멀티캐스트 화상회의를 위한 3-D 음향시스템 설계

준회원 김 영 오*, 정회원 고 대 식**

Design of a Three Dimensional Audio System for Multicast Conferencing

Young-oh Kim*, Dae-sik Ko** *Regular Members*

요 약

다수의 참여자가 존재하는 멀티미디어 화상회의 시스템에서, 참여자의 얼굴은 화상을 통하여 쉽게 구별할 수 있지만, 음성의 경우는 모든 참여자의 음성이 1차원적으로 처리되기 때문에 참여자의 구분이 어렵고 공간적인 실감을 느끼지 못한다. 본 논문에서는 HRTF(Head Related Transfer Function; 머리전달 함수)와 거리감 재생 기법을 이용한 3-D 음향재현 시스템을 구현하고, 멀티캐스트 화상회의 시스템의 적절한 화자 배치를 연구분석하였다.

고도각과 수평각을 이용한 청취실험결과, 수평각이 고도각에 비하여 양호한 방향감 구별 인지도를 보였으며, 특히 4명의 참여자가 존재하는 화상회의 시스템의 경우, 수평각 10°, 90°, 270°, 350°의 HRTF를 이용한 공간배치가 효율적인 것을 확인하였다. 끝으로 5인 이상의 참여자가 존재하는 경우와 현실감의 개선을 위하여 거리감이 이용될 수 있음을 제안하였다.

ABSTRACT

On multimedia teleconferencing system existing a number of participants, face of the participants can be perceived by visual image. However, differentiation of each participant's voice and spaciousness sense are very hard since voice of all participants is processed with one dimensional data.

In this paper, we implemented three dimensional audio rendering system using the HRTF(Head Related Transfer Function) and distance sense reproduction method and determined the optimal location of the participants for teleconferencing system.

In the results of the listening test using elevation and azimuth angle, we showed that directional perception of the azimuth angles were better than that of the elevation ang

les. Specially, we showed that participant location using the HRTFs of the azimuth angle 10°, 90°, 270° and 350° was efficient in teleconferencing system existing four participants. We also proposed that distance cue was used for enhancement of the reality and location of many participants more than five.

I. 서 론

21세기 정보화 시대를 구현하기 위하여 나라마다 국가차원의 초고속 정보 통신망이 구축되고 있으며 원격교육, 인터넷방송, 원격화상회의 시스템등의 활용이 확대되어 가고 있다. 인터넷을 이용한 원격화상회의 시스템의 구현은 인터넷이 공유망이기 때문에 충분한 대역폭 확보를 통한 QoS(Quality of service)를 보장받을 수는 없지만 ISDN이나 다중화

* 목원대학교 전자 및 컴퓨터 공학과(youngman@mokwon.ac.kr)
** 목원대학교 전자공학과 부교수
논문번호 : 99106-0328, 접수일자 : 1999년 3월 28일

기 같은 특별한 시스템의 요구 없이 저렴하게 구성할 수 있는 장점이 있다.

MBONE(Multicast backBONE)은 인터넷상에서 멀티미디어 화상회의를 위해 개발된 가상의 네트워크 환경이다[1]. MBONE에서는 멀티캐스팅 기술과 RSVP등을 통하여 인터넷 서비스에 필요한 대역폭을 확보할 수 있기 때문에 서로 다른 원격지의 회의 참여자들은 가상의 공간에서 회의를 진행할 수 있다. 이와 같이 다수의 참여자가 존재하는 화상회의 시스템에서, 음성은 1차원적으로 청취되기 때문에 각 참여자의 구분이 어렵고, 또한 공간적인 실감을 느끼지 못하는 단점이 있다. 이러한 문제를 해결하기 위하여 기본적으로 화자(speaker)의 화면을 특징적으로 표시해주는 시각적인 방법을 이용할 수 있으나 최근 활발하게 연구되고 있는 3-D 입체음향 기술[2,3,4]의 방향감 및 거리감을 통하여 화자의 구별은 물론 각각의 참여자가 마치 같은 한 공간에 있는 것과 같은 실감통신 방법을 이용할 수 있다. 즉, 화상회의 시스템에 3차원 입체음향 제어기술을 추가하여 화면의 화자위치마다 각각 음상을 정위시킨다면 현실감을 증가시킬 수 있을 뿐만 아니라 화자를 시각을 통해 판별하지 않고 음성만으로 화자를 판별하여 화자를 주시할 수 있기 때문에 화상회의는 물론 오디오회의 시스템에도 유용할 것이다.

본 논문에서는 3-D 입체음향 기술에서 방향감과 거리감 지각단서를 이용하여 화상회의 참여자의 음성을 특정한 위치에 정위함으로써 청취자가 화자를 즉시 구별함은 물론 현실감을 증가시킬 수 있는 PC 기반의 인터넷 멀티캐스트 화상회의 시스템을 설계하였다. 실험은 3-D 입체음향 연구에 널리 사용되는 있는 HRTF(Head Related Transfer Function; 머리전달 함수)를 이용한 음상제어 기술과 거리감 재생방법을 이용하였다. 또한, 제한된 대역폭을 고려하여 저비트율을 구현하기 위한 방법으로 HRTF의 표본화율을 낮추는 방법을 연구분석하였으며, 다수의 회의 참여자의 위치가 서로 잘 구별되는 음상의 효율적인 배치 방법을 제안하였다.

Ⅱ. 화상회의를 위한 3-D 음향시스템 설계

1. 3-D 입체음향

3-D 입체 음향은 음원에 방향감, 거리감, 공간감을 인위적으로 생성함으로써 현장감을 부여해주는 것이다[2,3,4]. 인간은 두 귀에 입사되는 소리를 방향에 대해 필터링을 하고 중이, 내이를 거쳐 인식하게 된다. 이와 같은 인간이 소리를 인식하는 과정을 더미헤드를 이용하여, 두 귀의 방향에 의존하는 필터링을 주파수 응답으로써 표현할 수 있으며, 이것을 HRTF 또는 대응하는 말로 Head-Related Impulse Response(HRIR)라고 부른다. HRTF는 한 쌍을 이루며 한 위치로부터의 소리가 두 귀로 어떻게 도달되는가를 설명할 수 있다. 그림 1은 HRTF의 측정 블록도로서 측정 장치는 스피커 시스템에 MLS (Maximum length pseudo-random binary sequences)신호를 가해 주는 구동부와 더미헤드 내에 설치되어 있는 마이크로폰의 출력신호를 측정하여 임펄스 응답을 결정할 수 있는 계측부로 구성된다. 그림 1에서 스피커는 더미헤드로부터 1.5m 떨어진 곳에 위치하며 패도를 따라 고도각을 조절할 수 있게 되어있으며, 수평각은 더미헤드 밑에 설치된 턴테이블에 의해서 조정된다. 이렇게 그림 1과 같은 장치를 이용해 측정된 HRTF는 그림 2와 같이 원음과 컨벌루션 연산으로 임의의 음원에 대해 원하는 방향감을 갖는 입체음향을 재생할 수 있다[5,6,7].

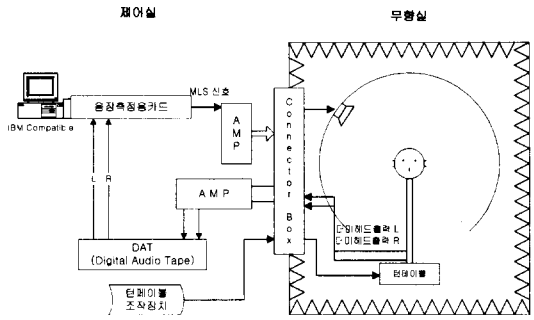


그림 1. HRTF 측정 블록도

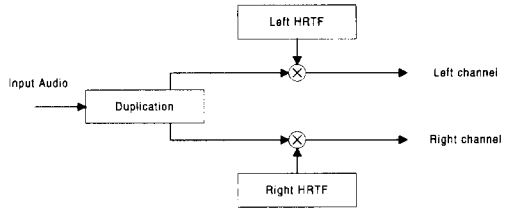


그림 2. 음원의 방향감 재생을 위한 컨벌루션

두 번째로 거리 변화에 따른 음향 신호의 거리 지각을 위해서는 소리의 크기, 직접음과 간접음의 비, 스펙트럼 모양, 그리고 두 귀사이의 차이 단서들이 가장 중요하게 다루어지고 있다. 이중에서도

음원과의 거리가 멀어짐에 따라 거리변화에 역비례하여 지수적으로 감소하는 소리의 크기 단서가 가장 영향력이 있다고 알려져 있다[4]. 또한 음 밀도와 잔향은 청각적 거리지각의 매우 중요한 단서로써 직접음과 반사음의 비를 이용하여 거리감을 변화시킬 수 있다. 그림 3은 지연과 이득소자를 이용해 소리의 크기와 잔향을 변화시키는 거리감 재생 필터 중의 하나이다.

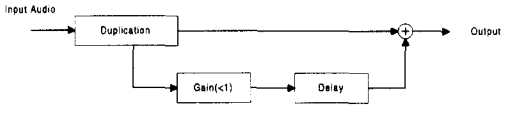


그림 3. 거리감 재생 필터

한편, 공간감은 물리적 공간의 음향 특성 정보를 나타내는 것이다. 공간감 생성을 위한 음향 환경은 크게 실내와 실외로 구분할 수 있으며, 주로 실내에서의 공간감 재현을 주로 다루고 있다. 실제 어떤 공간에서 측정된 실내 임펄스 응답(Room Impulse Response; RIR)과 같은 응답패턴을 통해 특정 공간(실내)이 갖는 음향 정보들을 추출해 낼 수 있으며, 이 정보들을 이용해 인위적인 공간 특성을 재현할 수 있다. 현재 공간감 재생을 위해서는 지연과 이득 파라미터를 이용한 콤(Comb) 필터가 주로 이용된다[8].

이상과 같은 입체음향 기술을 개별적으로 이용할 경우에도 참여자를 서로 구별할 수 있지만, 참여자가 많아질수록 음상배치 간격이 좁아지기 때문에 구별성이 감소하므로 방향감에 의한 참여자의 구분이 어려워진다. 이러한 경우, 거리감 및 방향감을 동시에 사용하여 참여자의 수가 많은 경우에도 참여자간의 구별성을 증가시킬 수 있다.

MBONE은 멀티캐스트를 위한 IP 멀티캐스트를 실제 공용망인 인터넷상에 구현한 가상 네트워크이며 호스트간에 2개의 병행 브로드캐스팅 채널(오디오, 비디오)을 이용한 회의 시스템의 멀티캐스트망 구축을 목적으로 개발되어 주로 실시간 멀티미디어 정보인 방송이나 화상 회의 용도로 사용된다[9]. 이것은 멀티캐스트가 1:다 또는 다:다 전송시 대역폭을 매우 효과적으로 사용할 수 있는 방법을 제공하고 있기 때문이다. 현재 MBONE상에서 화상 회의를 위한 오디오 정보는 입체음향과 같은 참여자의 구분이나 공간적인 특성을 고려하지 않은 1차원적인 음성 출력을 구현하고 있다.

2. 화상회의를 위한 3-D 음향 시스템 설계

인터넷상에서 화상회의 등의 응용을 위해 신뢰성 있는 데이터를 전송하기 위해서는 전송률을 가능한 저비트율로 유지하여야 하고 패킷손실도 감소시켜야 한다. 또한, 트래픽이 심한 경우 불가피하게 발생하는 패킷손실을 복구하기 위한 손실패킷 복구기술도 필요하다[10]. 그림 4는 네트워크상에서의 멀티캐스트 화상회의 시스템 구현을 위한 블록선도이다. 단, 그림 4는 화상 송수신 부분이 제외되었다.

그림 4의 각 송신단에서는 마이크로 입력되는 신호중 사용자의 음성 정보만을 인코딩하게 된다. 이 과정은 묵음검출 및 제거부와 음성압축부에서 수행하게 되고 이를 통해 전송률은 감소된다. 묵음검출 및 제거부에서는 마이크로부터의 입력이 없을 경우에 이를 검출하고 이를 전송하지 않게 한다. 이는 화상회의 중에 모든 사람이 동시에 말을 하는 것이 아니고, 사람의 말 또한 문장이나 단어 사이사이에 묵음에 해당하는 부분이 존재하기 때문에 이를 제거하여 전송률을 감소시킴으로써 대역폭을 효율적으로 사용할 수 있게 한다. 그러나, 음성정보는 기본적으로 8kHz, 8비트, 모노로 표본화되므로 이 음성정보는 전송률이 매우 높다. 따라서, 추가적으로 전송률을 감소시키기 위해 압축을 수행한다.

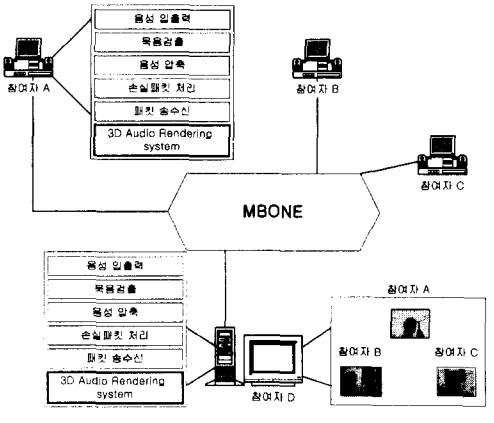


그림 4. 3-D 입체음향을 이용한 멀티캐스트 화상회의 시스템

압축기술로는 ADPCM(32, 16kb/s), GSM (13.2kb/s), LPC(2.4kb/s) 및 G.723.1(5.3, 6.3kb/s)등을 사용할 수 있다. 또한, 전송중에 발생될 수 있는 패킷손실을 복구하기 위한 처리를 수행하게 된다. 손실패킷 복구 기술로는 크게 송신자 기반의 복구기술과 수신자 기반의 복구기술로 구분할 수 있으

며, 화상회의에서는 송신자 기반 복구기술 중 코덱 기반의 복구기술인 [10]에서 제안된 잉여오디오 정보를 이용한 방법을 이용할 수 있다. 이렇게 생성된 정보들은 패킷화되고 연결중인 다른 사용자들에게 전송된다.

반대로 수신단에서는 수신된 패킷의 오디오정보를 디코딩하고, 전송중 발생된 손실패킷을 복구하게 된다. 복구과정을 거친 오디오정보는 재생되기 전에 3-D 음향 시스템에 입력된다. 3-D 음향 시스템에서는 오디오정보에 방향감과 거리감을 부여하기 위한 처리를 함으로써 각 참여자의 음성을 화자의 화상에 배치하는 기능을 수행한다. 우선 입력되는 오디오정보를 그림 2와 같이 HRTF와 컨벌루션함으로써 오디오정보에 방향감을 부여한다. MIT HRTF 데이터는 44.1kHz로 되어있기 때문에 오디오 정보와 컨벌루션을 수행하여 방향감을 부여하기 위해서는 오디오정보의 샘플링률과 HRTF 데이터의 샘플링률을 동일하게 하여야 한다[11,12]. 그러나, 오디오정보의 샘플링률을 높이는 것은 상대적으로 전송률이 높아지는 문제가 발생한다. 따라서, HRTF를 다운샘플링하여 오디오정보의 샘플링률과 동일하게 한다. 이때, HRTF가 갖고 있는 지각 단서의 손실을 적게 하는 것이 가장 중요하다. 이렇게 방향감이 부여된 신호는 다시 그림 3의 거리감 재생 필터를 거쳐 거리감을 추가할 수 있다. 거리감은 거리감 필터의 지연 및 이득값에 의해 결정된다. 따라서, 지연 및 이득값을 변화시켜가면서 적절한 파라미터값을 찾는 것이 중요하다.

이상의 과정을 통해 재생되는 오디오정보는 화상회의 시스템에서 화면의 참여자 위치에 대응되어 현실감을 증가시킬 수 있을 뿐만 아니라 화자를 시각을 통해 판별하지 않고 음성만으로 구별할 수 있도록 도와줄 것이다.

Ⅲ. 시뮬레이션 및 고찰

시뮬레이션을 위하여 한글 Windows 95를 운영체제로 하는 Pentium 200과 3-D 입체 음향을 재생하기 위한 사운드 카드로 Sound Blaster 32EL을 사용하였고, 입체음향 알고리즘(거리감, 방향감)의 시뮬레이션과 그래픽 해석을 위해 윈도우용 MatLab을 이용하였다. 음원은 사운드 카드에 의해 녹음된 오디오(음성) 데이터를 이용하여 시뮬레이션 하였다. 그림 5는 입체음향 구현을 위한 시뮬레이션 과정으로 PC로 입력된 음원은 MatLab에 의해 방향감과

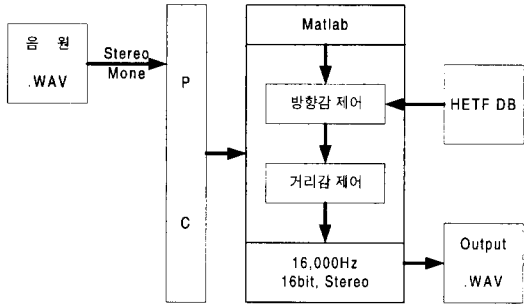


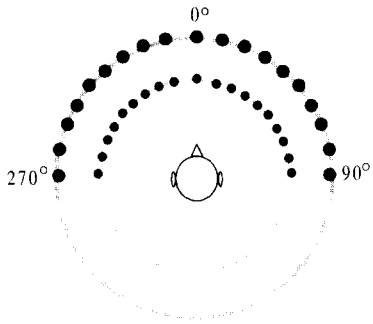
그림 5. 3-D 음향 재생 시스템

거리감을 갖는 입체음향으로 재구성되어 출력된다. 이때 방향감 제어를 위한 HRTF는 DB의 형태로 PC에 저장되어 진다.

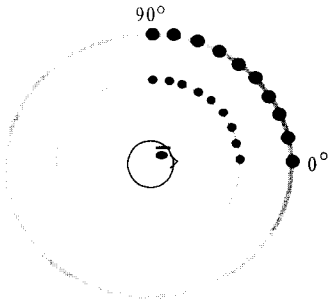
시뮬레이션을 위해 먼저 HRTF 데이터를 다운샘플링하였다. HRTF 데이터로는 MIT HRTF (44.1kHz)를 사용하였으며, 32kHz, 16kHz, 11kHz, 8kHz 순으로 다운샘플링을 반복하면서 입체음향 효과를 시뮬레이션한 결과, 16kHz 미만에서는 입체음향 효과가 거의 나타나지 않았다. 따라서, 입체음향을 위해서는 16kHz 이상으로 다운샘플링된 HRTF를 사용해야만 한다. 이때, 오디오 데이터 역시 16kHz, 모노, 16비트로 녹음된 것을 사용하였다. 이렇게 함으로써 PC 기반에서 부가적인 DSP 하드웨어 없이 실시간 컨벌루션 연산이 가능했다.

두 번째로 주관평가를 위한 데이터를 생성하였다. 그림 5의 시뮬레이션 흐름도와 같이 네트워크를 통해 수신되는 오디오 데이터는 16kHz로 다운샘플링된 HRTF와 컨벌류션 연산을 수행함으로써 음원에 대해 방향감이 부여된다. 시뮬레이션에 사용된 MIT HRTF 데이터는 -40° ~ 90°의 고도각 범위와 0° ~ 360°의 수평각 범위를 갖고 있지만, 본 실험에서의 방향감 재생 범위는 단일 청취자가 전방에 있는 화상의 중앙에 위치하고, 눈 높이 보다 높게 존재한다는 조건하에 시계방향으로 수평각 0° ~ 90°, 270° ~ 360°, 고도각 0° ~ 90° 사이의 HRTF만을 이용하였다. 또한, 방향성이 부여된 음원에 대해 그림 3에서와 같이 지연과 이득을 이용한 필터를 이용하여 거리감을 부여하였다. 이때 사용된 거리감 재생을 위한 파라미터들은 반복실험을 통하여 20 ~ 40msec의 지연과 0.5 ~ 0.7의 이득값을 갖도록 조정하였으며, 25msec의 지연과 0.7의 감쇠 파라미터를 사용하였을 때 거리감 지각이 잘 되었다.

그림 6은 음상의 구별성이 커다란 방향각도를 결정하는 실험을 위해 음상을 임의의 수평각과 고도



(a) 수평각 배치



(b) 고도각 배치

그림 6. 방향감 인지 실험을 위한 음상의 배치

각에 배치한 것을 보여준다. 그림 6에서 안쪽 부분은 HRTF만을 이용해 음상을 배치한 것이고, 바깥쪽 부분은 거리감을 추가로 부여한 음상의 배치이다.

음원에 고도각과 수평각의 방향성 및 거리감을 부여한 입체음향에 대해 일반 실내공간에서 10명의 피실험자를 대상으로 주관평가를 실시하였다. 평가의 목적은 정확한 재생방향에 대한 지각보다 음상 구분에 주안점을 두었다. 주관평가 결과, 피실험자들은 고도각을 거의 구별하지 못했으며, 고도각 0°에서의 수평각에 대한 실험에서는 대부분의 피실험자가 10°와 350°의 HRTF를 사용한 음원을 45°와 315°로 지각하였고, 90°와 270°의 HRTF를 사용한 음원은 80°와 280°로 지각하였다. 특히, 20°~90°, 270°~340° 사이의 HRTF를 사용한 음원에 대해서 거의 80°와 280°에 가깝게 지각하였다. 거리감 실험에서는 대부분의 참여자가 동일한 방향에 대해 거리감을 부여한 음원을 거리감을 부여하지 않은 음원보다 멀리 지각하였다. 이상의 실험 결과, 고도각에 비해 수평각에 대한 방향지각이 더 잘되는 것을 알 수 있었으며, 고도각 0°에서 10°, 90°, 270°, 350°의 HRTF를 사용할 경우, 음원이 가장 잘 구별되었다. 그러므로 4명의 참여자가 존재하는 화상회의 시스템의 경우, 10°, 90°, 270°, 350°의 HRTF를 이용하여 만들어진 음상을 각각 45°, 90°, 270°, 315°에 위치한 참여자의 화상에 배치하는 것이 효율적이다. 한편, 동일한 방향각도에 대하여 거리감을 이용하면 5명 이상의 참여자가 존재할 경우에도 참여자를 배치할 수 있고 현실감을 증가시킬 수도 있다. 그림 7은 실험 결과를 바탕으로 제안된 화상회의 시스템의 화상 배치와 음상 배치를 보여준다.

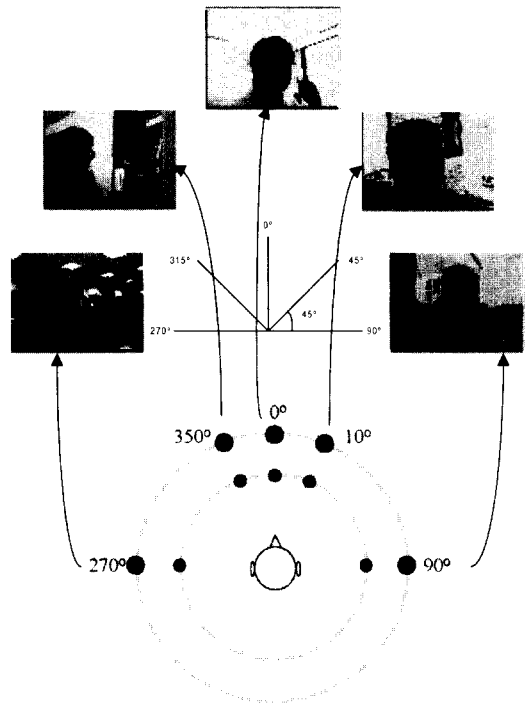


그림 7. 제안된 화상회의 참여자의 음상 배치와 화상 배치

Ⅳ. 결 론

본 연구에서는 MBONE 환경하의 다수 참여자가 존재하는 회의 시스템에서 3-D 입체 음향 기술인 방향감 및 거리감을 부여하여 실감있는 화상회의 시스템을 구축할 수 있음을 확인하였다. MBONE은 현재의 인터넷공용망에서 저비용으로 멀티미디어 화상회의가 가능한 환경이며 3-D 입체음향을 재생시킬 수 있는 최소 샘플링율은 16kHz임을 확인하였다.

입체음향에 대한 주관평가 결과, 고도각 0°에서 1 0°, 90°, 270°그리고 350°의 HRTF를 사용하였을 때 음상은 각각 45°, 80°, 280°, 315°로 지각되어 구별이 가장 크게 나타났으며, 25msec의 지연과 0.7의 감쇠 파라미터를 이용하였을 때 거리감을 지각하였다. 시뮬레이션 결과를 이용하여 화상회의 참여자를 배치시키는 방법은 그림 7과 같이 3명의 경우는 0°, 90°, 그리고 270°, 4명의 경우는 10°, 90°, 270°, 그리고 350°, 5명의 경우는 0°, 10°, 90°, 270°, 그리고 350°의 HRTF를 이용하여 참여자의 음상을 배치하는 것이 효과적임을 알 수 있다. 한편, 6명 이상의 참여자가 존재할 경우에는 거리감과 방향감을 조합하여 음상을 배치할 수 있으며, 5명 이하의 경우에도 거리감을 이용하면 현장감을 더 증가시킬 수 있다.

참 고 문 헌

[1] Deering S., "Host Extensions for IP Multicasting," *RFC* 1112, August 1989.

[2] Begault, D., *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Boston, MA., 1994.

[3] Durand R. Begault, *3D Sound*, Academic Press, Inc., 1994.

[4] 강성훈, *입체음향*, 기전연구사, 1997.

[5] Bill Gardner and Keith Martin, "HRTF Measurements of a Kemar Dummy-Head Microphone," *MIT Media Lab Perceptual Computing-Technical Report #280*, May, 1994.

[6] 김진욱, 고대식, 강성훈 외 1, "공간감, 거리감, 방향감 지각단서들을 이용한 실감 입체 음상 제어," *한국음향학회*, 추계 학술대회, pp. 411-414, 1997. 11.

[7] Middlebrooks, J.C., and Green, D.M., "Sound localization by human listeners," *Annual Review of Psychology*, Vol. 42, 1991.

[8] James A. Moorer, "About This Reverberation Business," *Computer Music Journal*, Vol. 3, No. 2, 1979.

[9] Roediger GA, Lidinsky WP, "The Multisession Bridge," *Computer Physics Communications* , V.110 N.1-3, pp. 149-154, 1998. 5

[10] 박준석, 고대식, "잉여 오디오 정보와 인터리빙을 이용한 패킷 복구 기술," *한국통신학회*, 추계 학술 발표 대회, pp. 399-401, 1997. 11.

[11] Wightman F. L., Kistler D. J., "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis," *Journal of the Acoustical Society of America*, Vol. 85, No. 2, pp. 858-867, 1989.

[12] Van Den Enden A.W.M., Verhoeckx N.A.M, *Discrete Signal Processing-An Introduction*, Pub., Prentice Hall, 1989.

김 영 오(Young-oh Kim) 준회원

1998. 2 : 목원대학교 전자공학과 공학사

1998. 3~현재 : 목원대학교 전자 및 컴퓨터공학과 석사과정

<주관심 분야> 신호처리, 인터넷 실시간 통신

고 대 식(Dae-sik Ko) 정회원

1982. 2 : 경희대학교 전자공학과 (학사)

1997. 2 : 경희대학교 전자공학과 (공학석사)

1991. 2 : 경희대학교 전자공학과 (공학박사)

1995년~1996년 : UCSB Post Doc.

1989년~현재 : 목원대학교 전자공학과 부교수

<주관심 분야> 신호처리, 인터넷 실시간 통신, 입체음향