

사용자 정보를 반영한 객체 기반 오디오 재생 기술

문재원, 정종진, 김경원, 임태범, 이석필

전자부품연구원

jwmoon@keti.re.kr, mazzalt@keti.re.kr, kwkim@keti.re.kr, tblim@keti.re.kr, lspbio@keti.re.kr

Object-based Audio Player using a User Information

Jaewon Moon, Jongjin Jung Kim, Kyungwon Kim, Taebeom Lim, Seokpil Lee

Korea Electronics Technology Institute

요 약

멀티미디어 서비스는 신호 처리 기술의 발달 및 전송 환경의 개선으로 정보 전달의 기존 역할 뿐 아니라 사용자의 다양한 요구 및 재생 환경을 반영하는 맞춤형 서비스로 진화하고 있다. 본 논문에서는 사용자의 다양한 청취 환경 변화, 선호도 및 감정을 네트워크상 입력 디바이스를 통해 전송하고, 이를 바탕으로 처리한 객체 기반 음원을 다채널 스피커를 통해 출력하는 능동형 재생 플랫폼을 제안한다. 다수의 청취자는 각각의 감성 및 환경 정보 등 음원 처리에 관련 데이터베이스를 실시간 저장하여 동일한 음원으로 단일 플랫폼에서 다양한 청취감으로 음원 재생이 가능하다.

1. 서 론

멀티미디어 서비스는 신호 처리 기술의 발달 및 전송 환경의 개선으로 정보 전달이라는 기존 역할을 수행할뿐만 아니라 사용자의 다양한 요구 및 재생 환경을 반영한 맞춤형 서비스로 진화 하고 있다. 또한 각종 IT 기기가 고도화 되고 콘텐츠 저장 공간의 제약은 줄어들어 고품질 멀티미디어 콘텐츠의 생성 및 재생이 가능하게 변화하였다. 이와 같은 트렌드를 반영하여 멀티미디어 가전 시장에서는 콘텐츠를 수동적으로 재생하는 제품이 아닌 인간의 감성 및 이동성을 모두 고려한 멀티미디어 서비스 제품이 경쟁력을 가질 것으로 예상된다. 이에 따라 IT 기술 뿐만 아니라 네트워크 및 감성 공학 기술을 고려한 멀티미디어 기반 요소 기술 연구가 필요하다. 고품질 콘텐츠 기반 실감형 멀티미디어 서비스와 관련하여 오디오 기술 분야에서는 고품질 무손실 오디오 저작(authoring) 기술 및 객체 기반 오디오 처리 기술, 사용자의 감성 및 환경 정보를 반영한 실감 오디오 전송 및 표현 기술 등에 관한 연구가 진행되고 있다.

기존 2 채널 기반 스트레오 음원 콘텐츠는 저용량으로 음원 저장이 가능하고 음장의 공간적 특성을 쉽게 재생할 수 있지만 멀티 채널 스피커를 지원하는 음원 소스로는 적합하지 않다. 콘텐츠 저장 공간의 제약이 줄어들든 현재 무손실 다채널 오디오 저작 기술은 왜곡 없이 고품질 콘텐츠를 재생 함으로써 풍부한 음향의 오디오 청취를 가능하게 한다.

객체 기반 오디오 처리 기술은 하나의 음원을 구성하는 다양한 종류의 객체 오디오 소스를 각각 제어 가능하게 하여 객체 기반 오디오 콘텐츠 생성 및 저작,

수정을 가능하게 하는 기술이다. 초기에 음원을 생성할 때는 독립적인 오디오 소스를 저작하기 보다 여러 개의 악기로 만든 오디오 객체를 생성하고 이를 정합(mixing)하여 사용하는 경우가 대부분이다. 음원 저작 특성을 이용하여 오디오 객체를 기반으로 제어 및 수정하는 기술은 제작자와 청취자가 원하는 방향으로 능동형 오디오 서비스를 가능하게 한다. 현재 사용자 능동형 오디오 서비스는 ETRI(한국 전자 통신 연구원)이 2008년 세계 최초로 개발하였으며[1] 이용자가 각각의 음원 객체를 제어해 가수와 악기의 소리를 원하는 대로 개별 선택해 감상할 수 있다. 국내에서는 객체기반 인터랙티브 뮤직(interactive music) 기술이 적용된 MUSIC 2.0 음반이 상용화 되고 있다.

심리 음향학적 특성을 반영하여 청취자의 위치 및 방향, 스피커의 공간 특성 및 채널 수 등 청취자와 연관된 다양한 환경 정보를 획득하고, 이를 기반으로 다채널 사운드 스트림의 채널을 분배하고 음원을 정위하여 청취자에게 최적의 사운드 서비스를 제공하는 기술 또한 연구되고 있다.

본 논문에서는 사용자의 능동적인 감정 및 환경 변화 입력 처리가 가능한 통합 플랫폼을 다양한 네트워크 디바이스에 적용하여 객체 기반의 독립 음원을 처리하는 고품질 능동형 음원 서비스를 제안한다. 서비스 통합 플랫폼은 사용자의 입력 정보를 데이터 베이스에 저장하고 재생용 스피커의 다채널 사운드 스트림 채널 분배를 제어하여 동일한 음원으로 다양한 청취 환경 변화를 가능하게 한다. 또한 다수의 사용자를 고려하여 사용자의 기본 프로파일에 각각의 감성 정보 및 음원 처리 관련 정보를 저장하여 음원 처리시 활용



[그림 1] 뮤직 2.0 플레이어 재생 화면

및 저장 가능하도록 한다.

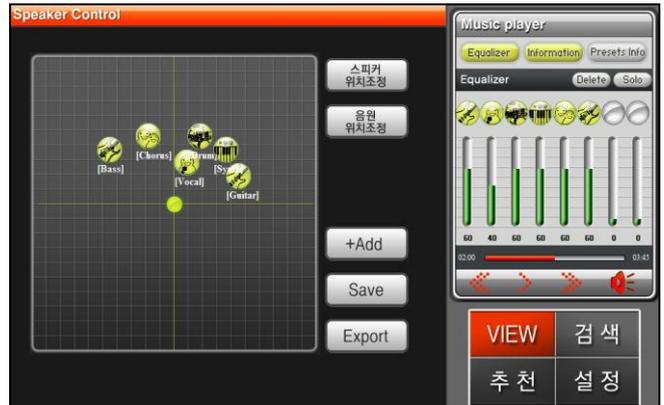
본 논문의 구성은 다음과 같다. 2장에서 기존의 객체 지향 멀티 트랙 오디오 콘텐츠 저작 및 재생 기술을 소개하고 3장에서 객체 기반 콘텐츠를 청취자의 환경과 선호도 및 감성 정보에 적응적으로 독립적인 음원 트랙을 보정하는 기술을 제안한다. 4장에서는 각각의 사용자가 자신의 디바이스 환경에 맞게 통합적으로 제어 가능한 객체 기반의 음원 재생 입출력 인터페이스를 제안한다. 마지막으로 5장에서 결론을 맺는다.

2. 기존 객체 음원 제어 기술

현재 오디오 음원은 보컬을 비롯한 악기 등 여러 개의 개별 음원에 독립적인 오디오 소스를 사용하기 보다는 제작자가 하나의 오디오 음원으로 정합하여 제공하며 사용자가 수동적으로 청취하는 포맷이 대부분이다. 객체 기반 오디오 처리 기술은 하나의 음원을 구성하는 다양한 객체 오디오 소스를 각각 제어 가능하게 하여 객체 기반 오디오 콘텐츠 생성 및 재생을 가능하게 하는 기술이다. 이는 제작자와 청취자가 원하는 오디오 객체를 능동적으로 제어 가능하게 하고 별도로 음원 세트를 조합 가능하므로 재생산이 가능하다. Music 2.0은 MPEG-A 멀티미디어 응용 포맷 분야(Multimedia Application Format) 국제표준을 주도한 객체 지향 멀티미디어 음원 기술로서, 세계 최초로 제품화 및 서비스 구현을 실현하였다. 그림[1]은 Music 2.0 플레이어 재생 화면으로 보컬과 여러 악기소리가 각자의 독립된 트랙을 구성하여 한 곡을 이루고 있으며, 플레이어를 통해 각 음악에 포함된 보컬, 피아노, 기타, 바이올린, 드럼 등 각 악기 음의 객체를 선택하여 조합 재생이 가능하다.

3. 사용자 환경을 반영한 음원 보정 시스템

3장에서는 오디오 콘텐츠의 객체 음원을 재생기에서 제어, 조합하여 재생하는 기능에서 나아가 사용자의



[그림 2] 음원 객체 조절 입력 장치

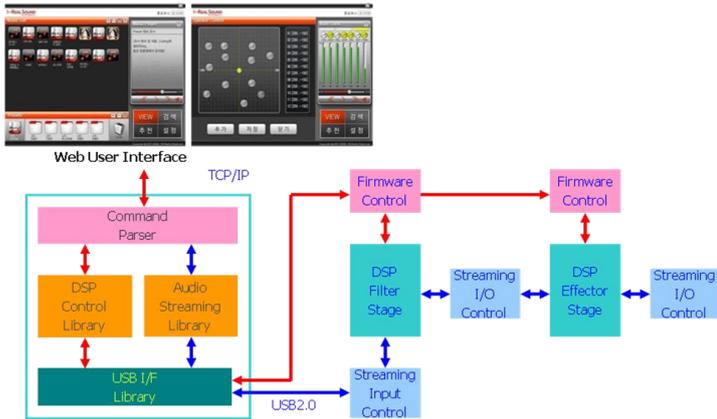
기본 청취 환경 및 기존 장면 묘사 정보, 음상 정위 정보를 이용하여 사용자의 다채널 스피커 환경에 맞게 조절 청취하는 시스템을 제안한다. 또한 객체 기반 음원 트랙 각각에 청취자의 선호도와 감성 정보를 반영하고 2차원 및 3차원 공간 파라미터와 실감 다채널 오디오 파라미터를 사용하여 단일 음원 소스에 다양한 효과를 적용한다. 마지막으로 트랙별 음원과 개별 스피커의 정보를 사용자의 입력에 맞게 조절 가능 하는 시스템을 소개한다. 위와 같은 사용자의 선호 및 환경 정보는 제어 파라미터를 새로운 프리셋 그룹으로 지정하고 원하는 시기에 언제든지 재생 가능하다.

3.1 사용자 기본 환경 반영

객체 기반 음원을 사용자 선호도에 기반하여 재생 하기 위하여 먼저 사용자의 기본 설정 환경을 반영 가능하도록 설계하였다. 먼저 사용자의 기본 가입 정보를 정의하고 관련 메타 데이터의 스키마를 설계하였다. 또한 사용자가 음원 객체를 제어 하는 세부 파라미터 값과 개별 음원에 따른 맞춤 재생 방식을 실시간으로 데이터 베이스에 저장하여 입출력 디바이스가 언제든지 접근 가능하게 하였다. 이렇게 저장된 데이터는 개인 청취 선호 패턴 분석 엔진에 사용하고 사용자가 히스토리에 근거한 음원 추천 결과를 요구 할 경우 사용자의 기본 정보 및 객체를 제어하는 기존 히스토리 데이터베이스에 근거하여 개인 맞춤형 음원 추천 결과 출력이 가능하도록 설계하였다.

3.2 음원 및 스피커 객체 조절 반영

사용자가 음원 파일을 선택 재생 할 경우 플레이어에서는 Scene description[2][3]을 이용해 음원의 종류 및 재생 관련한 기본 정보를 제공한다. 이에 기반하여 객체 기반 음원의 경우 세부 개별 트랙 정보 확인 및 개인 맞춤 제어가 가능하므로 더욱 풍부하고 실감 있는 음원 재생이 가능하다. 또한 다채널 출력 스피커의 위치를 포함한 기본 설정 정보를 알고



[그림 3] 객체 기반 음원 신호 처리 모듈

있는 경우 이에 맞추어 음원 스트리밍 출력을 조절하도록 설계 구현하였다.

그림[2]는 선택한 음원을 구성하는 개별 객체 종류와 볼륨 크기, 출력 스피커를 기준으로 한 스트리밍 재생을 제어하는 입력 모듈이다. 사용자는 각 음원 객체의 위치와 음원 크기를 직접 제어 가능하다. 각 조절 파라미터 값은 다채널 스피커 공간 정보와 함께 음원 처리 DSP로 전달 되고 기존 스피커 위치 메타 데이터와 결합하여 다채널 사운드 스트림 채널 분배를 제어한다. 음원 객체를 제어하는 동시에 실시간으로 음원 분배 스트리머에 반영하여 사용자 청취 환경의 다양한 변화가 가능하다.

객체 음원 조절과 유사하게 다채널 출력 스피터를 기준으로 사용자의 음원 제어가 가능하다. 기존에 구축된 스피커의 위치와 채널 분배 방식을 기준으로 사용자가 특정 스피커 위치의 음원 스트리밍 방향을 임의로 변경 할 경우 가상으로 개별 스피커의 위치를 조절하거나 제거할 수 있다. 이는 물리적으로 변경이

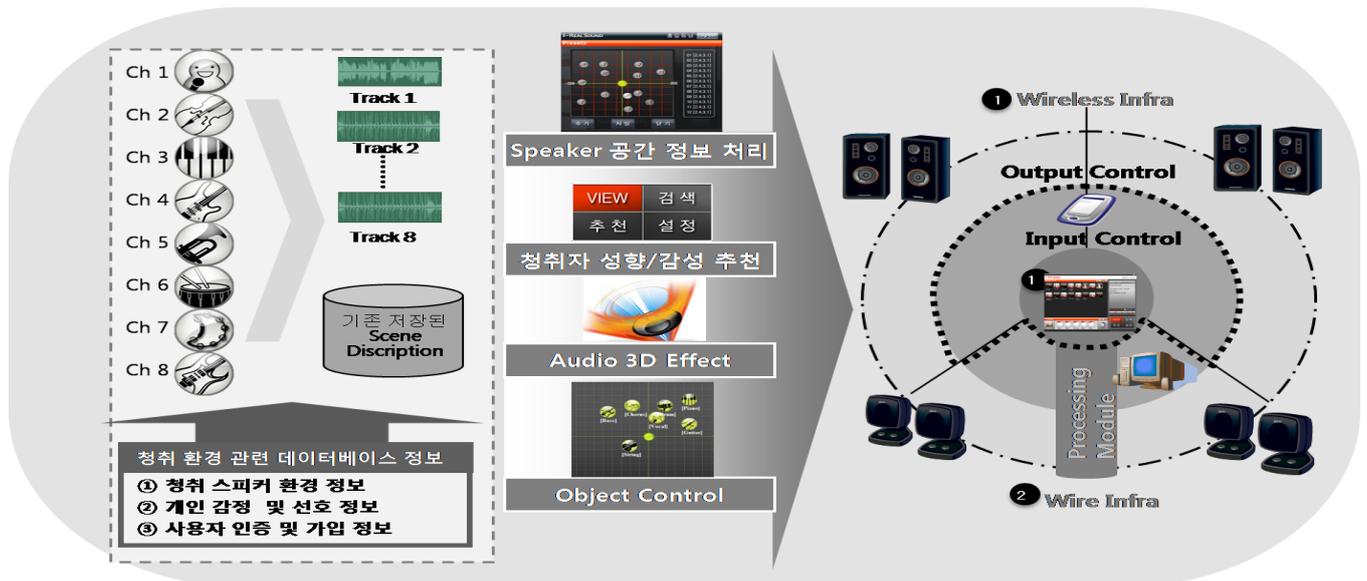
어려운 스피커의 위치 및 성능을 단순 조작만으로 변경 청취 시뮬레이션이 가능하게 한다.

음원 객체 위치 및 출력 스피커에 관련한 변경 요구 사항은 스피커 위치 좌표에 관한 움직임 벡터 값을 계산 하여 플랫폼의 음원 처리 모듈로 전송된다. 이를 반영하여 실시간으로 재생 가능하며, 사용자가 청취 중 선호하는 스피커 및 음원 제어 셋을 저장하여 언제든지 다음 재생시 같은 음향감으로 청취 가능하다.

3.3 감성 및 3D 환경 선호 정보 반영

사용자의 감성 및 선호 정보를 객체 음원에 반영하고 이를 저장 및 재생 가능하도록 설계하였다. 플랫폼에서 사용하는 객체 기반 음원은 오리지널 감성 정보를 갖고 있으며 이는 Scene description 에 정의되어 있다. 사용자는 감성 데이터 정보를 수집하기 위해 미리 감성 상태를 저장해 놓는다. 만약 사용자의 감성이 변할 경우 실시간으로 자신의 감성 정보를 변경하여 반영할 수 있다. 음원 처리 모듈은 입력된 감성 파라미터 값을 바탕으로 음원을 후처리 하여 재생에 반영한다.

인간의 청각 기관은 3차원 환경에서 전달되는 소리의 위치와 방향 크기를 인지한다. 이에 착안하여 다채널 환경을 고려한 3D 실감 오디오 재생 기술이 연구 되고 있다. 제안하는 시스템에서는 10.2 채널 환경을 고려하여 source perception, Source room interaction, late decay 관련한 9개의 Perceptual Parameter를 정의하고 사용자가 제어하는 파라미터 값을 재생에 적용 가능하도록 하였다. Source perception 관련 하여 Source presence, warmth, brilliance 파라미터, Source room interaction 관련 재생 공간 배치, 잔향 효과, envelopment 파라미터를 사용하며, 후 감쇠 효과 관련 하여 후잔향감, 중량감, 실감 파라미터를 사용하였다.



[그림 4] 객체 기반 음원 제어 통합 입출력 장치

또한 3차원 재생 공간셋을 미리 정의하고 선택 가능하게 하여 사용자가 특정 공간 가상 재생을 선택할 경우 미리 정의된 3D 파라미터 값으로 음원 출력을 변경하고 12채널 TDM 방식으로 출력하여 공간감을 주도록 설계하였다.

4. 객체 기반 음원 제어 통합 입출력 장치

제안하는 객체 지향 음원 재생 및 저작 기술을 적용하기 위해서는 사용자가 자신의 디바이스 환경에 맞게 통합 제어 가능한 객체 기반의 음원 재생 입출력 인터페이스가 필요하다. 이는 오디오를 단순히 선택 재생 하는 일반 오디오 기기와는 달리 사용자의 감성 및 선호도 입력 선택, 스피커 공간 정보 확인 및 가상 위치 조절, 3D 효과 제어 등의 정보를 제공 가능해야 한다. 그림 [3]은 관련 정보를 이용하여 능동형 실감 오디오를 복호화, 사용자 음원을 보정(3D 입체 음향 처리 및 입체 음상 정위 처리, 가상 공간 효과 처리) 하는 신호 처리 모듈을 보여준다. 그림[4]는 객체 기반 음원 제어 통합 입출력 장치를 이용한 시스템 구성도이다. 저작자가 제공하는 N개의 음원은 곡의 특성을 반영하는 Scene description 정보와 정합하여 객체 기반의 독립 음원 포맷으로 저장된다. 사용자 정보를 반영한 객체 기반 오디오 재생 플랫폼 구성을 위해 청취 환경에 관련한 기본 데이터 베이스 정보는 실시간으로 저장 및 검색이 가능하게 설계하였다. 기본 데이터베이스 정보는 청취 스피커의 환경 정보 및 개인 감정 정보, 곡 선호 정보와 사용자 인증 및 가입 정보를 포함한다. 기본 재생 정보를 이용하여 사용자가 원하는 곡을 선택하여 재생 가능하며 사용자의 감정 변화 및 재생 환경 변화 요구 입력 시 사용자의 기본 환경 정보를 바탕으로 스피커의 위치 및 객체 음원 재생 정보 변경, 재생 공간 및 청각 특성을 고려한 3D Effect 추가가 가능하다.

통합 인터페이스 장치는 실감 오디오에 대한 정보 확인, 사용자의 취향 입력 및 제어, 출력 콘트롤을 수행한다. 지능형 실감 오디오 재생 서비스뿐 아니라 유무선 전송 환경에 상관 없이 다수의 독립적인 기기에서 단일한 플랫폼으로 사용자의 입력을 수신하고 음원을 송출하는 통합 표시 장치를 개발하였다.

5. 결론

본 논문에서는 사용자의 감정 및 선호도, 청취 환경 변화를 다양한 네트워크상 입력 디바이스를 통해 전송하고 이를 바탕으로 처리한 객체 기반 음원을 다채널 스피커를 통해 출력하는 능동형 재생 플랫폼을 제안한다. 다양한 청취자의 개별 감성 및 환경 정보 등 음원 처리에 관련 데이터베이스를 실시간 저장하여 동일한 음원으로 단일 플랫폼에서 다양한 청취감으로 음원 재생이 가능하다. 청취자 맞춤형 감성 서비스를 비롯한 실감 사운드 관련 기술은 고화질 비디오 산업과

더불어 고부가가치 미래형 가전 시장의 핵심 기술이 될 것으로 기대한다.

[참고 문헌]

- [1] 홍진우, 강경옥, ETRI, 음향 선별 UCC 오디오 기술 개발, 보도자료, 2007.10
- [2] ISO/IEC 14496-11, Information technology - Coding of audio-visual objects, Part 11: Scene description and Application engine (BIFs, XMT, MPEG-J), Dec. 2005.
- [3] ISO/IEC 14496-20, Information technology - Coding of audio-visual objects, Part 20: Lightweight Scene Representation (Laser), June 2006.
- [4] D. R. Begault, 3D sound for virtual reality and multimedia, Academic Press, Cambridge. MA, ch. 4, pp. 132-136, 1994.