

특집논문-05-10-3-07

저비트율 멀티채널 오디오 부호화

장인선^{a)*}, 서정일^{a)}, 백승권^{a)}, 강경욱^{a)}

Low-bitrate Multichannel Audio Coding

Inseon Jang^{a)*}, Jeongil Seo^{a)}, Seungkwon Beack^{a)} and Kyeongok Kang^{a)}

요 약

멀티채널 오디오 콘텐츠에 대한 사용자의 욕구가 증가하고 있는 가운데 서비스를 위한 저비트율 멀티채널 오디오 부호화 기술에 대한 표준화가 진행되고 있다. 본 논문에서는 협대역폭 채널환경에 적합한 고압축율의 멀티채널 오디오 부호화 기술인 음원위치단서 부호화 기술(Sound Source Location Cue Coding; SSLCC)을 제안하고, SSLCC가 기존 BCC (Binaural Cue Coding)의 압축 성능을 향상시키기 위하여 단서로 이용한 VSLI (Virtual Source Location Information)를 설명한다. 또한 객관적 평가로 기존의 BCC 시스템과의 평균 비트율을 비교하고, MUSHRA (Multi-Stimulus test with Hidden Reference and Anchor) 방법을 이용한 주관 청취 평가를 수행하여 제안된 SSLCC 부호화 기법의 우수성을 확인하였다.

Abstract

Technology for compressing low-bitrate multichannel audio coding is being standardized owing to the increasing need of consumer for multichannel audio contents. In this paper we propose the sound source location cue coding (SSLCC) for extremely compressing multichannel audio to be suitable at the narrow bandwidth transmission environment. To improve the compression capability of the conventional binaural cue coding(BCC), the SSLCC adopts the virtual source location information (VSLI) as a spatial cue parameter, a symmetric uniform quantizer, and Huffman coder. The objective and subjective assessment results show that the SSLCC provides lower bitrate and better audio quality than conventional BCC method.

Keywords : Spatial Audio Coding, 가상음원 위치정보, Binaural Cue Coding

I. 서 론

오늘날 DVD 등 멀티채널 소스의 폭발적인 증가와 사용

자들의 멀티채널에 대한 욕구 증가에 따라 통신환경을 통한 멀티채널 오디오 서비스의 필요성이 증가되고 있다. 하지만 신호전송을 위한 대역폭의 제약성, 저장 용량의 한계성과 사용자가 수용할 수 있는 비용 등의 제약 조건 등으로 인하여 멀티채널 오디오 콘텐츠를 효율적으로 전송하기 위해서는 새로운 압축 기술이 필요하다. 이를 위해 저비트율 부호화 기술로 binaural cue coding(BCC) 기법이 연구되어 왔으며, 이는 현재 MPEG에서 Spatial Audio Coding (SAC)라는 이

a) 한국전자통신연구원 디지털방송연구단 방송미디어연구그룹
Broadcasting Media Research Group, Digital Broadcasting Research
Division, ETRI

※ 본 논문은 정보통신부의 연구사업인 "지능형 통합정보 방송 기술개발" 과제의 일환으로 수행한 결과로서 정보통신부 담당자 및 관련 연구원들의 노력에 감사를 드립니다.

름으로 표준화가 진행 중이다^[1].

SAC는 최소의 채널(모노 혹은 스테레오)로 합쳐진 다운믹스 오디오 신호와 멀티채널 오디오 신호에 대한 인간의 지각 특성을 나타내는 공간 단서(spatial cue) 파라미터들로 구성된 부가정보를 이용하여 멀티채널 오디오 신호를 부호화하는 기술이다^[2]. 이때 사용된 다운믹스 오디오 신호는 독립적으로 서비스가 가능하므로 기존의 모노 또는 스테레오 오디오 부호화 시스템과 호환성을 제공할 수 있다. 그리고 멀티채널 오디오 신호를 표현하기 위한 부가정보는 매우 작은 정보량(채널당 4kbps 이내)으로 전송이 가능하므로 모노 또는 스테레오 오디오 서비스만 제공할 수 있는 대역폭에서도 고품질의 멀티채널 오디오 서비스를 제공할 수 있다.

대표적인 SAC 기술인 binaural cue coding (BCC)^{[3][4]} 기술은 사람이 오디오 신호의 공간적 위치를 인지하는 특성인 양이효과(binaural effect)를 멀티채널 오디오에 적용한다. 즉, 양이효과 파라미터인 두 귀 사이의 소리 크기 차이(interaural level difference; ILD), 두 귀 사이의 도달 시간 차이(interaural time difference; ITD) 및 두 귀 사이의 상관도(interaural correlation: IC)를 각 채널간의 크기 차이(inter-channel level difference: ICLD), 채널간의 시간 차이(inter-channel time difference; ICTD) 및 채널간의 상관도(inter-channel coherence: ICC)로 표현하여 이를 멀티채널 오디오의 지각 단서로 사용한다. 이들 중 ICLD는 기준이 되는 채널과 해당하는 채널간의 전력비로 계산되며, BCC 시스템의 성능을 결정하는 데 가장 중요한 역할을 담당한다.

최근에 스피커의 기하학적 배치로부터 얻어진 가상음원의 위치로써 채널간의 전력비를 나타내는 파라미터인 가상음원 위치정보(Virtual Source Location Information;

VSLI)^[5,6,7,8,9]가 제안되었다. VSLI는 중앙 채널로부터의 각도로 표현되며, 이는 멀티채널 오디오 신호의 각 채널별 전력으로부터 constant power panning (CPP)^[10] 기법을 역으로 적용하여 구할 수 있다. 특히 VSLI는 양자화 할 때 최대 각도로 고정된 범위범위를 사용하므로 이론상 무한대의 동적 범위(dynamic range)가 필요한 ICLD에 비해 양자화 오차에 강한 특성이 있다.

본 논문에서는 협대역폭 채널환경에 적합한 고압축율의 멀티채널 오디오 부호화 기술인 음원위치단서 부호화(sound source location cue coding; SSLCC)기술을 제안한다. SSLCC는 기존 BCC의 압축 성능을 향상시키기 위해 양자화 오차에 강한 VSLI를 지각단서로 이용하고, 대칭형 균일 양자화기 허프만 부호화를 이용한다.

제안한 시스템의 우수성을 검증하기 위하여 객관적 평가로 기존의 ICLD에 기반한 BCC 시스템과 SSLCC 시스템의 실험데이터에 대한 평균 비트율을 비교하였으며, MUSHRA (Multi-Stimulus test with Hidden Reference and Anchor) 방법을 이용하여 주관 청취 평가를 수행하였다.

II. Binaural Cue Coding

Faller에 의해 제안된 BCC는 멀티채널 오디오 신호를 최소한의 채널로 다운믹스 된 오디오 신호와 부가정보로 표현되는 채널들에 대한 BCC 파라미터로 분리하여 전송하고, 다시 합성하는 파라미터 기반 멀티채널 오디오 부호화 기술이다(그림 1 참조)^{[3][4]}. 각 채널에 대한 BCC 파라미터는 채널당 4kbps 정도의 대역폭에서 표현이 가능하므로 멀티채

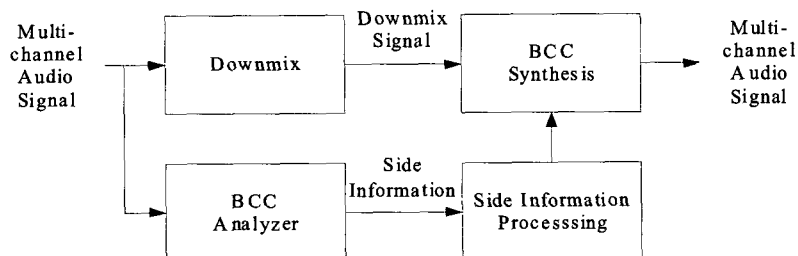


그림 1. Binaural cue coding 시스템 개념도
Fig. 1. Block diagram of binaural cue coding system

널 오디오 서비스를 모노 또는 스테레오 오디오 서비스의 대역폭에서도 제공할 수 있다^[11]. 또한 멀티채널 신호에 대한 정보는 BCC 파라미터들로 구성된 부가정보의 형태로써 독립적으로 처리 및 전송되므로, 기존의 모노 또는 스테레오 오디오 서비스와 호환성을 유지할 수 있는 장점이 있다^[12].

일반적인 BCC 부호화기에서는 시간-주파수 변환과 Equivalent Rectangular bandwidth (ERB) 필터뱅크를 통하여 시간-주파수 격자로 분리된 멀티채널 오디오 신호에 대해서 기준 채널과 다른 채널들 간의 상대적인 공간단서 파라미터들을 BCC 파라미터로 사용한다. 기준 채널과 다른 채널들 간의 BCC 파라미터는 ICLD, ICTD, ICC 등이 있다. 그림 2은 ERB 부대역 기준채널과 임의의 채널간의 BCC 파라미터를 추정하는 과정을 도시한 것이다. 임의의 ERB 부대역에서의 스펙트럼들의 상관도를 계산하여 최대 상관도를 ICC로 결정하고, 이때의 시간 지연값을 ICTD로 예측한다. 또한 ICLD는 ICTD값에 의한 시간지연을 보정한 후 두 채널간의 전력비를 이용하여 계산된다.

다채널 및 다객체 환경에서는 전송을 위한 대역폭이 채널

수나 오디오 객체 수에 의해 비례하여 증가한다. 그러나 다 채널 및 다객체 오디오 신호를 하나의 모노 채널로 다운믹스하여 압축하고 객체들과 채널들에 대한 정보는 별도의 부가정보로 표현하여 전송하면 압축효율을 극대화 할 수 있다. BCC는 멀티채널 오디오 신호를 최소한의 채널로 다운믹스한 오디오 신호와 멀티채널 오디오 신호의 음향 공간상의 단서로 이루어진 부가정보로 분할하여 압축하는 파라미터 오디오 부호화 기술이다.

이와 관련하여 MPEG 오디오 그룹에서는 MPEG-4 SAC 라는 이름으로 표준화가 진행 중이며 채널당 4-8kbps 정도의 부가정보로써 멀티채널 오디오 신호를 표현하고 기존의 모노 또는 스테레오 복호화에 대한 호환성을 제공하는 것을 목적으로 한다. MPEG-4 SAC 부호화기는 멀티채널 오디오 신호를 모노 또는 스테레오 채널로 다운믹스한 후 기존의 MPEG-4 오디오 도구(MPEG-4 AAC, MPEG-4 HE-AAC 등)로 압축하고 멀티채널 오디오 신호로부터 공간단서를 추출하여 다운믹스 오디오 신호와 다중화 한다. MPEG-4 SAC 복호화기는 역다중화기를 이용하여 다운믹스 오디오

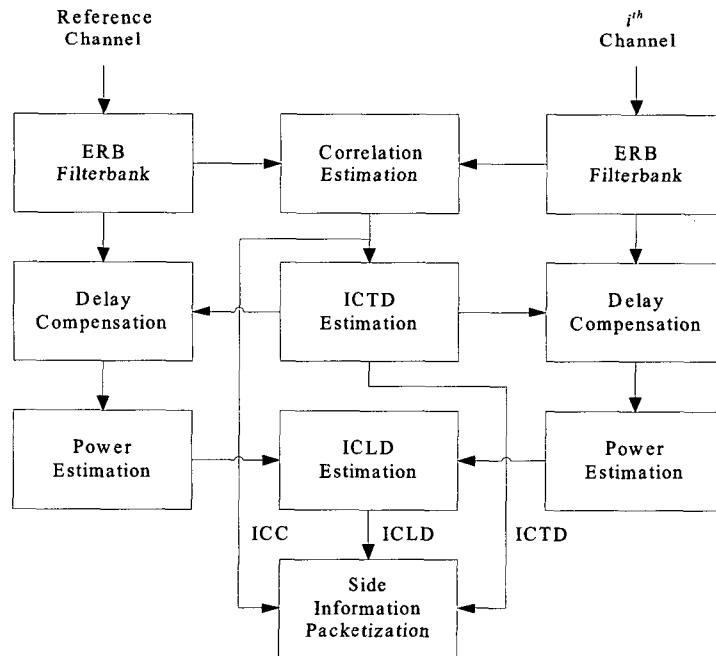


그림 2. BCC 파라미터 추정과정 개념도
 Fig. 2. The conceptual diagram of BCC parameter estimation

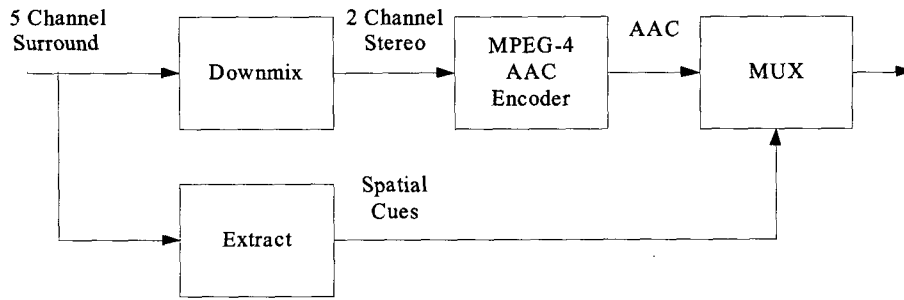


그림 3. MPEG-4 SAC 부호화기 개념도.
Fig. 3. The scheme of MPEG-4 SAC encoder

신호와 공간단서를 분리하고 다운믹스 오디오 신호에 공간단서를 적용하여 멀티채널 오디오 신호를 합성한다. 그림 3, 4는 멀티채널 오디오 신호를 스테레오로 다운믹스하고 MPEG-4 AAC를 이용하여 압축할 때의 MPEG-4 SAC 부호화기와 복호화기의 처리과정을 도시한 것이다.

의 채널 간의 전력벡터는 채널 사이에서의 음상의 위치를 결정하므로 최대 5개의 음상(S1~S5)이 존재할 수 있으며, 음상의 위치는 각도로 표현할 수 있다.

Ⅲ. 음원위치단서 부호화 시스템

1. 가상음원 위치정보

BCC 파라미터 중 음원의 위치를 인지하는 데에 가장 중요한 영향을 미치는 것은 ICLD이다. 양자화 하는 과정에서 ICLD 값의 동적범위(dynamic range)를 최대한 크게 할당하여야 음질의 왜곡을 최소화 할 수 있으며 일반적으로는±18dB로 설정하여 사용하고 있다. 그러나 클래식 음악과 같이 채널 간의 신호 크기의 차이가 큰 신호에서는 최대 변위를 초과하는 신호크기 차이가 존재하며 이로 인한 합성 음질의 왜곡이 발생한다.

가상음원 위치정보(Virtual Source Location Information; VSLI)는 멀티채널 오디오 신호의 크기 정보를 멀티채널 스피커 배치로부터 예측되는 음원의 위치를 중앙 채널과의 방위각으로 표현하므로 양자화 과정에서의 분해능(resolution)을 높일 수 있고, 양자화기의 변위범위를 파라미터가 표현할 수 있는 변위범위와 동일하게 정의할 수 있다.

그림 5와 같은 ITU-R BS. 775-1 규격^[13]으로 스피커들이 위치가 고정되어 있다고 가정하자. 인접한 스피커 사이에서

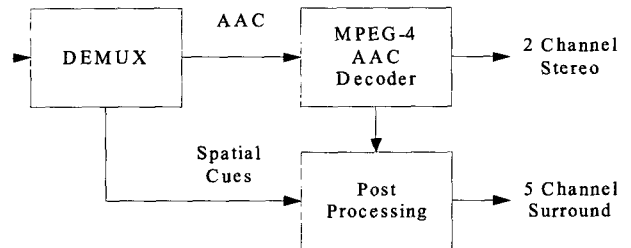


그림 4. MPEG-4 SAC 복호화기 개념도
Fig. 4. The scheme of MPEG-4 SAC decoder

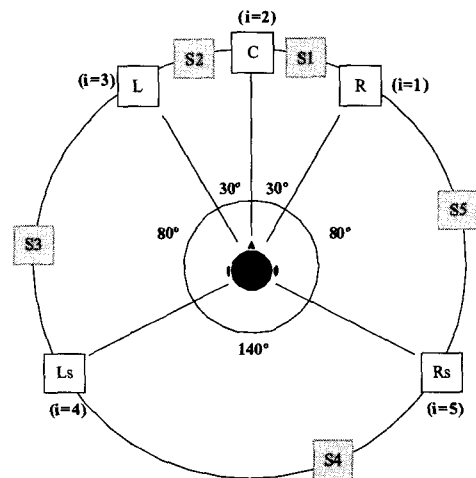


그림 5. 5채널 시스템에서의 스피커 배치
Fig. 5. Playback loudspeaker layout for five-channel system.

시간-주파수 변환과 ERB 필터뱅크를 통해 시간-주파수 격자로 분리된 멀티채널 오디오 신호에 대해서 k번째 프레임, i번째 채널, 시간-주파수 격자 b에서의 신호 크기를 L_{bk}^i 라 하면, i+1번째 채널에서의 신호크기는 L_{bk}^{i+1} 라 할 수 있다. CPP 기법^[10]을 역으로 적용하면 두 채널의 신호크기로부터 패닝 각도를 추정할 수 있다. CPP 기법을 적용하기 위해 두 채널간의 신호크기 정보를 정규화 하는 과정은 다음 수식과 같다.

$$g_{bk}^i = \frac{L_{bk}^i}{\sqrt{(L_{bk}^i)^2 + (L_{bk}^{i+1})^2}} \quad (1)$$

$$g_{bk}^{i+1} = \frac{L_{bk}^{i+1}}{\sqrt{(L_{bk}^i)^2 + (L_{bk}^{i+1})^2}} \quad (2)$$

일정전력패닝기법을 역으로 적용하면 크기 정보로부터 음원과 인접한 두 채널의 방위각 $\theta_{bk}^{i,i+1}$ 을 아래 수식과 같이 추정할 수 있다.

$$\theta_{bk}^{i,i+1} = \cos^{-1}(g_{bk}^i) \quad (3)$$

$$\theta_{bk}^{i,i+1} = \sin^{-1}(g_{bk}^{i+1}) \quad (4)$$

따라서 패닝 각도로 표현되는 i번째 채널과 i+1 사이의 가상음원의 위치 (θ_{vs})는 아래 수식과 같이 추정된다.

$$\theta_{vs} = \left(\theta_{bk}^{i,i+1} \times \frac{2}{\pi} \right) \times (\theta_{i+1} - \theta_i) + \theta_i \quad (5)$$

다섯 채널 재생 시스템에서는 다섯 개의 채널 간 전력벡터(inter-channel power vector)가 존재한다. 그러나 네 개의 채널 간 전력벡터를 알고 있으면 나머지 한 개의 채널 간 전력 벡터는 그림 6에서 도시한 바와 같이 좌평면 보조 벡터(left half plane subsequent vector; LSv), 좌평면 벡터(left half plane vector; LHv), 우평면 보조벡터(right half plane-subsequent vector; RSv), 우평면 벡터(right half plane vec

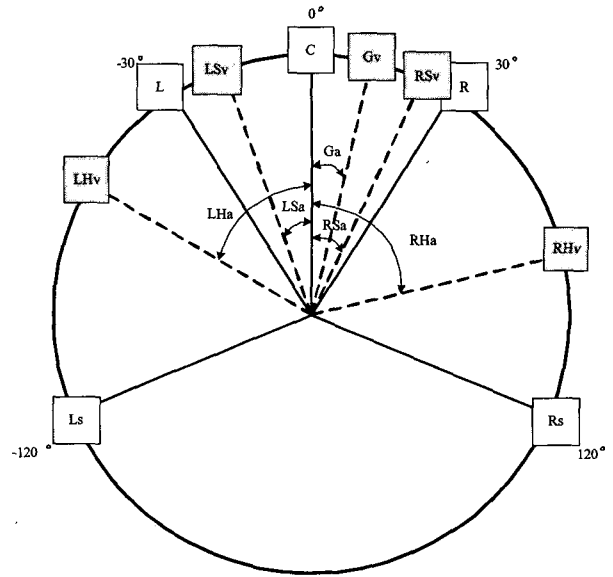


그림 6. MV-VSLI에서의 방위각들
Fig. 6. Component angles of MV-VSLI.

tor; RHv)를 이용하여 정의할 수 있다.

다섯 채널 신호를 두 채널 스테레오 신호로 다운믹스한 후 다섯 채널로 합성하는 5-2-5 시스템에서는 좌평면 벡터와 우평면 벡터의 전력을 스테레오 신호로부터 예측할 수 있으므로 네 개의 방위각만으로 모든 채널에 대한 스펙트럼 계수를 추정할 수 있다. 하지만 다섯 채널 신호를 모노 신호로 다운믹스한 후 다섯 채널로 합성하는 5-1-5 시스템에서는 모노 신호에 대한 전력만을 얻을 수 있으므로 네 개의

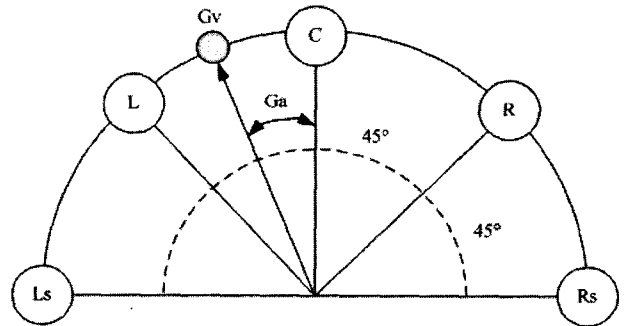


그림 7. 반원면에서의 음원 벡터 표현방법과 전체벡터의 각도^[30]
Fig. 7. Representation of source vectors on the semicircle plane with global vector angle (Ga).

방위각에 해당하는 전력을 얻기 위하여 네 개의 채널 간 전력 벡터의 합 벡터인 전체벡터(global vector; Gv)가 추가로 필요하다. 이에 반해 5-1-5 시스템에서의 ICLD 파라미터는 기준 채널과 나머지 채널간의 전력비를 표현하므로 네 개의 ICLD 파라미터만으로 모든 채널에 대한 이득을 예측할 수 있다. 따라서 5-1-5 시스템에서 VSLI는 ICLD에 비하여 멀티채널 오디오 신호를 위한 부가정보 크기가 증가하는 단점이 있다. 이러한 단점을 보완하기 위하여, 다섯 채널 스피커 배치를 그림 7과 같이 반원면 상으로 변형하고 채널간의 간격을 45도로 동일하게 함으로써 ICLD와 동일한 개수의 부가정보를 가지도록 하는 전체벡터 분할 방식 VSLI (global vector split VSLI; GS-VSLI) 파라미터가 소개되었다^[14].

GS-VSLI는 전체벡터의 위치에 따라 좌분면 벡터 각도 (LHa)와 우분면 벡터 각도(RHa)가 선택적으로 예측된다.

2. 음원위치단서 부호화 시스템

본 논문에서 제안하는 음원위치단서 부호화(SSLCC) 시스템은 VSLI와 ICC를 공간단서로 사용하는 저비트율 멀티채널 오디오 부호화 시스템이다. SSLCC 부호화기는 멀티채널 오디오 신호를 모노 또는 스테레오 신호로 다운믹스한 후 전력 동일화 과정을 거쳐 MPEG-4 AAC 또는 MPEG-4 BSAC과 같은 일반적인 지각 오디오 부호화기(perceptual audio coder)^[15]를 이용하여 압축하여 다운믹스 스트림을 생

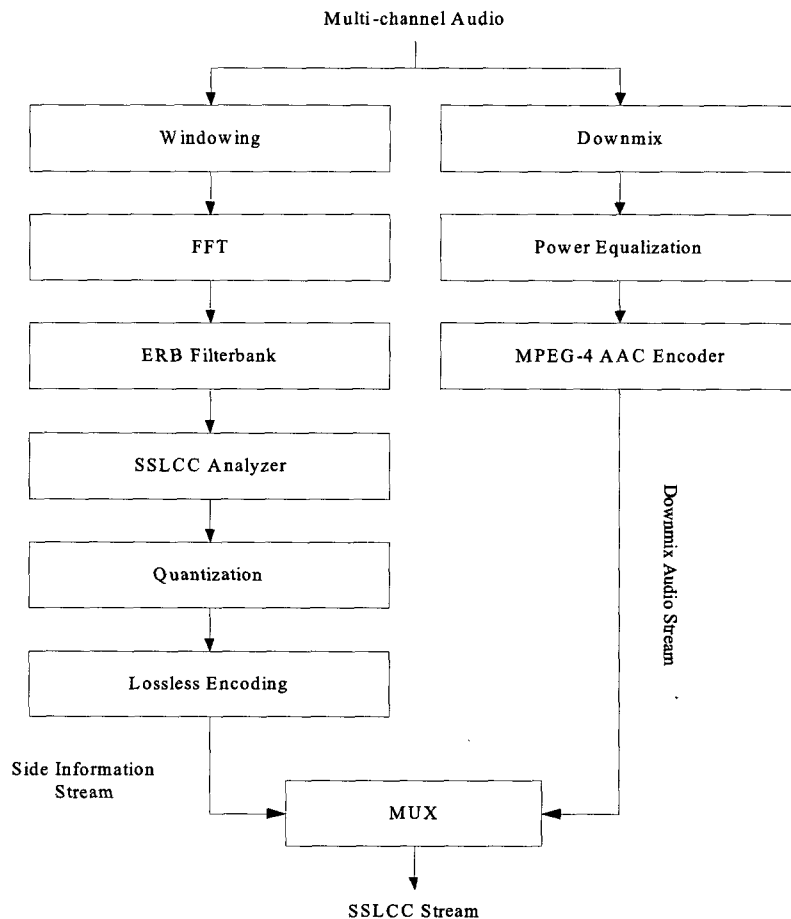


그림 8. SSLCC 부호화기 상세도

Fig. 8. The schematic diagram of SSLCC encoder

성한다. 또한 멀티채널 신호를 창 함수를 이용하여 프레임 단위로 분할하고 FFT 변환을 이용하여 주파수축 상에서 VSLI 파라미터와 ICC를 ERB 부대역별로 추출한 후 양자화 과정을 거쳐 부가정보 스트림을 생성한다. 생성된 다운믹스 스트림과 부가정보 스트림을 다중화 하여 SSLCC 스트림을 생성한다. 그림 8은 상기의 과정을 나타내는 블록도이다.

SSLCC 부호화기에서 적용되는 시간-주파수 변환 기법은 BCC 시스템과 동일한 FFT와 20대역 ERB 필터뱅크를 이용하였다. 본 논문에서 제안하는 SSLCC 부호화 시스템에서는 가상적인 반원면상에서 전체벡터를 분할하여 각도를 예측하는 GS-VSLI (global vector split VSLI) 파라미터를 공간 단서로 이용하여, 5-1-5 및 5-2-5 시스템 모두 동일한 개수의 GS-VSLI 파라미터를 예측한다.

SSLCC 해석기는 그림 9와 같이 주파수 대역으로 변환된 멀티채널 신호를 20대역 ERB 필터뱅크를 이용하여 대역별로 분할하고, 대역별 각 채널들의 전력을 계산한다. 계산된 대역별 채널의 전력은 삼각 창함수를 이용하여 평탄화 과정

을 거친 후 에너지 벡터를 예측하고 GS-VSLI 각도와 ICC 값을 부대역 별로 예측한다.

SSLCC 복호화기는 SSLCC 스트림을 다운믹스 오디오 스트림과 부가정보 스트림으로 분리한 후 다운믹스 오디오 스트림으로부터 다운믹스 신호의 스펙트럼을 추출하고 부가정보 스트림으로부터 획득된 GS-VSLI 파라미터와 ICC 파라미터를 이용하여 멀티채널 오디오 스펙트럼을 합성한다. 합성된 멀티채널 오디오 스펙트럼은 주파수-시간 변환을 통하여 시간 축으로 변환되어 멀티채널 오디오 신호를 생성한다. 다운믹스 오디오 스트림으로부터 MPEG-4 AAC 복호화기를 통해 해석된 모노 또는 스테레오 오디오 시스템과 호환성을 유지할 수 있다. 그림 10은 SSLCC 복호화기의 상세 블록도이다.

SSLCC 복호화기에서는 BCC 복호화기와 동일한 시간-주파수 변환 기법이 적용되며, ICLD를 대신하여 GS-VSLI 파라미터에 의해 멀티채널 오디오 스펙트럼에 대한 레벨 조절이 이루어지며, ICC 값에 의한 상관도 조절이 이루어진다.

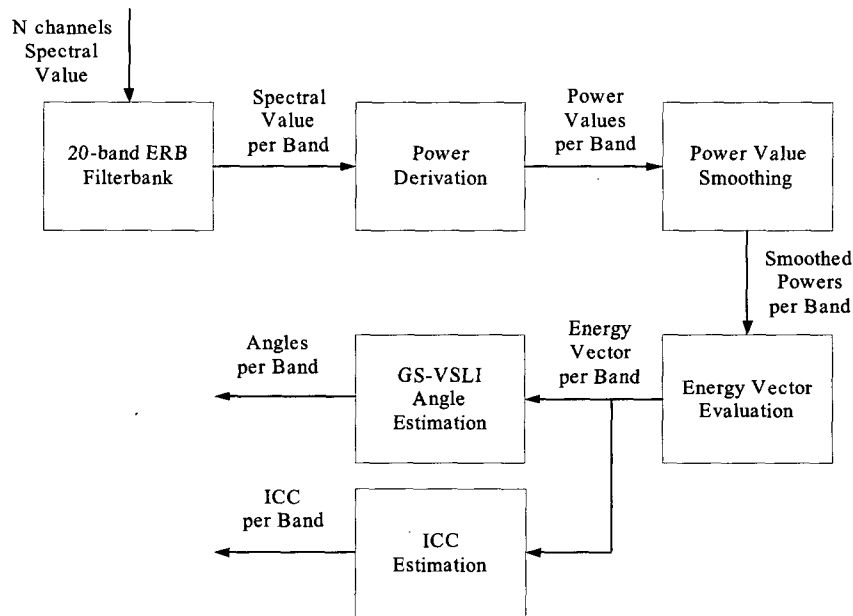


그림 9. SSLCC 해석기 구조
Fig. 9. Block diagram of SSLCC analyzer

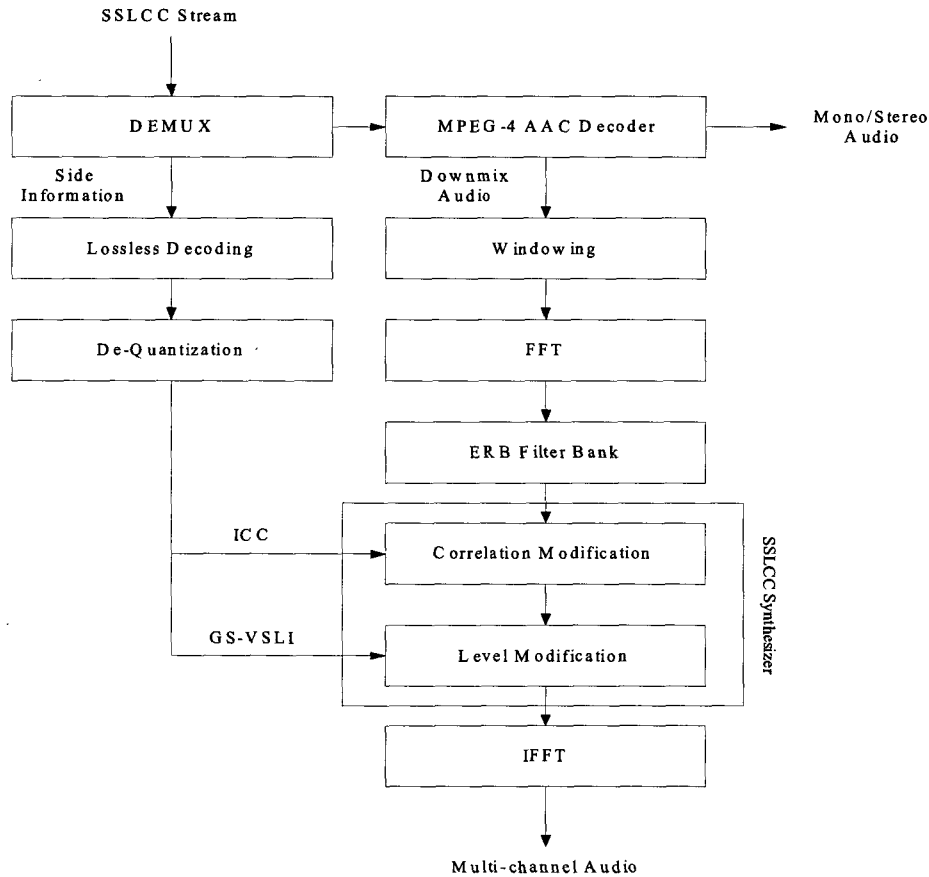


그림 10. SSLCC 복호화기 상세도
 Fig. 10. The schematic diagram of SSLCC decoder

IV. 실험 및 결과

제안한 SSLCC의 우수성을 입증하기 위하여 본 실험에서는 객관적/주관적 평가를 실시하였다. 우선 객관적 평가로써 기존의 ICLD를 이용한 BCC 시스템과 SSLCC 시스템에 동일한 양자화 기법과 허프만 무손실 압축 기법을 적용하여 평균 비트율을 비교하였고, 주관청취평가를 실시하여 제안한 SSLCC 시스템의 합성음질을 평가하였다. 실험에서는 MPEG 오디오 그룹에서 MPEG-4 SAC 기술의 성능을 평가하기 위해 사용된 11개의 멀티채널 데이터를 이용하였다¹⁶⁾. 표 1은 실험 데이터의 이름과 그 특징을 정리한 것이다.

1. 객관 평가

사용된 양자화 레벨에 따라 SSLCC와 BCC 시스템에 동일한 양자화 기법과 허프만 무손실 압축기법을 적용하였을 때 11개 실험 데이터에 대한 평균 비트율은 표 2와 같다. 제시된 비트율은 5-1-5 SSLCC 시스템의 11개 실험 데이터에 대한 평균 비트율과 5-1-5 BCC 시스템의 평균 비트율을 비교한 것이다. 표에서 보는 바와 같이, 모든 양자화 레벨에서 SSLCC가 BCC에 비해 약 1kbps 정도 비트율이 감소함을 확인하였다.

2. 주관 평가

2.1 실험 환경

청취평가는 서울대학교 음향공학연구소 청취룸과 한국 전자통신연구원 AV 시연룸에서 진행되었다. 두 평가 장소에서 사용된 멀티채널 오디오 재생시스템 및 개략적인 구성은 표 3과 그림 11과 같다. 제안된 SSLCC 기법의 음질을 평가하기 위해 ITU-R BS. 1543-1 MUSHRA (Multi-Stimulus test with Hidden Reference and Anchor) 방법을 이용하여 청취 평가를 수행하였다. MUSHRA는 MPEG-4 SAC 기술 성능 평가를 위해 선정된 음질평가 방법으로써 중간 정도의 음질을 가지는 오디오 시스템의 성능을 평가하기 위하여, 숨겨진 기준 신호와 저역필터로 음질을 열화시킨 anchor 신호들로 실험 시스템을 구성하여 주관적으로 음질을 평가하는 방법이다.

적절한 주관 청취평가 결과를 도출하기 위해 ISO 에서 제시한 실험 전 훈련과정^[7]을 통해 피실험자들이 실험 데이터에 익숙하도록 유도하였으며, 음질 판단을 위한 주요 인자들(합성음질의 명료도, 음상의 정위감 및 음장감)에 대한 사전교육도 수행하였다. 전체 20명에 대한 실험 결과 중에 일관성을 유지하지 못하거나 기준신호와 앵커신호를

표 1. 실험 데이터
Table 1. Test materials

Number	Index	Name	Category	Remarks
1	A	applause	pathological & ambience	
2	B	ARL applause	pathological & ambience	
3	C	chostakovitch	music(back: direct)	
4	D	fountain music	pathological & ambience	
5	E	Glock	pathological & ambience	
6	F	Indie2	movie sound	
7	G	Jackson1	Music(back: ambience)	
8	H	pops	music(back: direct)	
9	I	Poulenc	music(back: direct)	
10	J	rock concert	Music(back: ambience)	
11	K	Stomp	movie sound	LFE

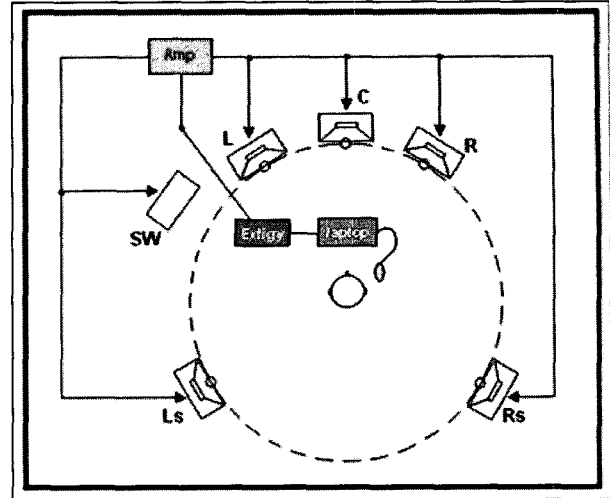


그림 11. 멀티채널 청취평가 시스템의 구성
Fig. 11. Configuration of the 5.1 channel listening test system.

표 2. 양자화 레벨(Q)에 따른 평균 비트율(kbps)
Table 2. Average bitrate (kbps) according to quantization level (Q).

Quantization Level (Q)	7	15	31	63
BCC (ICLD)	12.47	18.62	24.86	31.50
SSLCC (GS-VSLI)	11.85	17.67	23.83	29.45

구별하지 못한 실험자의 결과는 제외하여 총 10명의 실험 결과로 SSLCC 부호화기의 성능을 분석하였다.

2.2. 청취 실험 결과

스테레오로 다운믹스 된 오디오 신호로부터 다섯 채널 신호를 합성하는 5-2-5 SSLCC 시스템의 합성 음질에 대한 주관 청취평가를 위하여 표 4에서 나열한 네 개의 시스템이 사용되었다. 다운믹스 스테레오 신호는 MPEG-4 AAC를 이용하여 160kbps로 부호화하였다. 일반적으로 MPEG-4 AAC 스테레오 신호일 경우 96kbps 정도면 원음과 구분할 수 없는 음질을 내지만 다운믹스 오디오 신호를 부호화하는 과정에서 발생하는 음질의 열화가 제안한 시스템을 평가하는 데 영향을 끼치지 않게 하기 위하여 충분한 대역폭으로 다운믹스 오디오 신호를 압축하였다.

표 3. 멀티채널 오디오 재생 시스템의 구성

Table 3. The configuration of multi-channel audio playback system.

	서울대학교	한국전자통신연구원
Multi-Channel Amp.	YAMAHA DSP AZ1	DENON AVC-1
Loudspeaker	ELAC 330 JET x 5	ELAC 330 JET x 5
Subwoofer	YAMAHA YST-SW215	Velodyne HGS15
Multi-Channel Soundcard	SoundBlaster Extegy	SoundBlaster Extegy

표 4. 5-2-5 SSLCC 청취실험에 사용된 멀티채널 오디오 시스템들

Table 4. Multi-channel audio systems for 5-2-5 SSLCC listening test

Classification	Description
Hidden Reference	Original
Anchor	3.5 kHz band-limited
Additional Anchor	Software-based commercial Dolby ProLogic II encoder/decoder
5-2-5 SSLCC System	Stereo AAC codec (160 kbps) + SSLCC codec

표 5. 5-1-5 SSLCC 청취실험에 사용된 멀티채널 오디오 시스템들

Table 5. Multi-channel audio systems for 5-1-5 SSLCC listening test

Classification	Description
Hidden Reference	Original
Anchor	3.5 kHz band-limited
Additional Anchor	Software-based commercial Dolby ProLogic II encoder/decoder
5-1-5 SSLCC System	Stereo AAC codec (80 kbps) + SSLCC codec

그림 12는 MUSHRA 청취 평가 결과를 평균값과 95% 신뢰구간으로 나타낸 것이다. 실험 결과 5-2-5 SSLCC 시스템이 Dolby ProLogicII에 비해 음질이 두드러지게 우수함을 나타내었다. 그리고 11개 시험 데이터에 대한 음질 평가 평균 점수가 86점으로 측정되어 상당히 우수한 음질을 보임을 확인하였다.

모노로 다운믹스 된 오디오 신호로부터 다섯 채널 신호를 합성하는 5-1-5 SSLCC 시스템의 합성 음질에 대한 주관 청취 평가를 위하여 표 5에 나열한 네 개의 시스템이 사용되었다. 다운믹스 모노 신호는 MPEG-4 AAC를 이용하여

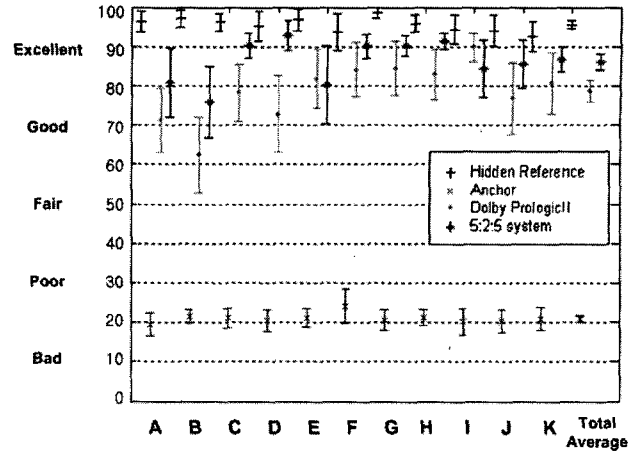


그림 12. 5-2-5 SSLCC 시스템 청취실험 결과

Fig. 12. Listening test results on 5-2-5 SSLCC System.

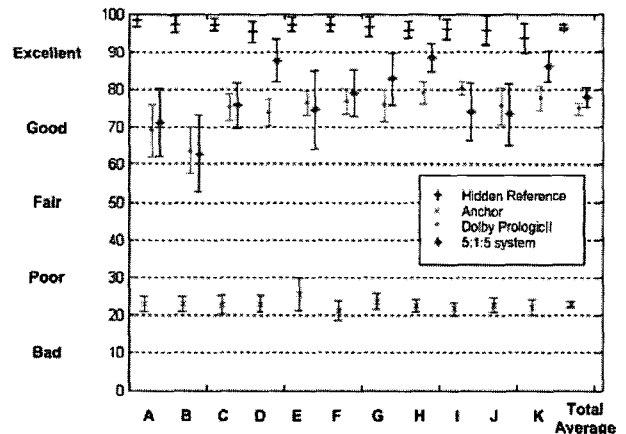


그림 13. 5-1-5 SSLCC 시스템 청취실험 결과

Fig. 13. Listening test results on 5-1-5 SSLCC System

80kbps로 부호화 하였고, 청취 평가 시 피실험자에게는 5-2-5 SSLCC 시스템 실험과 동일하게 실험데이터에 대한 일반적인 오디오 음질을 평가하도록 요구하였다. 그림 13은 청취평가결과를 평균값과 95% 신뢰구간으로 나타낸 것이다. 실험 결과 5-1-5 SSLCC 시스템이 5-2-5 시스템에 비해서는 음질의 열화가 느껴지지만 적절한 음질을 유지함을 확인하였고, Dolby ProLogicII 보다는 우수한 음질을 나타내었다.

V. 결론

오늘날 멀티채널 서비스의 필요성은 증가하고 있는 반면 신호전송을 위한 대역폭의 제약성 등으로 인해 멀티채널 오디오 콘텐츠를 효율적으로 전송하기 위한 압축 기술이 요구되고 있다. 본 논문에서는 협대역폭 채널환경에 적합한 고압축율의 멀티채널 오디오 부호화 기법인 음원위치단서 부호화(Sound Source Location Cue Coding; SSLCC) 기술을 제안하였다. SSLCC는 BCC와 동일한 구조를 이용하여 멀티채널 오디오 신호를 다음믹스 오디오 신호와 공간단서로 구성되는 부가정보로 분리하여 압축한다. 또한, 공간단서 파라미터로서 VSLI와 ICC를 이용하며, 대칭형 균일 양자화기와 허프만 부호화기를 이용하여 멀티채널 오디오 신호를 저비트율로 압축한다. 제안된 SSLCC 기법의 성능을 검증하기 위한 객관평가를 수행한 결과/주관 평가를 수행하였다. 객관평가로써 BCC 기법과 SSLCC 기법에 동일한 허프만 부호화기를 적용했을 때의 평균 비트율을 비교한 결과, SSLCC가 BCC에 비해 비트율이 약 1kbps 정도 감소하였다. 그리고 주관평가 결과 5-2-5 SSLCC 시스템 및 5-1-5 SSLCC 시스템 모두 상용되고 있는 Dolby ProLogic II에 비해 월등히 우수한 음질을 보였다.

본 연구를 바탕으로 앞으로는 멀티채널 오디오 콘텐츠의 효율적 전송을 위한 부호화 기술에 추가적으로 멀티채널 재생환경을 최대한 활용하여 청취자에게 고품질의 3차원 음향을 재현할 수 있는 기술을 개발 할 계획이다.

참 고 문 헌

- [1] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N6455, "Call for Proposals on Spatial Audio Coding," Munich, Mar. 2004.
- [2] C. Faller and F. Baumgarte, "Efficient Representation of Spatial Audio using Perceptual Parameterization," IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust., Oct. 2001.
- [3] C. Faller and F. Baumgarte, "Binaural Cue Coding: A Novel and Efficient Representation of Spatial Audio," Proceeding ICASSP 2002, Orlando, FL, May 2002.
- [4] F. Baumgarte and C. Faller, "Estimation of Auditory Spatial Cues for Binaural Cue Coding (BCC)," Proceeding ICASSP 2002, Orlando, FL, May 2002.
- [5] ISO/IEC JTC1/SC29/WG11, Document M11264, "A New Cue Parameter for Spatial Audio Coding," Redmond, July 2004.
- [6] 문한길, 이정석, 서정일, 성광모, "멀티채널 오디오 재생 시스템에서 가상 음원의 위치 정보를 이용한 압축 재생 방법," 2004년도 한국음향학회 추계학술발표대회, 제23권 제2(s)호, 2004.
- [7] 문한길, "가상음원 정보를 이용한 MPEG-4 SAC 다채널 오디오 압축 기법," 서울대학교박사학위논문, 2005년 2월.
- [8] S. Beack, J. Seo, H. Moon, and M. Hahn, "Angle-based Virtual Source Location Representation for Spatial Audio Coding," submitted to IEE Letter.
- [9] H. Moon, J. Seo, S. Beack, and K. Sung, "A Multi-Channel Audio Compression Method with Virtual Source Location Information for MPEG-4 SAC," submitted to IEEE Trans. on Consumer Electronics.
- [10] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources I: Three-dimensional Panning," Journal Audio Engineering Society, vol. 49, no. 9, pp. 753-767, September 2001.
- [11] C. Faller and F. Baumgarte, "Binaural Cue Coding Applied to Audio Compression with Flexible Rendering," Proceeding AES 113th Convention, Los Angeles, CA, Oct. 2002.
- [12] C. Faller and F. Baumgarte, "Binaural Cue Coding Part II: Schemes and Applications," IEEE Trans. Speech Audio Processing, vol. 11, pp. 520-531, Nov. 2003.
- [13] ITU-R Recommendation, "Multichannel Stereophonic Sound System with and Without Accompanying Picture," International Telecommunication Union, BS. 775-1, Geneva, 1994.
- [14] S. Beack, J. Seo, H. Moon, and M. Hahn, "Angle-based Virtual Source Location Representation for Spatial Audio Coding," submitted to IEE Letter.
- [15] ISO/IEC 14496-3:2001, "Information Technology-Generic Coding of Audio-Visual Objects, Part 3: Audio," International Standard Organization, 2001.
- [16] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N6691, "Procedures for the Evaluation of Spatial Audio Coding Systems," Redmond, July 2004.
- [17] G. A. Souldre, and M. C. Lavoie, "Subjective Evaluation of Large and Small Impairments in Audio Codecs," AES 17th International Conference, Florence, pp. 329-336, Sep. 1999.

— 저 자 소 개 —



장 인 선

- 2001년 2월 : 충북대학교 전기전자공학부 졸업 (학사)
- 2004년 2월 : 포항공과대학교 대학원 컴퓨터공학과 졸업 (석사)
- 2004년 8월 ~ 현재 : 한국전자통신연구원 방송미디어연구그룹 연구원
- 주관심분야 : 오디오 신호처리, 음원분리, MPEG, DMB



서 정 일

- 1994년 2월 : 경북대학교 전자공학과 졸업 (학사)
- 1996년 2월 : 경북대학교 대학원 전자공학과 졸업 (석사)
- 2005년 8월 : 경북대학교 대학원 전자공학과 졸업 (박사)
- 1988년 3월 ~ 1999년 5월 : LG 반도체 근무
- 1999년 6월 ~ 2000년 10월 : 현대 전자 근무
- 2000년 11월 ~ 현재 : 한국전자통신연구원 방송미디어연구그룹 선임연구원
- 주관심분야 : 음향신호처리, 음성신호처리, 객체기반 3차원 오디오 신호처리, MPEG, DMB



백 승 권

- 1999년 2월 : 한국항공대학교 항공전자공학과 졸업 (학사)
- 2001년 2월 : 한국정보통신대학교 대학원 전자공학과 졸업 (석사)
- 2005년 8월 : 한국정보통신대학교 대학원 전자공학과 졸업 (박사)
- 2005년 8월 ~ 현재 : 한국전자통신연구원 방송미디어연구그룹 선임연구원
- 주관심분야 : 오디오 신호처리, 음원분리, MPEG, DMB



강 경 옥

- 1985년 2월 : 부산대학교 물리학과 졸업 (학사)
- 1988년 2월 : 부산대학교 대학원 물리학과 졸업 (석사)
- 2004년 2월 : 한국항공대학교 대학원 항공전자공학과 졸업 (박사)
- 1991년 2월 ~ 현재 : 한국전자통신연구원 방송미디어연구그룹 3D미디어연구팀장, 책임연구원
- 주관심분야 : 오디오 신호처리, MPEG-7 및 TV-Anytime 기반 맞춤형 방송