비압축 3D HD 영상 및 다채널 음성 전송

Uncompressed 3D HD Video and Multi-channel Sound Transport

채종권, JongKwon Chae, 이영한, Young Han Lee, 김종원, JongWon Kim, 김홍국, Hong Kook Kim
광주과학기술원 정보통신공학과

요약 국가간 연구목적으로 개설된 초고속 광 네트워크 기술의 발전은 새로운 응용 기술의 등장을 요구하고 있다. 고화질 저지연의 실감 협업 응용은 이러한 연구 목적에 부합할 뿐만 아니라 향후 커뮤니티 기반 응용 기술의 요구를 충족시킬 것으로 보인다. 본 논문에서는 실감 협업 응용 기술에 필요한 비압축 HD stereoscopic 영상 전송 시스템을 구성해 3D HD 영상을 사용자가 체감할 수 있도록 한다. 또한, 소프트웨어 기반 다채널 음성 재생을 다루며 실험을 통해 방향성 있는 협업 환경 구축의 가능성을 보여준다. 입체감 있는 미디어 재생을 위해 병렬 구조의 좌·우 송수신 시스템을 구축 후 stereoscopic 비압축 영상 전송을 수행하며, 좌·우 영상 세션간의 인터 미디어 동기화 기법의 설계방법을 제안한다. 음성 재생 소프트웨어는 ALSA를 이용하여 구현하였으며 가변 데이터 길이 및 프레임 손실로 인한 채널 뒤섞임(channel swapping)을 방지하기 위한 버퍼를 재생 모듈 전처리단에 추가하였다. 초고속 네트워크와 비압축 미디어 전송의 결합은 IP를 이용해 다채널 음성 지원의 실감 HDTV를 가능케 하므로 이를 유용하게 활용할 수 있는 사용 시나리오를 살펴본다.

핵심어: *Uncompressed HD transport, stereoscopic synchronization, multi-channel sound collaboration*

1. 서론

Gigabit 초고속 네트워크의 등장은 네트워크 자원을 충분히 이용할 수 있는 고대역폭의 응용 기술의 등장을 요구해왔다. 커뮤니티 기반의 실시간 협업은 이러한 응용의 하나이며, 특히 비압축 HD급 미디어 전송 기술을 이용한 고대역폭, 저지연, 고화질의 실시간 미디어 스트리밍은 고품질을 요구하는 사용자의 수요를 만족시킬 수 있는 응용으로 대두되어 왔다. 이를 위해, 비압축 HD 미디어 전송 시스템을 개선[11]해 음성 프레임의 패킷화를 위해 비압축 음성 전송을 위한 RTP 페이로드를 정의, 오디오 세션과 비디오 세션의 인터 미디어 동기화, RTP 패킷의 특정 패킷 손실상황에서의 버퍼 관리등을 다루었다. 본 논문은 한층 더 몰입감 있는 협업환경을 위해 3D HD 영상 및 5.1채널 음성의 각 스트림을 비압축 HD 전송기술에 접목함으로써 협업 연구, 엔터테인먼트, 원격 의료 등의 요구를 만족시키는 시스템을 제안한다. 3D HD 영상을 재현하기 위한 방법을 위해 본 논문에서는 stereoscopic 영상을 사용해 입체감 있는 영상을 재현한다.

본 논문에서는 좌·우 2개의 비압축 HD 영상 전송과 5.1 채널의 음성 재생을 이용해 실감 협업 응용을 목표로 한다. 특히, 2개의 영상 세션과 1개의 음성 세션의 동기화 기법 및 다채널의 음성을 소프트웨어적으로 재생하는 모듈의 구현에 초점을 둔다. 이를 위해, 비압축 HD 미디어 전송 시스템을 이용해 stereoscopic 영상 획득, 3Gbps급의 고속 네트워크를 이용한 전송, 3D 재생 장치를 통한 입체감 있는 영상 재현이 가능하도록 시스템을 구축했다. 제안하는 시스템은 좌, 우측 영상 획득을 서로 다른 송신 시스템에서 수행하며, 전송 후 재생 또한 서로 다른 수신 시스템에서 수행된다. 즉, 병렬 세션 구조를 가지게 되며, 이 세션들간의 동기화 기법을 제안한다.

기존의 비압축 전송 시스템에서 음성 재생을 이용하기 위해서는 이를 처리하기 위한 하드웨어의 설치가 필수적이었고, 이는 구축 비용 및 서비스의 확장을 막는 요소다. 따라서 본 논문에서는 음성 재생 하드웨어를 대체할 수 있는 다채널 음성 재생 소프트웨어를 구현한다. 이는 기존의 하드웨어로 재생된 음성을 일반 사운드카드를 이용하여 재생할 수 있기 때문에 수신부에서 음성 재생을 위한 하드웨어의 설치 없이 사용할 수 있다는 장점을 가진다. 제안하는 비압축 stereoscopic 영상 전송과 다채널 음성 재생 시스템을 활용하여 고품질 HD 스트리밍 서비스와 다채널을 활용한 다자간 화상회의 서비스가 가능하며 이를 위한 사용 시나리오를 살펴봄으로써 본 응용 기술의 가능성을 검토한다.

논문의 구성은 다음과 같다. 2절에서는 HDTV over IP 의 배경 지식 및 관련 연구에 대해 알아보며, 3절에서는 전체적인 시스템 구성을 다룬다. 4절과 5절에서는 병렬 세션에서의 동기화 기법과 다채널 음성 재생 모듈 구현을 각각 다룬다. 6절에서는 실험 결과를 밝히며, 7절에서 제안하는 시스템의 가능성 있는 사용 시나리오를 알아본다.

2. 배경 지식 및 관련 연구

이 절에서는 HD 미디어를 IP를 이용해 전송할 때 필요한 배경 지식을 알아보고, 입체감 있는 영상을 전송하기 위해 기존에 제안된 방법들과 본 연구와의 차이점을 밝힌다.

2.1 배경 지식

비압축 HD 영상 전송을 위한 시스템은 크게 획득, 전송, 재생의 세 부분으로 나뉘어 동작한다. 비압축 HD 신호를 로컬 장치들 간에 전송하기 위한 표준은 SMPTE-292M이다. SMPTE-292M 신호의 최대 전송률은 1.485Gbps이며 이 신호에는 영상 및 음성 정보가 들어있다. 카메라로부터 받은 SMPTE-292M 신호를 송신 시스템의 HD-SDI (High Definition Serial Digital Interface) 인터페이스를 이용해 획득하며, 이는 1920x1080i (interlaced)의 8bit/10bit 샘플링된 영상이다. 획득은 29.97pfs(progressive segmented frame)의 속도로 진행되며 8bit와 10bit의 샘플링의 경우 각각 995Mbps와 1.327Gbps의 대역폭을 필요로 한다. 24bit 48Khz의 비압축 음성도 또한 HD-SDI 인터페이스를 통해 획득되며, 획득된 영상 및 음성 프레임은 전송모듈을 통해 패킷화 된다.

전송은 RTP/RTCP[1] 프로토콜을 이용하며, RTP 세션을 통해 미디어가 전송되며, RTCP 세션을 통해 송수신측의 상태 정보의 교환이 이루어 진다. 송신측은 RTCP SR 메시지를 이용해 미디어 timestamp와 NTP timestamp를 보내며, 수신측은 이를 통해 영상 프레임과 음성 프레임의 송신 시간의 차이를 알 수 있다. 수신측은 RTCP RR 메시지를 보내며 수신한 패킷의 바이트수, 지터, 패킷 손실 등의 정보를 전송한다. [그림 1]은 비압축 HD 영상 프레임을 전송하기 위한 RTP 패킷의 필드들을 보여준다.

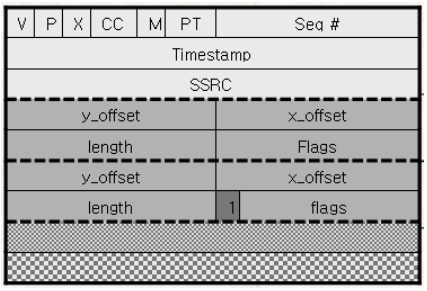

그림 1. 비압축 영상 전송을 위한 RTP 패킷.

점선 위의 필드는 RTP 패킷 헤더에 해당하는 12Bytes이며, 프레임의 마지막 패킷을 알리기 위한 M 비트 필드, sequence number, timestamp 필드를 통해 전송 순서 및 전송 시간을 반영하는 필드가 있다. SSRC필드는 사용자가 생성한 RTP 세션에 할당되는 고유 번호이다. RTP 헤더 다음의 offset 및 length는 RTP 페이로드 헤더로 RTP페이로드의 특정 부분이 비디오 프레임 내의 어느 위치에 해당하는지를 알려주는 역할을 한다. Flag 필드는 해당 페이로드 헤더가 마지막 인지 아닌지를 알려주는 역할을 한다. 이와 같은 RTP 패킷의 필드들을 프레임의 타입, 전송하는 미디어의 클럭 등의 정보를 반영해서 전송함으로써 각 세션 별로 미디어를 전송한다.

재생을 위해 수신측은 세션 참여자의 정보를 유지하고, 각 참여자 별로 재생 버퍼를 생성한다. RTP 패킷을 버퍼에 저장하고, M bit필드가 활성화된 패킷이 도착하는 즉시 재생 가능한 프레임으로 인식한다. 이때 재생시간이 계산되고 동기화 모듈에 의해 영상과 음성이 동기화된 후 재생 장치로 각 프레임이 보내져 재생되게 된다. 재생 시간 계산시에는 지터(jitter) 흡수 및 송수신 클럭 차이 보정[2]해주어야 하며 이렇게 계산된 재생시간은 동기화 모듈 부분에서 다른 세션의 프레임과의 동기를 맞춰주기 위해 추가적으로 보정이 이루어진다.

지금까지는 하나의 송신 시스템에서 하나의 수신 시스템으로 영상과 음성이 전송될 경우에 대해 알아보았다. 본 연구에서 다루는 병렬 세션간의 송신을 통한 stereoscopic 영상재생을 위해 세션을 병렬적으로 구성해 이용해 좌측과 우측의 영상을 각각의 송수신 시스템에서 받도록 한다.

2.2 관련 연구

입체감 있는 영상을 전송하기 위해 제안된 방법 중 하나는 UCLP (User Controlled LightPath)를 이용한 글로리아드망 위에서의 stereoscopic 영상 전송[8]이 있다. 이 연구에서는 1920x1080i의 MPEG-2 압축 영상으로 좌·우 영상을 획득하고 약 50Mbps의 대역폭을 사용해 전송을 수행했으며 stereoscopic 영상 전송 및 재생을 수행했다. 저비용의 HD 카메라로부터 영상을 IEEE1394 인터페이스를 통해 MPEG-2 TS(transport system)으로 받아 전송하는 해당 시스템은 3D 재생을 위해 두 개의 프로젝터를 사용해 재생을 한다. 이 연구와 본 연구에서 다루는 stereoscopic 영상 전송의 차이점은 비압축 영상의 장점인 저지연성에 있으며 이는 인터렉티브한 협업환경을 지원하기에 더 적합하다.

고속 IP망 위에서 1440x1080i의 HD MPEG-2 압축 영상을 이용해 입체 영상 구현 및 원격 의료에 활용[9]한 사례도 있다. HD 카메라에 비해 저비용의 HDV(High-definition Digital Video) 캠코더를 이용해 MPEG-2 MP@HL의 인코딩된 영상을 얻고, 이를 동기화된 다중화 기법을 이용해 stereoscopic 영상을 고속 네트워크를 통해 전송한다. 수신측에서는 역다중화 시켜 좌·우 영상을 획득한 후 소프트웨어적인 방법으로 병렬적인 디코딩을 수행 후 프로젝터를 통해 입체화면으로 재생한다. 여기서 사용한 방법도 또한 압축으로 인한 지연을 가지고 있으며, 영상의 해상도 또한 본 연구에서 목표 모델로 구축하는 1920x1080i 해상도에 비해 떨어진다.

비압축 영상을 이용해 입체감 있는 영상을 전송하는 시스템은 현재 상황에서 새로운 시도이며 가능성 있는 서비스 모델로서 연구의 가치가 있다고 본다.

3. 시스템 구성

이 절에서는 비압축 HD stereoscopic 영상 및 다채널 음성 전송 시스템의 전체 구성에 대해 소개한다. 하나의 비압축 HD 영상 세션을 전송/재생하기 위해 한 쌍의 송·수신 시스템을 이용하며, 좌·우측 영상을 전송해야 하므로 총 두 쌍의 독립된 송·수신 시스템이 사용된다.

Stereoscopic 영상을 전송하기 위해 송신 시스템과 수신 시스템은 기존의 획득, 재생장치를 좌·우 영상을 획득 재생할 수 있도록 구성된다. Stereoscopic 카메라의 추가비용을 없애기 위해 기존의 비압축 HD 카메라 두 대를 이용해 [그림 3 (a)]에서 보는 것과 같이 구성한다. 각 카메라는 좌측 또는 우측의 영상을 획득하고 이 영상 신호는 각각의 송신 시스템에 HD-SDI 인터페이스를 통해 입력된다.

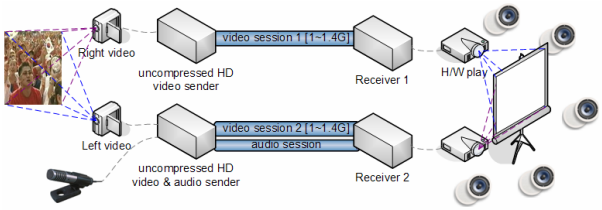

그림 2. Stereoscopic HD 미디어 전송 시스템 구성도.

음성은 하나의 세션만으로도 6채널을 전송하므로 하나의 송수신 시스템만을 이용해 전송한다. 해당 영상 세션들과 음성 세션을 전송하기 위해 본 연구에서는 8bit 샘플링 및 영상의 위아래 부분을 일부 잘라내는 방법을 사용해 2Gbps (1Gbps x 2) 네트워크 대역폭 요구사항 하에서 전송 실험을 수행했다.

그림 3. (a) Stereoscopic 획득 캠코더, (b) 3D displayer.

독립된 두 개의 송신 시스템은 획득한 영상 프레임 및 음성 프레임을 RTP/RTCP를 이용해 총 3개의 세션(영상-좌, 영상-우, 음성)을 맺은 후 전송한다. [그림 2]에서 보듯이, 수신 시스템은 하나의 영상 세션과 음성 세션을 받아들이며, 재생 버퍼에 저장 후, 재생 시간이 되면 프레임을 만든 후 출력장치를 통해 재생한다. 출력 장치는 [그림 3 (b)]에서 보이듯이 두 대의 고해상도 프로젝터를 이용해 하나의 뷰로 겹쳐 보임으로서 입체감 있는 영상을 보여준다. 고해상도의 비압축 HD영상을 입체감 있게 볼 수 있는 시스템은 이와 같은 형식으로 구성 된다.

보다 더 정확한 동기화를 고려한 모듈의 설계는 4절에서 소개된다. 요약하면, 한 수신 시스템은 다른 수신 시스템과 같은 시간에 획득된 프레임들의 재생시간 정보 교환을 통해, 먼저 도착한 프레임을 지연시키는 방법(RTP 프로토콜에서 기본적으로 제안하는 방법)을 이용하되 병렬적인 세션간의 프레임 송신 시간 정보 교환이 필요하므로 이 부분을 추가하는 방식으로 설계를 했으며, 구현을 위해 송수신측의 시스템 클럭, 미디어 클럭은 동일하게 설정한다.

음성 세션을 통해 수신된 6채널의 음성 프레임은 Advanced Linux Sound Architecture (ALSA) 기반으로 개발한 재생 모듈을 이용해 재생한다. 비압축 24bit 48Khz 6채널 오디오를 전송하기 위한 RTP 패킷 포맷의 정의는 [11]에서 다루었으며 본 논문에서는 수신측에서 전송 받은 오디오 프레임을 재생하기 위해 구현 시에 고려한 사항에 대해 자세히 다룬다. 이는 5절에서 소개되며, 요약하면, 가변적인 오디오 프레임의 구조로 인해 채널 변경을 막기 위해 메타데이터와 여분의 샘플을 저장할 수 있는 버퍼를 이용해 전송 에러나 지연에 의한 프레임 손실로 인한 채널 변경을 막을 수 있다는 것이며 이를 통해 6채널 음성의 재생이 수행된다.

4. Stereoscopic 영상을 위한 인터 미디어 동기화 설계

이 장에서는 병렬 구조를 이루는 미디어 세션간의 동기화 모듈 및 개발한 음성 재생모듈의 설계를 다룬다. 병렬 세션의 인터 미디어 동기화는 송신측과 수신측 별로 다루며, 좌·우 영상의 두 송수신 시스템들의 시스템 클럭 (system clock)은 서로 동일하지 않아도 되며, 시스템 클럭의 속도는 같다고 가정한다. 공통 레퍼런스 클럭 (common reference clock)은 NTP를 사용한다.

기본적인 설계는 송신측의 프레임 별로 할당되는 RTP timestamp와 RTCP SR메시지에 들어있는 NTP timestamp를 이용해 수신측은 해당 미디어가 NTP시간을 기준으로 어느때 보내졌는지를 판단하고, 다른 수신 시스템과의 프레임 시간 정보 교환을 통해 가장 근접한 시간에 보내진 또 다른 비디오 프레임을 찾아, 먼저 도착한 프레임을 늦게 도착한 프레임과 같이 재생시키기 위해 지연시키는 방법을 사용한다.

4.1 송신측 동작

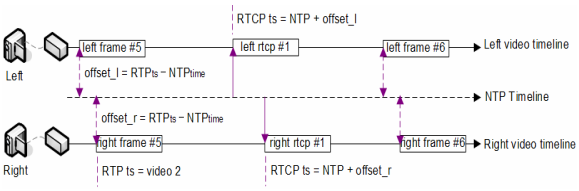

그림 4. 송신측 동작.

송신측에서의 동작은 [그림 4]에 나타나 있다. 그림의 좌측에는 두 개의 송신시스템이 카메라 신호로부터 획득한 영상 프레임을 각각의 독립된 세션을 통해 보낸다. 프레임은 RTP 패킷으로 분할(frame #5, #6)되며 이 패킷의 timestamp 필드에 미디어 클럭이 설정(*RTPts*)된다. 각각의 송신 시스템은 RTCP세션도 또한 유지하는데 이 세션을 통해 송신측은 Sender Report (SR) 메시지를 보내며, 이 메시지 내에 미디어 클럭(*RTCPts*)과 NTP 시간을 같이 보낸다. 이 정보를 이용해, RTCP SR 메시지를 받는 수신측은 수신되는 RTP 패킷의 timestamp만 보고도 이 패킷의 영상 데이터가 어느 시점에 획득되었는지를 NTP 시간으로 알 수 있게 된다. 수신측은 유지되는 RTCP 세션으로부터 RTP 패킷의 timestamp를 NTP시간으로 바꿀 수 있기 때문에 좌·우 송신측간의 시스템 클럭의 동기화는 필요조건이 아니다.

4.2 수신측 동작

수신측의 동기화를 위한 설계는 [그림 5]에 나타나 있다. 인터 미디어 동기화의 목표는 같은 시간에 획득된 좌·우 프레임이 수신되었을 때 늦게 수신된 프레임과 같이 재생 될 수 있게 일찍 수신된 프레임의 재생시간을 보정해주는 것이다. 그림에서 좌측 영상 세션을 기준으로 살펴보면, RTCP 메시지(left RTCP#1)가 수신되고 해당 메시지로부터 RTCP timestamp (*RTPts_L*)와 그에 해당하는 NTP timestamp (*NTPts_L*)를 알 수 있다. 때문에 RTCP메시지 이후에 들어

오는 RTP 패킷 (left frame #6)의 timestamp (*RTPts_L*)를 보고 해당 프레임의 송신측에서의 획득 시간 (*NTPts_L(Sender)* = NTPts_L + (RTPts_L-RTCPts_L) / video_clockrate_L*)을 계산 할 수 있다.

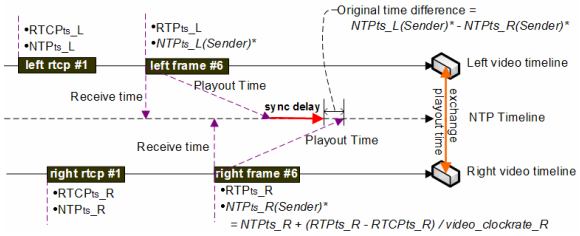

그림 5. 수신측 동기화.

이 획득시간을 바탕으로 해당 프레임의 재생시간을 계산할 수 있고, 우측 세션의 프레임 (right frame #6)의 재생시간과 비교해 먼저 들어온 좌측의 프레임을 synchronization delay만큼 지연을 시켜주어 송신측에서 획득된 두 프레임의 시간 차이 (*NTPts_L(Sender)* - NTPts_R(Sender)**)만큼 유지시켜 재생한다. 이 과정에서, 전체 시스템 구조에서 알아보았듯이 좌·우 영상 세션들이 독립된 상태의 병렬적인 구조를 갖고 있기 때문에 우측 세션의 프레임의 재생시간을 즉시 알지 못하는 제약이 생긴다. 이를 해결 하기 위해서 두 수신측은 전송 모듈의 시작과 동시에 재생시간 교환을 위한 추가적인 세션을 유지하는 방법을 사용할 수 있다. 좌·우 수신측은 동일한 로컬 영역에 있으므로 재생 시간 교환을 위한 세션의 측정값의 변동은 심하지 않을 것으로 예상되지만, 메시지 교환에 의한 추가적인 시간을 계속적으로 측정하며 보정해주어야 보다 정확한 동기화가 가능하다.

설계를 위해 추가적으로 고려해야 할 부분은 NTP의 정확도가 동기화에 미치는 영향으로, 더 정확한 공통 클럭을 사용하는 것이 전체적인 동기를 위해 중요하며, 이 경우에도 RTCP 메시지의 NTP timestamp 필드를 그대로 이용해 위의 절차를 그대로 사용 가능하다.

5. 다채널 음성 재생 모듈의 설계 및 구현

제안한 시스템은 비압축 영상 및 음성을 전송한다. 따라서 별도의 음성 코덱를 필요로 하지 않으며 RAW 데이터를 이용하여 재생한다. 사용되는 음성 형식을 정리하면 다음과 같다[3].

- 채널수: 6
- 방식: Little Endian
- 표본화 주파수: 48 kHz
- 샘플당 비트수: 24 bits

음성 포맷에서 샘플당 비트수가 24 bits으로 정의되어 있지만 본 시스템에서는 일반적인 24 bits 와는 다른 형식으로 저장되어 있다. 본 시스템에서 사용하는 음성 샘플의 구조는 [그림 6]과 같이 정의되어 있다.

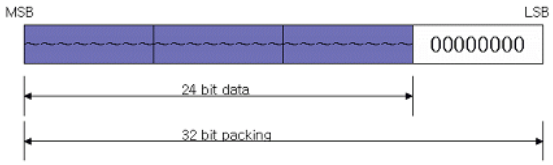

그림 6. 음성 샘플 구조.

비압축 전송 시스템에서는 24 bits 음성 데이터임에도 불구하고 32 bits 의 공간에 음성 샘플을 할당하는 방식을 사용한다. MSB (Most Significant Bit) 24 bits은 음성 데이터, LSB (Least Significant Bit) 8 bits은 zero-padding이 된 형식을 가지기 때문에 소프트웨어로 재생하기 위해서는 LSB 8 bits을 제거하는 과정이 추가되거나 음성 샘플을 32 bits으로 재생해야 올바른 재생 결과를 얻을 수 있다. 본 논문에서는 음성 재생 소프트웨어의 연산에 대한 부담을 줄이기 위해 LSB 8 bits을 제거하는 방식을 사용하지 않고 음성 포맷을 32 bits 로 설정하는 방식을 선택하였다.

본 논문에서 개발한 음성 재생 소프트웨어는 ALSA에서 제공하는 라이브러리를 이용하여 구현하였다[4]. ALSA는 리눅스에서 사운드카드의 설치를 위한 드라이버 및 사운드 관련 유틸리티, 사운드 프로그래밍을 위한 라이브러리 등을 제공하는 공개 소스 코드이다. 구현된 음성 재생 소프트웨어의 구성도를 간략하게 정리하면 다음 [그림 7]과 같다.

그림 7. 음성 재생 모듈 구성도.

비압축 전송 시스템에서 받은 음성 프레임은 음성 모듈로 전달되며 음성 모듈에서는 음성 프레임을 음성 데이터와 메타 데이터(meta data)로 구분해 해석한다. 본 시스템에서 정의된 음성 프레임의 구조는 다음 [그림 8]와 같다.

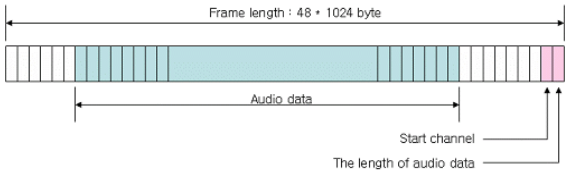

그림 8. 음성 프레임 구조.

[그림 8]와 같이 음성 프레임은 음성 데이터와 시작 채널, 프레임 내의 음성 데이터 길이를 알려주는 메타 데이터로 구성되어 있다. 앞의 6개의 샘플은 항상 사용하지 않고 7번째 샘플부터 메타 데이터에 기록된 오디오 데이터의 길이만큼을 오디오 데이터로 사용한다. 본 시스템에서 음성 데이터는 영상 데이터과 동일한 프레임 재생률인 29.97 psf를 가지기 때문에, 오디오 데이터의 길이가 1600 samples/data, 1605.33 samples/data의 2 가지이다. 따라서 음성 프레임의 길이는 같지만 내부에 존재하는 음성 데이터의 길이는 가변적인 구조를 가지고 있다. 또한, 0.33 샘플은 6채널 샘플 중에서 2채널 샘플에 해당하기 때문에 맨 처음 샘플에 해당하는 채널이 주기적으로 변하게 된다. 즉, 처음 샘플에 해당하는 채널이 1번째, 3번째, 5번째 채널로 변한다. 따라서, 시작 채널 메타

데이터를 이용하여 첫 번째 샘플의 채널 정보를 얻는다.

음성 데이터를 소프트웨어로 처리하는데 있어서 데이터의 길이가 가변적이기 때문에 발생하는 문제점은 재생 소프트웨어에서의 채널 뒤섞임이다. 비압축 전송 시스템에서 음성 모듈로 전달되는 프레임당 샘플 수는 1600개, 1605.33개로 2가지이지만 디바이스 버퍼의 크기는 설정한 음성 포맷에 따라 일정한 크기로 설정되기 때문에 [그림 9]와 같이 채널 뒤섞임이 발생하게 된다. 이러한 현상은 1, 2 채널에서 나오던 소리가 3, 4 채널 또는 5, 6 채널에서 들리게 되어 심각한 품질의 저하를 일으키게 된다. 따라서, [그림 9]에서와 같이 0.33샘플이 도착할 경우 다음 프레임이 도착한 후에 처리하는 버퍼가 필수적이다.

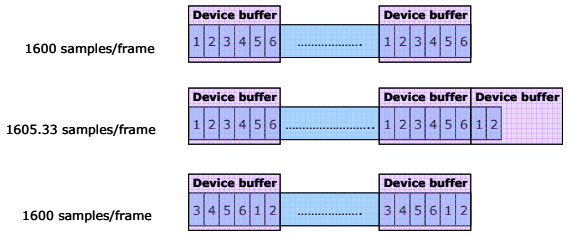

그림 9. 버퍼를 사용하지 않은 음성 재생 모듈.

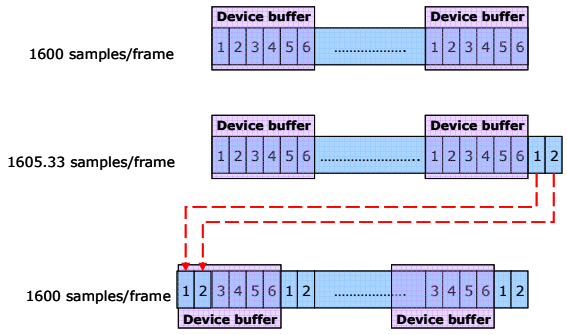

그림 10. 버퍼를 사용한 음성 재생 모듈.

본 논문에서 개발한 음성 재생 모듈에서는 [그림 10]과 같이 전송된 음성 데이터에 대한 버퍼를 이용하여 채널 뒤섞임을 방지하였다. 버퍼를 이용할 경우 0.33 샘플에 대한 처리를 다음 프레임의 오디오 데이터와 함께 처리할 수 있기 때문에 채널 뒤섞임을 방지할 수 있다. 또한, 오디오 프레임의 손실에 대비하여 [그림 11]과 같이 시작 채널의 메타 데이터와 버퍼에 입력되는 첫 샘플의 채널을 비교하는 과정을 추가하였다. 본 개발에서는 버퍼의 크기를 4800 샘플로 정의하였고 링 버퍼의 구조를 이용하여 버퍼를 설계하였다.

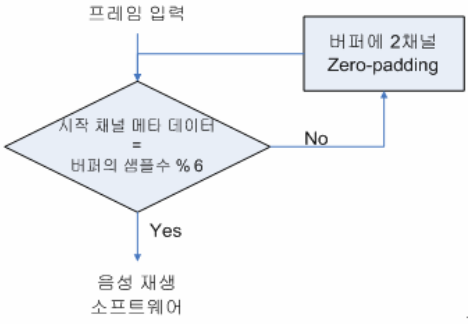

그림 11. 프레임 손실에 의한 채널 뒤섞임 방지.

6. 실험결과

6.1 비압축 stereoscopic HD 미디어 전송

저지연성의 입체감 있는 stereoscopic HD 전송을 위해 3절에서 언급한 시스템 구성을 갖추고 실험을 수행했다. GIST가 stereoscopic 영상의 송신측이 되며 KISTI가 수신측이 됐다. CPU 부하를 줄이기 위해 송수신 측의 네트워크 인터페이스 카드의 MTU 크기를 9180 bytes (점보 프레임)로 설정해 전송되는 패킷의 개수를 줄였고, 송신큐의 크기는 3000 byte로 설정했다. 수신측에서 프로젝터를 이용해 재생하기 위해 8bit로 샘플링해 송신하게 되었고, 좌·우 영상 및 음성을 보내기 전송하기 위해 약 2 Gbps의 네트워크 대역폭이 사용되었다. 수신측에서는 두 대의 프로젝터를 이용해 좌측 또는 우측의 영상을 재생함에 따라 두 프로젝터의 출력은 한 화면에 비춰지게 된다. 사용자는 좌·우측을 구별할 수 있는 안경을 사용해 스크린을 응시하게 되면 입체적인 영상을 볼 수 있게 된다.

그림 12. 비압축 stereoscopic HD 전송 실험 결과.

수신측의 출력 결과화면은 [그림 12]에 표현되어 있다. 해당 영상은 GIST의 실험실을 촬영한 것이며 좌·우 영상이 한 화면에 겹쳐져 있음을 확인할 수 있다. 몰입도를 높이기 위해서는 송신측 및 수신측에서 카메라의 간격, 각도, 높낮이 및 양쪽 영상의 간격을 조정해주어야 한다.

6.2 다채널 음성 재생 모듈

개발된 음성 모듈을 사용해 한국에서 보낸 다채널 음성을 미국 SuperComputing06 행사에서 재생하는 시연을 수행 했다. 시연은 비압축 HD 미디어 전송 시스템의 다채널 비압축 오디오 재생부분에 중점을 맞춰 참석자들에게 방향성 있는 HD 미디어를 체감하게 해주는 것에 목적을 두고 진행되었다. 다채널 오디오를 재생하기 위해 영상은 8bit로 보내진다. 영상과 음성의 재생은 소프트웨어 모듈을 통해 이루어지며, 그렇기 때문에 수신측은 HD-SDI 인터페이스 및 A/V 컨버터 없이 시스템의 그래픽 카드와 사운드 카드를 이용해 HDTV를 재생하게 된다.

전송을 위해 사용된 네트워크는 L3 대륙간 IP 망을 할당받았고, L2가 아닌 L3이므로 네트워크 상황은 다른 트래픽에 의해 변하며 이는 미디어의 재생 품질에 영향을 주는 요소가 된다. 전송을 위해 네트워크 인터페이스 설정은 점보 프레임을 사용했다. [그림 13, 14]는 수신측에서 재생시 측정된 프

레임 재생률 및 지터를 표현하고 있다. 프레임 재생률은 29.97이 대다수를 차지했으며 29.5 ~ 30.5프레임 내에 95% 이상의 프레임이 제때 정상 재생되었음을 알 수 있다. 송수신 과정에서 손실은 0% 였다. 측정된 지터는 오디오와 비디오 별로 차이가 있다. 비디오 프레임 송수신시 지터는 92% 이상이 10usec 이내였다. 이는 비디오 패킷의 전송이 지터의 영향을 거의 받지 않고 안정적으로 전송되었음을 의미하며 그러므로 일정한 재생률도 또한 보일 수 있었다. 오디오 세션을 통한 전송의 경우 대부분의 지터는 100usec ~ 300usec에서 발생되는데 이는 비디오 세션의 지터 측정값에 비교해보면 큰 값이다. 이는 송신 시스템에서 오디오 전송은 비디오 전송 후 시작되기 때문에 생기는 영향 및 오디오 세션이 상대적으로 비디오 세션에 비해 낮은 대역폭을 차지하기 때문에 생기는 결과로 해석된다.

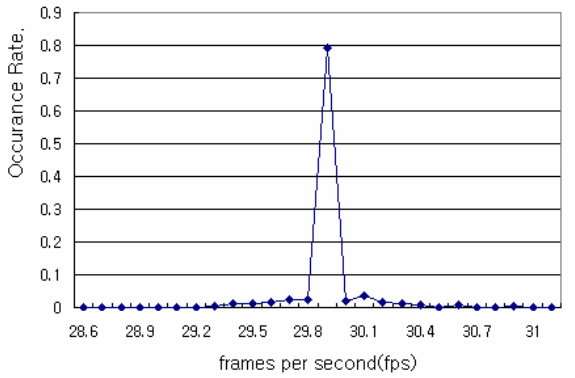

그림 13. Frames per second (fps) estimation.

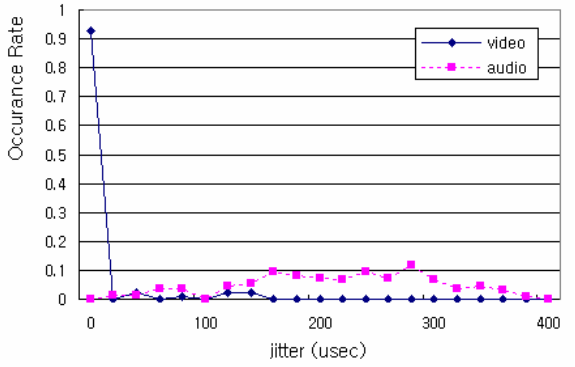

그림 14. jitter estimation.

시연은 5.1채널 음성이 들어있는 컨텐츠와 스테레오 음성이 있는 영상을 송수신하는 것이었으며, 네트워크 상황에 따라 가끔 음성 재생의 품질 저하가 있기는 했지만 분리된 채널이 각 스피커 별로 재생됨에 따라 입체적인 음향을 갖는 HDTV를 체감할 수 있는 가능성을 보여주었다.

7. 사용 시나리오

비압축 stereoscopic HD 전송 및 다채널 음성 재생 소프트웨어를 이용한 서비스로 고품질 실감형 HD 스트리밍 서비스와 5.1채널 음성 할당을 통한 다자간 화상회의 서비스를 고려해 볼 수 있다.

7.1 고품질 실감형 HD 스트리밍 서비스

고품질 실감형 HD 서비스는 HD 영상과 5.1 채널 오디오로 구성된다. 기존의 오디오 서비스는 모노, 스테레오, 5.1 채널 오디오로 서비스의 품질이 향상되어 왔고, 이와 함께 주어진 대역폭에서 최대 품질을 제공할 수 있는 압축기술이 발달해왔다. 특히 5.1 채널 압축을 위해서는 Dolby Digital, MPEG AAC 등이 사용되었다[5]. 하지만 이러한 압축기술은 비트율을 낮추기 위해 손실 압축(loss compression) 코덱이 주로 사용되었다. 따라서, Dolby Digital이나 MPEG AAC는 낮은 비트율을 가지지만 원음과 비교하여 낮은 음질을 제공하였다. 최근 네트워크의 발달로 인해 충분한 대역폭이 확보되면서 낮은 비트율보다 음질의 중요성이 높아지고 있다. 즉, 무손실 오디오 서비스에 대한 사용자의 요구가 높아지고 있다. 본 시스템은 비압축 전송 서비스를 제공하기 때문에 원음을 그대로 재생하는 고품질 오디오 스트리밍 서비스가 가능하다. 특히, 48 kHz로 표본화된 24 bits의 6 채널 오디오를 제공하기 때문에 고품질 오디오 재생에 적합하다. 또한, 구현된 다채널 음성 재생 소프트웨어를 이용할 경우 추가적인 AV (Audio/Video)장비 없이 음성 재생이 가능하기 때문에 서비스 확산에 큰 도움을 줄 수 있다.

7.2 채널 할당을 통한 다자간 화상회의 서비스

본 논문에서 개발한 음성 재생 소프트웨어는 멀티캐스트를 이용한 다자간 화상회의 서비스에 활용이 가능하다. 다자간 화상회의 서비스는 1:1 화상회의와 달리 회의에서 여러 명이 동시에 이야기하는 경우가 발생한다. 이와 같은 현상이 발생하였을 때 청취자는 대화 내용의 대부분을 이해하지 못하거나 자신이 집중하는 소리만 듣는다[6]. 하지만 음원을 다른 공간에 분리하여 분리하여 배치할 경우 동일한 위치에서 음원을 재생하는 것보다 대화의 이해도를 높일 수 있다.

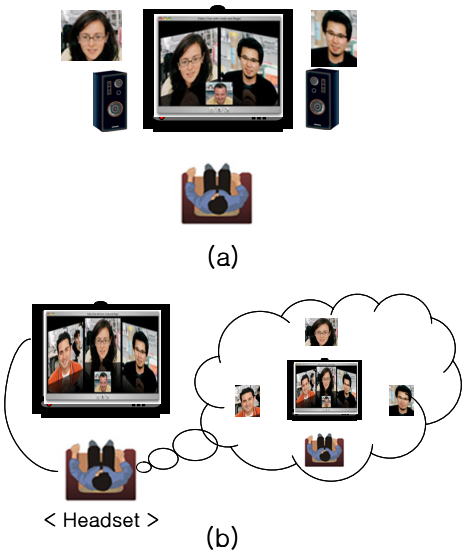

그림 15. (a) 스테레오 할당 방식 (b) HRTF 정위 방식.

기존에는 스테레오 채널을 이용하여 다자간 화상회의 서비스를 제공하였기 때문에 [그림 15 (a)]와 같이 좌·우 각각 1명씩을 할당하여 다자간 화상회의에서 이해도를 높이는 스테레오 할당 방식과 [그림 15 (b)]와 같이 HRTF (Head-

Related Transfer Function)을 이용하여 가상으로 스테레오 채널에 여러 명의 음성을 정위(localization)하는 HRTF 정위 방식이 사용되어 왔다[7]. 하지만 각 방식에는 다음과 같은 제약이 존재한다. 스테레오 채널을 이용하여 좌·우 각각 1명씩을 할당하는 경우 최대 2명에 대한 음성을 할당할 수 있고, 그 이상의 음성을 할당할 경우 회의에서 오가는 내용에 대한 이해도를 낮추게 된다. HRTF를 이용하는 방법은 보다 많은 음성을 분리시켜 정위함으로써 스테레오 할당 방식에 비해 많은 음성을 할당할 수 있지만 HRTF가 개인의 신체특성에 민감하기 때문에 모든 사용자에게 품질을 보장하지 못한다. 또한 헤드셋을 사용할 경우 음상 정위 효과가 가능하지만, 스피커를 이용할 경우 상호간섭 효과(crosstalk effect)가 발생하기 때문에 품질이 저하된다.

본 논문에서 개발한 다채널 음성 재생 소프트웨어는 최대 6 개의 독립된 채널을 지원하기 때문에 [그림 16]과 같이 각 채널에 음성을 할당하여 최대 6명의 다자간 화상회의에서 이해도를 향상시킬 수 있다는 장점이 있다. 특히, 상호간섭 효과 및 사용자 특성에 민감한 HRTF를 이용하지 않고 독립된 채널에 음성을 할당하기 때문에 스피커 환경에서도 회의의 이해도를 높일 수 있는 서비스의 지원이 가능하다.

그림 16. 다채널 음성 재생을 이용한 화상회의.

8. 결론

본 논문에서 실감 협업 환경을 구성하는데 필요한 비압축 stereoscopic HD 미디어 전송 및 다채널 오디오 재생 시스템을 다루었다. 또한, 좌·우 영상 세션의 인터 미디어 동기화에 대한 설계 및 다채널 음성 재생 소프트웨어를 구현을 하고 이를 실제 환경에서 테스트를 수행해보았다.

입체감 있는 영상을 획득, 전송, 재생해 고화질 저지연성을 갖는 HD 협업 환경의 가능성을 실험을 통해 보였으며 동기화를 위한 설계를 송수신 각각의 측면에서 다루었다. 다채널 음성 재생을 위해서, 본 논문에서는 오디오 프레임의 구조를 분석하고 채널 뒤섞임을 방지하기 위한 버퍼를 전처리로 추가하였다. 또한 프레임 손실에 의한 채널 뒤섞임을 방지하기 위해 입력신호와 시작 채널 메타 데이터를 비교하는 과정을 추가하였다. 이를 통해 다채널 음성 재생을 위한 시스템 구성의 비용 절감을 이룰 수 있었다.

마지막으로 제안한 시스템의 가능성 있는 고품질 실감형 HD 스트리밍 서비스와 채널 할당을 통한 다자간 화상협업 서비스의 사용 시나리오를 알아 보았다. 이와 같은 고품질의 HD 실감 협업 시스템은 과학 기술뿐만 아니라 stereoscopic 스트리밍 서비스를 지원할 수 있으며 이를 위해서는 기본적으로 고성능의 네트워크 자원을 필요로 한다. 그렇기 때문에 개발된 시스템은 국가 간의 연구망을 활용한 입체 영상 스트리밍 서비스와 프리미엄급 미디어 서비스에 주로 활용될 것으로 전망된다.

Acknowledgment

본 연구는 한국과학기술연구원 지원 하에 GLORIAD 망의 활용을 위한 비압축 HD 미디어 전송 시스템 기능 개선을 위한 연구 협력으로 이루어졌습니다.

참고문헌

[1] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, “RTP: a transport protocol for real-time applications”, IETF RFC 3550, July 2003.

[2] C. Perkins, "RTP audio and video for the Internet,” Addison Wesly, Nov. 2005

[3] HD-SDI Capture Card User Manual (http://www.aja.com/pdf/support/XenaSDHD_manual.pdf).

[4] ALSA, http://www.alsa-project.org/.

[5] M. Bosi and R. E. Goldberg, “Introduction to digital audio coding and standards,” Kluwer academic, Norwell, USA, 2003.

[6] A, Barry, “A review of the cocktail party effect,” *Journal of the American Voice I/O Society*, pp. 35-50, July, 1992.

[7] Robust Audio Tool (RAT), http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/

[8] J. Jo, W. Hong, S. Lee, D. Kim, J. Kim, and O. Byeon, "Interactive 3D HD video transport for e-science collaboration over UCLP-enabled GLORIAD lightpath," *Journal of Future Generation Computer Systems (FGCS), Elsevier*, vol. 22, pp. 884-891, 2006.

[9] Kiyoung Lee and JongWon Kim, “Software-based realization of secure stereoscopic HD video delivery over IP networks,” *Proc. of SPIE*, vol. 6016, 2005.

[10] L. Gharai, C. Perkins, and A. Saurin, "UltraGrid: A high definition collaboratory," USC/ISI, Sept. 2005, http://ultragrid.east.isi.edu/.

[11] 채종권, 조진용, 김종원, 변옥환, "비압축 HD 미디어 전송 시스템 개선을 위한 설계 및 구현," *KICS 추계종합학술발표회*, 34, pp. 36, 2006.