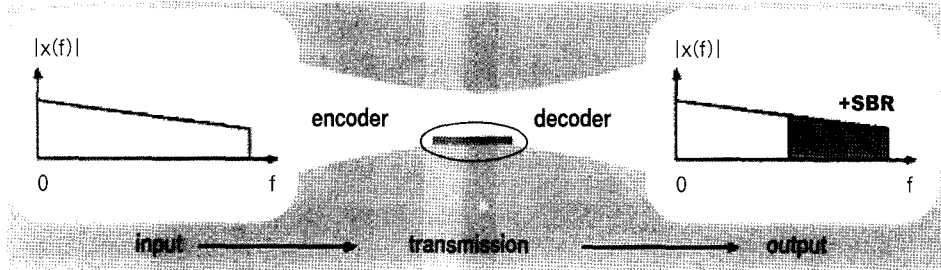
3. SBR

채널당 64kbps 부근에서 “CD-quality” 또는 “transparent” 오디오 음질을 제공하는 AAC와 같은 지각 오디오 부호화 방식에서 압축률을 계속해서 높일 경우 주파수 대역별로 할당되는 비트의 수가 마스킹 문턱치(masking threshold)를 표현 하는데 필요한 비트수보다 적어지기 때문에 음질이 저하된다.

대역폭 확장 도구인 SBR (Spectral Band Replication)은 대역폭이 제한된 오디오 신호의 오디오 대역폭을 확장한다. 기본적인 절차는 비트율을 감소 시키기 위해 제한되었던 고주파의 열을 인코더에서 부호화된 대역폭이 제한된 신호와 제어 데이터에 의해 복제하는 것이다. 주기 신호 성분과 잡음 성분 사이의 비율은 적응 역 필터와 잡음 및 정현파 성분을 추가함으로써 유지 된다.



〈그림 4〉 BSAC에서 비트 슬라이스 개념도



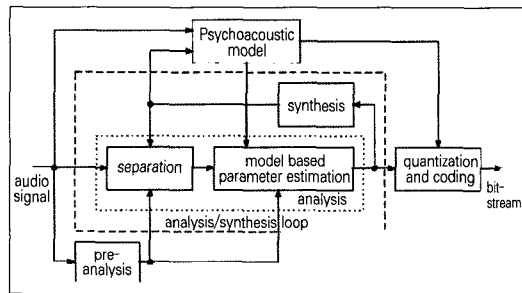
〈그림 5〉 SBR 알고리즘의 개념도

〈그림 5〉는 SBR 알고리즘을 이용하여 고주파수 대역을 복원하는 과정을 나타낸 개념도이다. 입력 오디오 신호를 다운 샘플링하여 주파수 대역을 줄인 후 저주파수 성분은 기존의 지각 오디오 부호화 방식 등을 이용하여 부호화하고 다운 샘플링 하지 않은 신호를 이용하여 SBR 알고리즘으로 고주파수 대역을 부호화하는 것이다. 여기서 원으로 표시된 부분이 추가되는 비트열과 복원된 고주파수 대역을 나타낸다.

CodingTechnologies 사에 의해서 제안된 SBR 알고리즘은 MP3에 적용되어 mp3PRO란 이름으로 사용되고 있으며, MPEG에서는 MPEG-4 Audio Extension 1: Bandwidth Extension (BWE)라고 명명되어 2003년도 하반기에 국제표준(ISO)으로 제정될 예정이다. AAC에 SBR이 적용된 것을 aacPlus라 부르는데, aacPlus는 AAC가 CD급의 음질을 제공하는 채널당 48kbps 이상에서는 AAC 비하여 음질 개선 효과가 없지만, AAC의 음질이 저하되는 32kbps 이하에서는 우수한 음질을 제공하는 장점을 가진다. aacPlus는 유럽에서 진행중인 아날로그 AM 라디오 방송을 대체하는 DRM (Digital Radio Mondail) 표준에서 표준 오디오 압축 방식으로 채택되어 있으며, 국내외 DMB 방송에서도 적용될 가능성이 높다.

4. High Quality Parametric Audio Coding

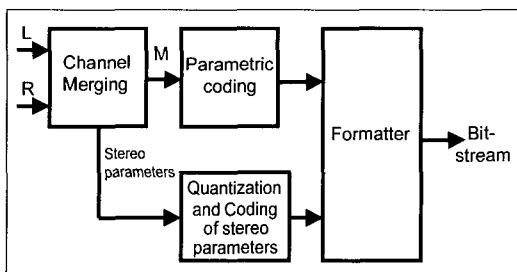
저 비트율 오디오 압축 방식으로 사용되는 파라메트릭 오디오 압축 방식의 기본적인 원리는 신호를 기술하는 파라미터들을 추출하기 위해 입력 신호를 분석하는 것이다. 프레임 단위의 입력신호는 신호의 포락선과 대략적인 기본주파수 추출을 포함하는 예비 신호 분석 과정을 거친 다음 고조파 성분, 개별적인 정현파 성분 그리고 잡음성분으로 구분한다. 고조파 성분과 개별적인 정현파 성분으로 구성되는 정현파 성분의 추출은 분석-합성 방식을 통한 페루프 방식을 통해 이루어진다. 이때 각 반복 주기에 심리음향 모델을 적용하여 심리음향적으로 가장 중요한 정현파 성분을 먼저 추출하게 되며 나



〈그림 6〉 HILN 부호화기의 블록도

머지 잔류 성분은 잡음으로 분류된다. 고조파 성분, 개별적인 정현파 성분, 잡음 성분으로 구분된 파라미터들의 측정이 완료된 다음 양자화된다. <그림 6>은 MPEG-4에 표준으로 채택된 HILN (Harmonic and Individual Lines plus Noise) 부호화기의 블록도이다.

고음질 파라메트릭 오디오 압축 방식은 스테레오 입력 오디오 신호를 하나의 모노 채널로 믹싱하고, 스테레오 신호에 대한 위치정보는 별도의 파라미터를 이용하여 처리한다. 스테레오 신호를 하나의 모노 채널로 믹싱함으로써 압축효율을 높이고 일련의 압축과정의 효율을 높임으로써 낮은 비트레이트에서도 고음질을 얻을 수 있다. <그림 7>은 고음질 파라메트릭 오디오 압축방식의 부호화기의 블록도이다.



<그림 7> 고음질 파라메트릭 오디오 부호화기

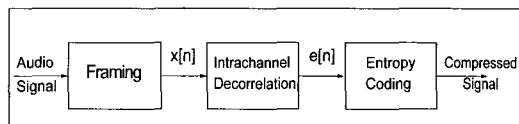
고음질 파라메트릭 압축방식은 MPEG-4 Audio Extension 2: High Quality Parametric Audio Coding 으로 명명되어 2003년도 하반기에 국제표준(ISO)으로 제정될 예정이다.

5. Lossless Audio Coding

손실 압축 방식(Lossy Audio Coding)이 약

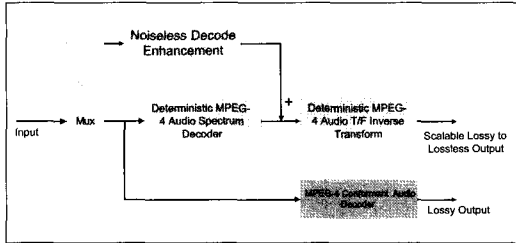
12:1정도의 압축효율을 얻을 수 있는 반면 무손실 압축 방식(Lossless Audio Coding)은 기껏해야 3:1정도의 압축효율을 얻을 수 있었기 때문에 손실 압축 방식에 비하여 각광받지 못했다. 그러나, 손실 압축 방식의 압축률을 높이면 음질의 저하는 피할 수 없고, CD로부터 취득할 수 있는 최고의 음질을 인터넷을 이용하여 자신의 하이파이 오디오 시스템으로 듣고 싶어하는 사용자의 요구가 증가하고, 디지털 시네마와 같은 네트워크를 통한 영화 배급 환경에서 최고의 오디오 음질을 유지하기 위하여 무손실 압축 방식이 대안으로 떠오르고 있다. 위와 같은 산업계의 요구로 MPEG에서는 MPEG-4 Extension 3: Lossless Audio Coding 이라는 이름으로 표준화 작업이 진행 중이다.

<그림 8>은 무손실 압축 방식 부호화기에서 하나의 채널에 대한 압축 과정을 나타낸다. 입력 오디오 신호를 적절한 크기(frame)로 나눈 후 선형 예측 기법(Linear Prediction)을 이용하여 신호내에 포함되어 있는 과잉 성분(redundancy)을 제거한다. 마지막으로 엔트로피 부호화를 이용하여 압축률을 높이게 된다.



<그림 8> 무손실 압축방식 부호화기 개념도

또한 사용자의 요구에 따라 <그림 9>와 같이 낮은 비트율의 손실 압축 방식 비트스트림과 무손실 압축 방식 비트스트림을 동시에 보낼 수 있게하는 스케일러블 무손실 부호화기 구조도 MPEG에서 활



〈그림 9〉 스케일러를 무손실 부호화기

발히 논의되고 있다.

6. Synthetic Audio Coding

합성 오디오 압축 방식(Synthetic Audio Coding)은 심리음향모델을 사용하지 않고 정밀한 파라메트릭 모델을 이용하여 오디오 신호를 기술(description)하여 수 kbps까지 오디오 신호를 압축하는 방식이다.

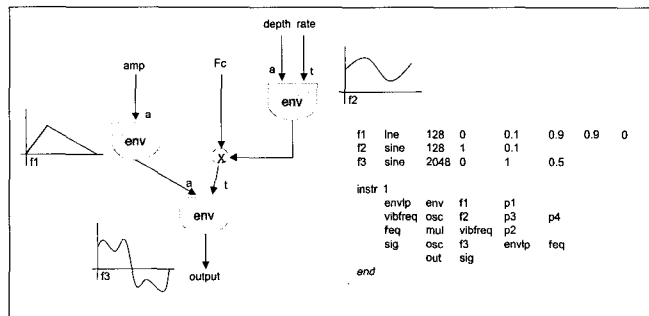
오디오 합성 알고리즘은 크게 학술적인 접근과 산업적인 접근으로 나눌 수 있다. 학술적인 접근은 1960년대 초 Mathew가 개발한 소프트웨어 기반 오디오 합성을 유닛 생성 모델(unit generation model)을 시초로 한다. 유닛 생성 모델에서는 오디오 신호 생성 알고리즘이 컴퓨터 언어로 표현되며, 특히 디지털 신호처리 네트워크를 통하여 오디오 신호가 생성되게 하였다. 유닛 생성 언어는 오디오 신호에 대한 일반적인 축약이라 볼 수 있으므로 어떠한 형태의 오디오 신호도 표현 가능하다.

산업적인 접근은 조금 늦게 1960년도 중반 Moog, Buchla, Oberheim과 같은 사람들이 아날로그 신서사이저(analog synthesizer)

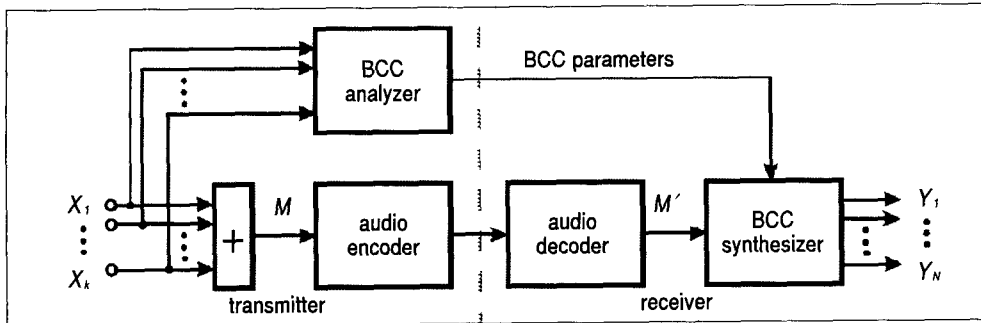
를 개발하면서 시작되었다. 아날로그 신서사이저는 한가지 파라메트릭 합성방법이 하드웨어로 제작된 형태이며, 음악가들이 프론트 패널의 노브를 이용하여 파라미터를 제어함으로써 다양한 합성음을 생성할 수 있게 하였다. 아날로그 신서사이저는 비록 단순한 구조를 가지고 있었으나 실시간 처리가 가능했으므로 음악가들이나 작곡가들에게 널리 사용되었다. 1980년대 중반부터 마이크로프로세서의 성장과 효율적인 합성 알고리즘의 개발에 힘입어 디지털 신서사이저가 개발되기 시작했다. 초기의 디지털 신서사이저는 주파수 변조와 샘플러에 기반한 웨이브 테이블 합성방법을 이용하였고, PC의 혁명적인 발전 후 하드웨어 디지털 신서사이저는 플러그인 카드 형태로 집적되게 되었다.

위와 같은 합성 오디오 압축방식의 발전에 힘입어 MPEG에서는 아래와 합성 오디오 압축방식 기술들을 표준화 하였다.

- SAOL (Structured Audio Orchestra Language): 사운드 유닛 생성 언어
- SASL (Structured Audio Score Language): 사운드 생성 명령 언어
- SASBF (Structured Audio Sample Bank Format): 웨이브 테이블을 합성 방식
- TTSI (Text-to-Speech Interface): 문자/음성 변환 시스템과의 인터페이스



〈그림 10〉 유닛 생성 모델의 신호흐름도와 유닛 생성 언어



〈그림 11〉 바이노럴 큐 압축방식 개념도

7. Binaural Cue Coding

바이노럴 큐 압축 방식은 멀티채널 오디오 신호의 음향공간상의 위치정보를 파라미터화하여 모노 채널로 믹싱하고 재구성하는 파라메트릭 압축 방식의 일종이다. 3차원 오디오 신호와 같이 다채널 다객체의 환경에서는 채널수나 오디오 신호의 객체수에 비례하여 오디오 신호를 전송하기 위한 대역폭 또한 커지게 된다. 이와 같은 오디오 신호를 하나의 모노채널로 믹싱하여 압축하고 객체와 채널에 대한 정보는 별도의 파라미터로 표현하여 전송하면 압축 효율을 극대화 할 수 있으므로 바이노럴 큐를 이용하는 압축방식이 최근 제안되었다.

〈그림 11〉은 바이노럴 큐를 이용한 압축방식의 부호화기와 복호화기를 나타낸 것이다. 멀티채널 오디오 신호는 하나의 채널로 믹싱되어 부호화되며, 각 채널의 음향공간상의 특징은 BCC 해석기를 통하여 파라미터화된다. 전송된 모노 채널에 대한 비트스트림은 복호화 과정을 거친 후 BCC 파라미터를 이용하여 원래의 채널로 분리되게 된다.

IV. 결론

본 논문에서는 오디오 신호를 압축하기 위하여 최근에 국제표준으로 채택되었거나 활발히 논의되고 있는 알고리즘들에 대하여 살펴 보았다. 방송과 통신의 융합과 사용자들의 다양한 욕구를 충족시키기 위해서는, 오디오 압축 방식도 저비트율 압축 방식에서부터 대용량 고품질 무손실 압축 방식에까지 다양한 성능을 요구하게 되었으며 이에 적절한 오디오 압축방식들이 계속해서 제안되고 MPEG을 비롯한 표준화 단체에서 논의되고 있다. 국내에서는 삼성전자에서 BSAC을 MPEG-4의 표준으로 채택시킨 것 외에는 뚜렷한 연구동향을 보이지 않았으나, 새로운 서비스 요구사항과 사용자 요구사항에 맞는 새로운 오디오 압축방식에 대한 지속적인 연구가 이루어지는 것이 바람직하다.

● 참고 문헌 ●

- (1) ISO/IEC Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 7: Advanced Audio Coding (AAC), ISO/IEC 13818-7
- (2) ISO/IEC Information Technology - Coding of Audio-Visual Objects - Part 3: Audio, ISO/IEC 14496-3
- (3) ISO/IEC, ISO/IEC 14496-1:2001 FDAM1 Bandwidth Extension, ISO/IEC JTC1 SC29 WG11 N5570, 2003-03
- (4) ISO/IEC, ISO/IEC 14496-1:2001 PDAM2 Parametric coding for High Quality Audio, ISO/IEC JTC1 SC29 WG11 N5381, 2002-12
- (5) Mat Hans and Ronald W. Schafer, "Lossless Compression of Digital Audio," IEEE Signal Processing Magazine, July 2001
- (6) Frank Baumgarte and Christo Faller, "Design and Evaluation of Binaural Cue Coding Schemes," 113th AES Convention, October 2002

필자 소개



서 정 일

- 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 방송미디어연구부 연구원
- 주관심분야 : 음향신호처리, 음성신호처리, 객체기반 3차원 오디오 신호처리, MPEG, DMB



강 경 옥

- 1995년 2월 : 부산대학교 물리학과 학사
- 1988년 2월 : 부산대학교 대학원 물리학과 석사
- 1993년 3월~현재 : 한국항공대학교 대학원 전자공학과 박사과정
- 1991년~현재 : 한국전자통신연구원 방송미디어연구부 방송컨텐츠응용연구팀장(선임연구원)
- 주관심분야 : 디지털 방송, 메타데이터 신호처리, 음향신호처리, 객체기반 3차원 오디오 신호처리 기술



장 대 영

- 1991년 2월 : 부경대학교 전자공학과 졸업(공학사)
- 2000년 2월 : 배재대학교 대학원 컴퓨터공학과 졸업(공학석사)
- 2000년 3월~현재 : 배재대학교 대학원 컴퓨터공학과 박사과정
- 1991년 1월~현재 : 한국전자통신연구원 방송미디어연구부 선임연구원
- 주관심분야 : 음향 신호처리, 멀티미디어 응용, 객체기반 3차원 오디오 신호처리



김 진 웅

- 1981년 2월 : 서울대학교 공과대학 전자공학과 졸업(학사)
- 1983년 2월 : 서울대학교 대학원 전자공학과 졸업(석사)
- 1993년 8월 : Texas A&M Univ. 전기전자공학과 졸업(박사)
- 1983년 3월~현재 : 한국전자통신연구원 책임연구원/방송미디어연구부장
- 주관심분야 : 영상통신, 멀티미디어 방송