

Spatial Audio Coding 기술의 멀티채널 부호화 성능 비교

장인선*, 서정일*, 문한길**, 강경옥*

*한국전자통신연구원, **서울대학교

Evaluation of Spatial Audio Coding Tools for Multichannel Audio

Inseon Jang*, Jeongil*, Hangil Mun**, Kyeongok Kang*

*ETRI, **SNU

*{jinsn, seoji, kokang}@etri.re.kr

**fullmoon@acoustics.snu.ac.kr

요약

Spatial Audio Coding (SAC)은 낮은 대역폭에서 다채널/다객체 오디오 신호를 전송하기 위해 제안된 기술이다. 본 논문에서는 MPEG 에서 SAC 기술의 평가 방법으로 채택된 Multi-Stimulus test with Hidden Reference and Anchor (MUSHRA) 실험 절차에 대해서 설명한다. 또한 제 69 차 MPEG 회의에서 제안된 4 개 기관의 SAC 기술에 대한 청취실험을 수행하고 그 결과를 분석한다.

1. 서론

낮은 대역폭에서 다채널 다객체 오디오 신호를 전송하기 위해 새롭게 제안된 Spatial Audio Coding (SAC) 기술은 다음과 같은 장점을 가진다. [3] 첫째로 SAC 기술은 채널당 4~8kbps 정도의 부정보 (side information)로 멀티채널 오디오 신호를 표현하기 때문에, 멀티채널 오디오 신호의 압축율을 극대화 할

수 있다. 둘째로, SAC 기술은 모노 또는 스테레오 디코더에 대한 역행 호환성(backward compatibility)을 제공하기 때문에 모노 또는 스테레오 디코더는 멀티채널 부정보 채널을 제외한 다운믹스 오디오 신호를 재생할 수 있다.

SAC 기술의 평가 방법과 관련하여 제 69 차 MPEG 회의에서 Multi-Stimulus test with Hidden Reference and Anchor (MUSHRA) [1, 4] 주관 청취평가 채택되었다. 이 방법에 기반하여 각 실험 기관에서는 현재 MPEG 에 제안된 4 개 기관의 SAC 기술을 평가해야 한다.

본 논문의 전체 구성은 다음과 같다. 2 장에서는 MUSHRA 실험의 전체 개요와 방법에 대해서 설명하고, 3 장에서는 한국전자통신연구원과 서울대학교에서 실시한 청취실험의 환경 및 방법에 대해서 설명하고 그 결과를 분석한다.

2. Multi-Stimulus test with Hidden Reference and Anchor

2.1 피험자의 선별

피험자의 선별은 신뢰할 수 있는 결과를 얻기 위해서 매우 중요한 단계이며, 실제 청취 실험 전 그리고/또는 후에 이루어진다. 청취 실험 이전의 피험자 선별 기준은 피험자가 정밀하고 감식력 있게 소리를 들은 경험이 있는지 여부이다. 청취 실험 이후의 피험자 선별은 두 가지 기준을 가지고 이루어지는데, 첫 번째 기준은 피험자가 반복되는 실험에서 일관적인 채점을 하는지 여부이며, 두 번째 기준은 피험자가 채점한 결과가 전체 피험자들이 채점한 실험 결과(평균 점수)와 일치하는지 여부이다. 이러한 기준에 맞는 피험자를 선별하면 의미 있는 실험결과를 얻을 수 있다. 그리고 전체 실험의 관리를 위해서 20 명 이하의 피험자가 적당하다.

2.2 실험 방법

2.2.1 Description of test signals

피험자들의 청취 실험 효율과 청취 실험의 전체 시간(약 30 분)을 고려해서 각 실험 신호는 20 초가 넘지 않게 한다. 각 실험에는 참조 신호(reference), 숨겨진 참조 신호 (hidden reference), 각 기관에서 재출한 신호들 그리고 숨겨진 앵커(hidden anchor)를 실험 신호로 사용한다. 이때 사용되는 숨겨진 참조신호와 숨겨진 앵커는 피험자들의 청취능력을 평가하는데 사용된다.

2.2.2 Training Phase

피험자들은 실제 청취 실험에 들어가기 전에 훈련 단계를 거치게 된다. 훈련 단계의 목적은 첫째, 피험자들이 실험 신호들에 익숙해지게 함이다. 훈련 단계에서 피험자들은 실험실 내부를 자유롭게 돌아다니며 반향(reverberation) 및 특정소리의 위치 등에 익숙해 질 수 있다. 둘째, 피험자들이 실험 신호를

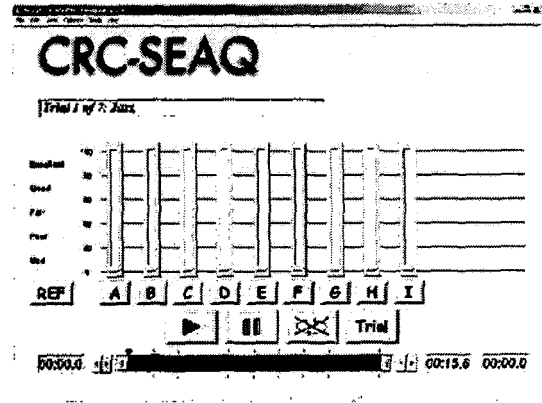


그림 1. User interface used in the blind grading phase

채점하기 위해 이용되는 사용자 인터페이스 틀과 청취 실험 장비를 사용하는 방법을 배우는데 있다.

2.2.3 Grading Phase

MUSHRA 방법에 따라 피험자들은 실험 신호들의 기본적인 오디오 품질(Basic Audio Quality)을 연속적인 음질 등급(Continuous Quality Scale: CQS)에 따라 채점한다. 그림 1 은 MUSHRA 를 위한 사용자 인터페이스 틀이다. 인터페이스 틀을 이용하여 피험자들은 참조신호와 다양한 실험 신호들을 횡수나 순서 면에서 마음대로 청취할 수 있고 청취 후 슬라이더를 움직여서 각각의 실험 신호들에 대한 채점을 할 수 있다.

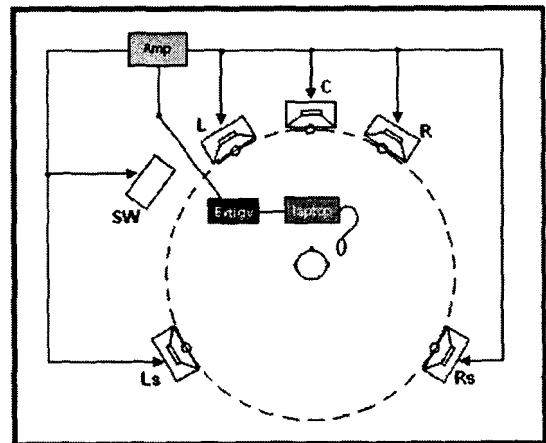


그림 2. Configuration for Listening Room

표 1. test equipments

Equipment	Model
5 Full-range loudspeakers	ELAC 330 JET
Subwoofer	YAMAHA YST-SW315
Amplifier	YAMAHA DSP-AX1
Sound card	Sound Blaster Extigy

3. 청취 실험

3.1 청취 실험 환경

전체적인 청취 실험 환경은 그림 2 와 같다. 청취 실험은 서울대학교 뉴미디어 연구소에 있는 무향실에서 실시되었고, 실험에 사용된 실험 장비의 모델명은 표 1 과 같고 무향실의 환경은 표 2 와 같다.

3.2 청취 실험

전체적으로 두 가지 실험을 실시했는데, 기존의 1)모노, 2)스테레오 오디오 시스템과 호환성을 유지하면서 고품질 멀티 채널 어플리케이션을 위한 성능을 측정하기 위한 실험이었다. 그리고 실험 신호들은 MPEG Audio 그룹에서 제공한 실험 신호들을 사용하였다. 이 실험 신호들에는 숨겨진 참조 신호와 3.5kHz 저주파 통과된 숨겨진 앵커 신호, Dolby Prologic II 를 통과한 앵커 신호, 그리고 4 개의 SAC 시스템을 통과한 신호들이 실험의 공정성을 위하여 실험 별 실험 신호 종류별로 암목적 무작위로 섞여 있었다.

청취 실험에 참여한 피험자의 수는 총 43 명이었고, 각 실험 시 소요되는 시간은 피험자 한 명당 약 30 분이였다. 실험에 사용된 MUSHRA 소프트웨어는 ARL 에서 제공한 Subjective Test Evaluation Program (STEP) v0.10 을 사용하였다.

실험 후, 신뢰할 만한 실험 결과를 선별하여 총 10 명분에 해당하는 결과를 MPEG 에 보고하였다. MPEG 에서는 각 실험 기관에서 보고한 실험 결과를 취합하여 표 3, 4 와 같이 발표 하였다. [2]

청취 실험 결과는 제안된 4 개의 SAC 기술 중 CTP 와 FhG/Agere 의 것이 우수하였다. 이 중 최우수 시스템을 선정하기 위하여 평가 기준을 적용한 결과, spatial coding side information 의 비율이 상대적으로 약 30% 적은 FhG/Agere 의 SAC 기술이 RMO 로 선정 되었다.

4. 결론

본 논문에서는 제한된 대역폭에서 다채널/다객체 오디오 신호를 효과적으로 압축하기 위해 제안된 Spatial Audio Coding (SAC) 기술의 평가 방법인 MUSHRA 주관 청취평가 방법에 대해서 설명하였다. 한국전자통신연구원과 서울대학교는 제안된 4 개 기관의 SAC 기술을 MUSHRA 주관 청취평가 방법을 이용하여 각 기술 결과물의 기본적인 오디오 음질을 연속적인 음질 등급에 따라 채점하였다.

감사의 글

본 논문은 정보통신부의 연구사업인 “지능형 통합 정보 방송 기술 개발”과제의 일환으로 수행한 결과로서 정보통신부 담당자 및 관련 연구원들의 노력에 감사를 드립니다.

참고문헌

1. ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate sound quality (MUSHRA)", International Telecommunications Union, Geneva, Switzerland, 2001.
2. M.11199, 70th MPEG meeting, "Draft report on spatial audio coding RMO selection tests."
3. N.6455, 68th MPEG meeting, "Call for proposals on spatial audio coding."
4. N.6691, 69th MPEG meeting, "Procedures for the evaluation of spatial audio coding systems."

II 2. Acoustic reference environment

	Reference	Studio
Floor area (m ²)	50±20	23.9
Room volume (m ³)	Over 80	55
Reverberation time (sec)	0.2±0.05	0.2~0.3
Static transfer frequency response	125Hz~4kHz: ±3dB	125Hz~4kHz: ±6dB
Diffusion	Over 50%	Over 60%
Noise criteria	NC 15	NC 25

II 3. Test results of 5:1:5

Multichannel listening test		Key	Upper	Lower	Mean
Systems under test	Coding Technologies/Philips	ctp	71.22	67.79	69.50
	Dolby	dol	70.18	66.60	68.39
	FhG/Agere	fng	73.95	70.79	72.37
	Panasonic	pan	62.50	58.64	60.57
Reference and anchor systems	Hidden reference	ref	98.66	98.00	98.33
	Dolby ProLogic II	Proll	65.33	61.90	63.62
	3.5 kHz lowpass filtered test items	lp35	17.43	15.97	16.70

II 4. Test results of 5:2:5

Multichannel listening test		Key	Upper	Lower	Mean
Systems under test	Coding Technologies/Philips	ctp	81.78	79.58	80.68
	Dolby	dol	76.29	73.86	75.07
	FhG/Agere	fng	76.93	74.55	75.74
	Panasonic	pan	72.69	70.10	71.39
Reference and anchor systems	Hidden reference	ref	98.61	97.95	98.28
	Dolby ProLogic II	Proll	57.25	54.55	55.90
	3.5 kHz lowpass filtered test items	lp35	15.46	14.05	14.76