

오디오의 심리음향

강성훈
대전보건전문대학 방송제작기술과

1. 서론

디지털 기술의 진보, 컴퓨터의 발달에 의해 CD나 DAT로 대표되는 오디오 기기의 기능이나 성능은 도약적으로 향상되었다. 예를 들면, S/N, 다이내믹 레인지는 아날로그 기기의 숫자와 비교하면 상당히 향상되고, 왜율은 10³ 오더 정도로 그 완성도는 최종단계까지 와 있다고 할 수 있다. 디지털 신호로 기록, 전송된 음성 신호는 아무런 노이즈의 영향을 받지 않고 그대로 전송되고 재생된다. 그러나, 최종 유저인 인간을 포함한 시스템을 생각하면, 사용 환경, 즉, 음과 인간이 직접 관계있는 음향 변환계는 종래의 구조로부터 탈피하지 못하고, 감성이나 휴머니즘에 호소하고 있는 영역이 결락되어 있으므로, 음질 결정의 중요한 포인트가 아직 해결되어 있지 않다. 그 중에서도 청각 지각과 물리 계측량과의 대응이다. 인간이 음을 들을 때 반응으로서 나타낸 감각량이 물리 데이터의 어느 부분에 대응하고 있는가를 아날로그 시대에는 매크로적으로 커버하였지만, 디지털 시대에는 종래와는 다른 해결 방법을 모색하여야 한다.

디지털 오디오에서는 데이터의 정밀도가 향상되어, 때로는 인간의 검지한계를 넘은 판단을 해야할 경우도 있다. 동시에 몇십 종류의 그리고 그것도 특성의 차이가 거의 없는 음을 들었을 경우에 정말로 인간이 정확히 검지할 수 없는 경우가 많다. 판단은 디지털적인 판단과 아날로그적인 판단이 있고, 0이나 1, 또는 있는가 없는가의 디지털적인 판단이라면, 별로 애매하지 않지만, 음질을 판단하는 것 같은 표현 감정이나 기술이 가미되는 것이고, 금방 아날로그적 판단이 되어 애매한 영역에 들어간다. 이와 같이 인간의 감각량을 수치화하여 통계처리함으로써, 관능을 과학적으로 해명하려고 하는 시도에서, 최근에 음질 평가법에 뇌파 등의 생체 신호를 측정하는 Biotechnology적인 접근이 이루어지고 있다. 인간을 대상으로 하기 때문에 그 분산을 고려하면, 막대한 데이터와 시간이 소요되며, 현재까지는 아

직 초보적인 단계이지만, 많은 측정 데이터가 축적되면 인간의 감성에 대한 데이터가 나올 가능성이 있을 것이다.

한편, 최근에는 영상 신호나 음성 신호의 대역을 압축하는 기술이 활발히 연구되고 있다[1~5] 이것은 멀티미디어의 정보로서 화상이나 음악을 효율적으로 저장하거나 전송하기 위해서이다. 이러한 신호를 디지털 화하여 그대로 저장하면 정보량이 너무 많기 때문에 처리시간이 길어지고 저장할 수 있는 정보량도 줄어든다. 따라서, 화질이나 음질의 열화없이 어떻게 정보량을 줄이는가 하는 점이 중요한 테마이다.

CD나 DAT와 같은 고품질 디지털 음향 신호를 기록하는 데는 아날로그의 음향신호에 비교하여 막대한 데이터 양이 필요하기 때문에, 한정된 전송대역이나 기록용량내에서 다 채널화 및 컴팩트화를 실현하기 위해서는 막대한 디지털 음향신호를 압축해야 한다. 따라서 최근의 기술 개발의 관심은 청각 감각상에서 CD품질을 유지하면서 데이터를 압축하는 것이다. 그 대표적인 것이 MPEG(Moving Picture Experts Group), SB-ADPCM(Sub-band Adaptive Difference PCM), Dolly Stereo Digital(AC-3), DCC(Digital Compact Cassette)나 MD(Mini Disk) 등이다. 이러한 오디오 신호 압축에 적극적으로 활용되고 있는 청각심리특성은 어느 주파수 성분이 존재하고 있을 때, 그 근방의 낮은 레벨의 주파수 성분은 지각되지 않는 마스크 현상과 최소 가청 한계의 특성 등이다. 그리고 최근에 레이저 디스크, Hi-Fi VTR, Video CD와 같이 Audio/Visual(A/V) 미디어를 통해서 음악을 즐기는 것이 일반화되어 가고 있다.

A/V미디어를 통한 정보 전달에 있어서는 소리와 영상이 중립되고, 시각과 청각의 상호작용이 일어나 시청자에게 보다 강한 인상을 주게 된다. 청각과 시각과 같이 Modality가 다른 감성들이 서로 어떠한 작용을 미치는가 하는 문제로 심리학의 분야에서도 오래전부터 연구되어 왔지만, A/V미디어의 보급 및 고품질 TV의 등장 등으로 그에 대한 관심은 점점 높아지고 있다.

본고에서는 최근에 디지털 오디오 개발에서 정보량을 압축하기 위하여 활용되고 있는 심리 음향 현상 및 멀티채널 오디오와 중저음에 대한 심리음향, 음향 재생계의 음질의 차이가 A/V 미디어를 통한 정보 전달에 있어서 청각과 시각의 상호작용에 어떠한 영향을 미치는가에 대해서 소개한다.

2 디지털 오디오와 心理音響

마이크로폰은 2개 사용하여도 전후나 상하 방향을 검출할 수 없지만, 인간의 청각은 2개의 귀를 가지고 정밀하게 음원의 방향이나 거리를 정확히 지각할 수 있다. 인간은 머리를 움직이어서 거리나 방향을 판정한다고 하지만, 머리를 움직이지 않고, 또 한쪽 귀를 막아도 방향 판별은 어느 정도 가능하며, 이것은 인간의 청각 능력이 우수함을 나타내는 한가지 특징이다. 컴퓨터가 급속히 발전되고 있는 상황이라고 해도 기계에는 인간과 같이 고도의 처리를 실시간으로 처리시키는 것은 불가능한 일이지만, 한발짝이라도 가까워지려는 노력은 계속되고 있다.

인간의 청각은 20Hz에서 20,000Hz 범위내에서 음을 들을 수 없는 최소 가청한계가 존재하며, 커다란 소리가 적은 소리를 마스킹(Masking)하는 현상이 있다. 마스킹 효과는 마스커(Masker)와 마스크(Maskee)의 주파수에 따라서 그 패턴이 달라진다. 이것은 최소가청한계와 마스킹 곡선 이하의 소리는 양자화 잡음을 허용하여도 들리지 않으므로, 이러한 신호성분은 전송하지 않아도 된다. 따라서 이러한 청각 특성을 이용하면, 정보량을 1/10정도까지 압축할 수 있다. 청각심리는 최소 가청한계, 마스킹 효과, 라우드니스, 임계 대역, 음색, 두 귀 효과(Binaural Effect) 등 많은 분야가 있지만, 여기에서는 오디오 데이터 압축에 사용되는 청각 성질만을 기술한다[1-4].

2-1. 최소 가청한계

청각은 20~20,000Hz의 음을 들을 수 있지만, 청각의 감도는 모든 주파수에 대하여 일정하지 않다. 그림 1은 등감곡선에서 최소 가청역치(minimum threshold)를 나타낸다. 그림에서 A와 같이 최소 가청역치보다 위에 있는 소리는 들을 수 있지만, A보다 레벨은 크지만 B와 같이 최소 가청역치 이하이면 들리지 않는다. 일반적으로 저역의 감도는 낮고, 주파수가 높아질수록 감도가 상승하여 3kHz 부근에서 가장 감도가 좋으며, 이 이상에서 점점 낮아지는 것을 알 수 있다.

어떤 음에 대해서 정상적인 청력을 갖는 사람이 그 음과 같은 크기로 들린다고 판단한 1,000Hz의 순음의

음압레벨의 수치를 라우드니스 레벨(Loudness Level, 단위:Phon)이라 한다. 저역에 대한 청각의 감도와 고역에 대한 청각의 감도 차이는 음의 레벨이 높아질수록 적어진다. 낮은 레벨에서 음악을 청취하면 저역이 잘 들리지 않지만, 레벨을 올리면 저음이 잘 울리는 것은 이와 같은 청각의 주파수 특성 때문이다. 다시말하면, 저역의 음은 고역의 음과 비교하여 음압 레벨이 상승하면 급격히 커지고, 저역에서는 다이내믹 레인지가 좁다는 것을 의미한다. 오디오 앰프의 라우드니스 스위치는 낮은 레벨에서 음악을 들을 때 이러한 청각특성을 보상하기 위한 것이다.

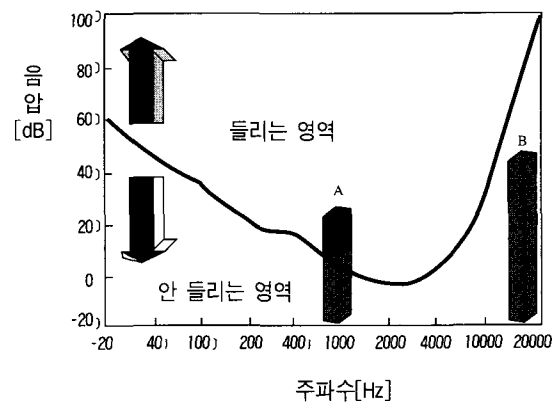


그림 1. 주파수에 따른 청각의 최소 가청한계

2-2 마스킹 현상(Masking Effect)

마스킹은 '어떤 음에 대한 최소 가청한계가 다른 음의 존재에 의해 상승하는 현상'으로 정의되고 그 상승치를 dB로 나타낸다. 2개의 음이 제시되는 조건에 따라서 동시 마스킹(Simultaneous Masking), 계속 마스킹(Temporal Masking), 두 귀사이의 마스킹(Contralateral Masking)으로 구분된다. 이러한 마스킹 현상의 연구는 오래전부터 연구되어 왔으나, 오디오 개발에 적극적으로 도입된 것은 MD나 DCC 개발시 정보량을 압축하기 위해서 도입되었다.

(1) 동시 마스킹

그림 2 같이 마스크하는 음(masker)과 마스크되는 음(maskee)이 동시에 제시되는 경우를 동시 마스킹이라고 한다. 그림은 순음 A가 존재하고 있는 상태에서 다른 순음이 어떻게 들리는가를 나타낸 것이다. A가 존재할 때 일정한 주파수 범위내에서는 다른 순음을 듣기 어렵다. 예를 들면, B는 레벨이 높은 순음이지만 들리지 않고, C는 레벨이 B보다 낮지만 들린다. 마스킹 효과는 마스커와 마스크의 주파수가 가까울수록 강해진다.

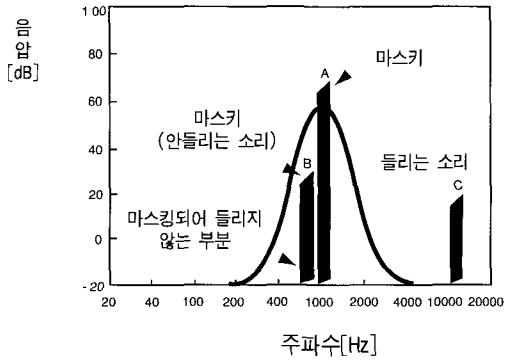


그림 2. 동시 마스킹 현상

(2) 계속 마스킹

그림 3과 같이 2개의 음이 시간적으로 계속 제시되었을 때 생기는 마스킹 현상이다. 시간적으로 선행하는 음의 후속되는 음을 마스킹하는 현상을 전방향 마스킹 (Forward Masing 또는 Post-masking), 후속되는 음이 선행하는 음을 마스키하는 현상을 역방향 마스키 (Backward Masking 또는 Pre-masking)이라 한다. 전방향, 역방향 마스킹에서 마스커와 마스키가 시간적으로 떨어지면 마스킹량은 급속히 감소된다. 음악이나 음성은 시시각각 레벨과 주파수가 변화되고 레벨이 큰 음 뒤에 레벨이 낮은 음이 계속되는 경우는 흔히 있는 일이다. 계속 마스킹은 일상생활에서도 음의 정보전달을 이해하는데도 중요한 현상이다.

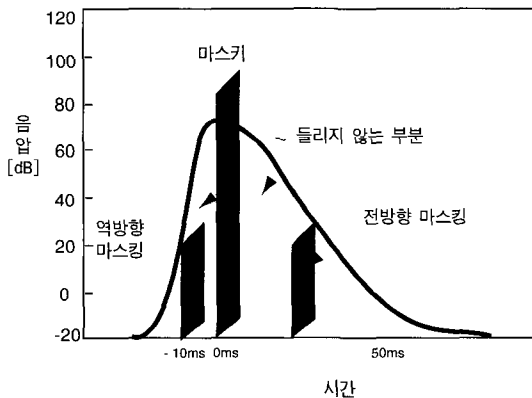


그림 3. 계속 마스킹 현상

(3) 두 귀사이의 마스킹

2개의 음을 좌우의 귀에 따라따로 들려줄 경우에 생기는 마스킹 현상을 말한다. 같은 쪽 귀 마스킹 (Ipsilateral Masking)과 비교하여 마스킹량은 50dB 정도

낮다. 이것은 마스킹하는 음이 體傳導(Bone Conduction)에 의해 반대쪽 귀에 전달되고, 거기에서 동시 마스킹이 발생하기 때문이다.

2-3. 임계 대역

청각계에 존재한다고 가정된 일종의 대역 필터이며, 라우드니스와 마스킹 현상 등에 있어서 그 양상이 급속히 변화되는 대역을 말한다. 임계대역의 개념은 1940년 Fletcher에 의해 제안된 임계비의 개념으로부터 시작된다. Fletcher는 순음의 대역 잡음에 의한 마스킹 역치를 측정할 때 대역 잡음(중심 주파수는 신호의 순음과 같음)의 파워를 일정하게 유지하고, 대역폭을 변화시켜 순음이 가까스로 마스킹될 때의 순음의 레벨과 대역잡음의 레벨 차이(Signal/Noise)를 구한 결과, 대역폭이 증가됨에 따라서 처음에는 S/N이 증가하였지만, 어느 폭을 넘어가면 S/N은 변화되지 않은 것을 발견하였다. 이것으로부터 Fletcher는 청각계의 필터의 존재를 제시하였고 그 후에 Zwicker는 체계적인 실험을 통하여 임계대역 개념을 명확히 하였다.

라우드니스 및 마스킹에 대해서는 다음과 같은 현상이 나타난다.

- 라우드니스 : 전체의 파워가 일정할 때 대역잡음의 크기는 임계대역내에서는 대역폭에 관계없이 일정하지만 임계대역을 이상으로 넓어지면 증가한다.
- 마스킹 : 순음을 대역잡음으로 마스킹할 경우에 순음의 파워와 순음의 주파수를 중심으로 임계 대역내의 잡음의 파워가 같을 때, 순음은 가까스로 마스킹된다. 대역 잡음외의 잡음은 마스킹에는 별로 기여하지 않는다.

Zwicker는 임계대역이 중요한 역할을 하는 라우드니스 레벨의 단위 'phon'을 도입한 Barkhausen의 이름을 사용하여 임계대역의 단위를 'Bark'라는 명칭을 사용하고 있다. 임계 대역 폭은 150Hz에서 100Hz, 1kHz에서 160Hz, 4kHz에서 700Hz 그리고 10.5kHz에서 25kHz로서 주파수가 높아짐에 따라서 넓어진다. 이 임계대역은 저역을 제외하고 280 Hz 이상에서는 1/3 옥타브 폭에 가깝다. 일반적으로 음향 분석에서 1/3 옥타브 필터를 사용하고 있는 이유는 이러한 임계대역의 개념 때문이다. 한편 DCC나 MPEG 오디오에서 사용되는 대역분할 부호화는 이 임계대역을 기본으로 하는 처리방식이다.

2-3 청각 심리의 적용 예

그림 1과 같이 인간이 들을 수 있는 주파수 대역은 20Hz~20,000Hz이고, 다이내믹레인지는 거의 90dB 정도이다. 오디오 기기는 이 가청 범위의 음을 재생하는 것이 요구되고 이러한 조건을 만족시키기 위해서는 1초

간에 14Mbit의 정보량이 필요하다. 직경 64mm의 MD에서 이 요구를 만족시키면, 불과 10분 정도 밖에 녹음 재생을 할 수 없으며, CD와 같이 72분간 녹음하기 위해서는 정보량을 1/5~1/10정도 압축하여야 한다. DCC도 아날로그 카세트와 대응성을 갖기 위해서는 1/4로 정보량을 압축하여야 한다. 음질을 열화시키지 않고 이러한 조건을 만족시키기 위해서는 청각특성을 고려한 기법들이 활용되어야 한다.

정보량을 압축시키기 위해서는 bit수를 줄여야 하지만 bit수와 양자화 잡음과의 사이에는 다음 식과 같은 관계가 있다.

$$L_{qn} = L_{sm} - 20 \log 2^n$$

L_{qn} : 양자화 잡음 레벨

L_{sm} : 최대 레벨

n: bit 수

이 식으로부터 양자화 잡음은 16bit에서 최대 레벨보다 96dB이고, 12bit에서는 -72dB, 8bit에서는 -48dB가 된다. 즉, 정보량을 압축하면 양자화 잡음이 증가하여 여러군데에서 양자화 잡음 레벨이 신호 레벨보다 높아지는 곳이 생겨 양자화 잡음이 들리게 된다.

MD에서는 bit수를 줄여도 이러한 양자화 잡음이 들리지 않도록 하는 것과 신호속에서도 다른 부분에 마스킹되어 들리지 않는 여분의 정보를 생략하는 것은 마스킹 이론을 이용하고 있다. 즉, 입력된 음원을 MD에서는 3개의 서브 밴드로 나누고, MDCT(Modified Discrete Cosine Transform)를 수행하며, DCC에서는 32 서브밴드의 필터로 분석하여, 신호음에 의한 마스킹 역치를 구한다. 이것을 기본으로 마스킹 역치 이하의 음은 코딩하지 않으며, 양자화 잡음은 마스킹 역치 이하가 되도록 한다. 또 이러한 처리는 인간의 귀의 신호처리 속도와 같도록 해야 하며 입력신호의 시간적, 주파수적인 특징이 있으며, 동시 마스킹뿐만 아니라 과도적인 신호에 대해서는 계속 마스킹도 고려하여 처리한다[4].

MPEG 오디오에서도 부호화 효율을 높이기 위해서 임계대역 등의 청각심리를 이용하고 있다. 우선, 신호를 주파수 대역을 32개의 등간격의 밴드로 분할하여 각각의 신호를 원래의 샘플링 주파수의 1/32로 서브 샘플링하여 부호화한다. 그리고 계속 마스킹 효과의 범위에서 가장 부호화 효율을 높은 값으로서 32의 정수배인 384샘플을 1프레임으로 하여, 각 서브밴드 12샘플, 32밴드로 변환하여 부호화한다. 한편 2개의 채널을 조합하여 부호화할 경우에는 청각의 위상 감지 능력이 높은 주파수에서는 낮다는 사실을 이용하여 높은 주파수에서는 신호를 모노로 전송하는 등의 부호화 효율을

한층 높이기 위해서 결합 스테레오 코딩 방법을 사용하고 있다.

이와같은 정보량을 감소시킬 경우에 청감은 실제로 양자화 잡음을 지각하지 못하는가가 문제이다. Houtsma는 bit수를 여러가지로 변화시켜 처리한 음과 완전한 것과 비교하는 청감 실험 결과에서 변별한계는 평균적으로 25~30bit/sample이다[2].

2-5. 들리지 않은 소리의 중요성

이상에서 기술한 바와 같이 음향 신호의 압축은 청각 심리 현상을 이용하고 있다. 즉, 음향신호의 압축이 각 시간마다 음향신호의 스펙트럼을 분석하여 그 중에서 스펙트럼이 큰 성분만을 남기고 이것에 의해 마스크되는 성분은 버리는 방식이다. 국제표준화 위원회에서는 시청실험 결과를 통해서, 1/4 정도로 압축하여도 CD 음질과 구별할 수 없다는 결론을 내리고 있다. 그러나 보다 중요한 점은 기준음질로 하고 있는 CD음질 자체가 실제 연주와는 상당히 다르다는 점이다. 화상의 압축기술과 음향신호의 압축기술을 비교하면, 화상의 경우에는 인간의 시각에 보이지 않는 곳은 주변의 밝기나 색을 얼버무리는 것이다. 이것에 대해 음성의 경우는 청각의 마스킹 현상을 이용하는 것으로서, 성분음이 적은 음을 버리는 것은 순간 순간의 음의 스펙트럼에 구멍을 뚫어버리는 것과 같다. 이것을 통화상의 경우와 비교하면 중간 중간에 드롭아웃이 있는 화상이 연결되는 것과 같다.

화상의 압축기술은 고주파 영역을 자르면 되지만, 이것을 음향 신호의 압축기술로 대치하면, 음성신호의 고역 성분을 자르는 것에 해당한다. 그러나 이것은 CD에 녹음할 때 대역을 22kHz로 제한하는 것으로서 이미 실행되고 있는 사실이다. 따라서 이 이상으로 음향신호의 대역을 압축한다는 것은 마스킹 효과와 같은 현상을 이용할 수 없는 것은 아닌가? 실제로 CD보다 1/4~1/5로 대역 압축한 음은 고주파 성분은 적고, 마치 넓은 초원에서 몇몇의 음각가가 질 나쁜 악기로 연주하고 있는 것처럼 들린다. 연주회장의 현장감이나 음장감을 전달하기 위해서는 연주 회장의 벽면으로부터 반사음이 중요한 역할을 한다. 그러나 이러한 반사음들은 레벨이 적기 때문에 레벨이 큰 직접음에 마스킹되어 버리므로, 이 성분들을 잘라 내버린다면 음장감이나 현장감이 재생되지 않는 것은 당연하다. 이것이 넓은 초원에서 연주하고 있는 것처럼 들리는 이유이다.

인간의 귀에는 마스킹 효과가 있다고는 하지만 인간은 레벨이 낮은 수 많은 성분음에 의해 악기의 섬세함이나 연주의 섬세한 뉴앙스 등을 느끼며 음악을 감상하고 있으며, 또 레벨이 낮은 많은 반사음이 음장감이나 현장감을 전달하고 있으므로 단지 들리지 않는다는

이유만으로 그냥 무시하여 버리기에는 너무 중요한 정보가 아닐까 하는 의견도 있다[5]. 이상에서 기술한 바와 같이 청각의 특성을 이용하여 데이터를 압축하여 많은 정보량을 적은 그릇에 담고자하는 노력이 이루어지고 있는 반면에 들리지 않은 소리의 중요성을 강조하면서 인간의 청각 주파수 특성을 훨씬 넘는 주파수 대역까지 소리를 재생하고 있는 연구도 진행되고 있다[6].

3. 멀티채널 오디오와 심리음향

입체음향에 관한 연구는 1930년대 초 영국의 Blumlein과 미국의 Snow에 의해 시작되었으며 상용적인 구현은 1950년대 미국 영화업자들에 의해 관객을 TV에 뺏기는 것을 막기 위한 방편으로 본격적으로 도입이 시작되었다. 입체음향은 모노포닉에 비해 크게 두가지 중요한 효과, 즉 입체음상(Stereophonic imaging) 효과와 청각적 효과(Auditory impression)가 있다[6]. 입체음상 효과란 오케스트라의 연주를 스테레오로 청취할 때 각 악기의 음상을 정확히 지각할 수 있는 것과 같이 여러 개의 음원(Sound source)의 위치를 독립적으로 분리하여 청각적으로 지각할 수 있도록 하는 음상(Sound image)을 제공하는 효과를 말한다. 청각적 효과란 입체음상 효과에 비하여 복잡한 청각적인 효과로서 크게 공간적 효과(Spatial impression)와 효과음(Effect sound)으로써 구분된다. 공간적 효과에는 여러 방향으로부터의 반사음을 재생함으로써 마치 콘서트 홀에 있는 것과 같은 느낌을 주는 방법과 잔향음을 이용하여 소리의 확산효과를 제공함으로써 실제로 존재하는 공간과 같은 곳에서 음을 듣는 듯한 느낌을 주는 방법이 있다. 전자는 초기 반사음(Early Reflections)에 의해, 그리고 후자는 지연 시간이 긴 잔향음(Reverberation)에 의해 효과를 얻을 수 있다. 효과음은 콘서트 홀에서 박수소리나 웃음소리, 빗소리, 바람소리 등을 현장감있게 제공하는 것을 말한다.

이상에 설명한 입체음향의 효과적인 재생은 여러 음원들로 구성된 음장(Sound Field)을 어떻게 효과적으로 재생하는가에 달려 있다. 입체음향의 실제적인 구현은 다수의 스피커를 이용한 음의 재생을 통해 이루어지는데 입체음상은 전면 스피커로 재생되고, 청각적 효과는 서라운드 스피커에 의해 재생된다. 현재까지 TV방송이나 FM 방송 등 대부분의 분야에서는 두 채널 스테레오가 무리없이 사용되어 왔지만 고선명 TV 등에서는 여러가지 문제점이 있다. 기존 두채널 스테레오의 가장 큰 문제점은 방향감이 불안정(Directional instability)하다는 점이다. 즉 두채널 스테레오에서는

두 스피커 사이의 중앙 위치에서 듣는 경우는 문제가 없지만 중앙에서 조금만 벗어나도 선행 효과(Precedence effect) 때문에 먼저 도달하는 스피커쪽에 음상의 정위되어 버린다.

그동안 두채널 스테레오가 무리없이 사용되어 왔던 이유는 입체음상 효과보다는 청각적 효과에 어느 정도 만족할 수 있었기 때문이지만, 고선명 TV에서는 현장감을 재현하기 위해 필요한 영상과 소리의 일치성(Directional coincidence)의 구현에 있어서 방향성의 불안정이 특히 문제가 된다. 즉 고선명 TV를 가정에서 3~4명 이상이 시청할 경우에 중앙에 위치한 사람을 제외하고는 나머지 사람들은 충분한 입체음상을 느낄 수 없기 때문에 고선명 TV의 원래 목적인 35mm 필름의 극장식 현장감을 실현하기 어렵다[7-10].

2채널 스테레오 재생이 가지고 있는 음상의 불안정함을 해결하기 위한 여러가지 실험과 방법들이 연구되어 왔다. 실험은 전면 스피커의 갯수와 위치를 변화시키면서 방향감의 오차를 측정하는 방법인데, 두채널에서 세채널(좌, 우, 중앙)로 증가시키면 방향감이 안정되는 효과가 얻어지지만, 네채널로 증가시킬 경우는 그다지 큰 효과가 없다는 점이 많은 실험들의 공통된 결과이다. 이들 실험을 종합해 보면 스피커의 수가 많을수록 입체음상을 재현하는데 유리하지만 가정에서 고선명 TV의 사용을 전제로 할 때 허용되는 방향성의 오차, 가격대 성능비, 전송 비용 등을 고려하면 대체로 세개의 전면 스피커를 사용하는 것이 적당하다.

또한 청각적 효과를 충분히 재현하는데 있어서도 전 방향 스피커만으로는 부족하다. 청각적 효과를 충실하게 재현하기 위해서는 서라운드 스피커를 이용하게 되는데, 이에 관한 실험도 역시 서라운드 스피커의 갯수와 위치를 가변시키면서 청각적 효과를 측정한다. 이러한 실험들을 종합해 보면 전면 스피커와 마찬가지로 가정에서의 사용을 전제로 하면 적당한 수준의 청각적 효과, 가격대 성능비, 전송 비용 등을 고려하면 한쌍의 서라운드 스피커의 사용이 적당하다.

전면 2채널 스피커와 서라운드 스피커 이외에도 저주파수 대역(200Hz 이하)의 오디오 신호를 재생하여 입체음향의 실감을 한층 더올리기 위해서 초저음(Sub-woofer) 스피커의 사용도 고려되고 있다. 저주파수 대역에서는 청각의 특성상 상대적으로 높은 신호레벨을 필요로 하며 또한 무지향성이므로 전면에 별도로 한개의 서브우퍼 스피커를 배치하는 것이 효과적이다. 이상을 종합하면 고선명 TV를 위한 오디오 채널의 구성은 그림 4와 같이 설치하는 것이 바람직하다.

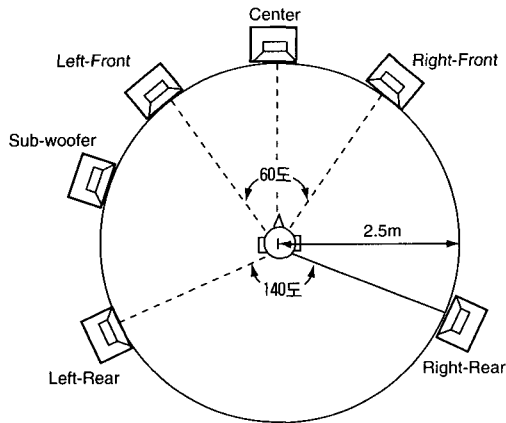


그림 4. 멀티채널 오디오 스피커 배치

4. 저음의 심리효과

인간에게 쾌적한 환경은 콘서트 홀이나 자연에 둘러싸인 환경 또는 모태안 등이다. 콘서트 홀 등 실제 음장에서의 현장감을 실현하려면 음악 신호만이 아니라 바닥을 전달하는 진동과 공기, 그것의 진동이 있는 수 Hz~100Hz 이하의 중저음 영역의 재생이 중요하다. 또 자연에 둘러싸인 환경에서, 예를 들어 산속에서의 바람과 나무의 술렁거림, 파도가 밀려오는 소리, 계곡물이 졸졸 흐르는 소리 등 모든 자연계의 음도 저음이 풍부하게 포함되어 있다. 마찬가지로 모태안의 환경도 스펙트럼은 20Hz에서 피크를 갖는 저음 우위의 환경인 것을 나타나고 있다. 그리고 이들 쾌적 환경은 모두 음으로 신체가 둘러싸이는 음장이다.

오디오에 있어서도 저음은 여러가지 의미를 가지며 파이프 오르간이나 대편성 관현악단과 같이 저음이 많은 음악에는 필요 불가결하고 더우기 소편성의 실내악에 있어서도 저역을 충분히 재생하면 음악이 갖는 분위기의 재현성이 한층 높아진다. 저음 악기를 사용하지 않는 음악에 있어서도 녹음 방법에 따라서는 음장 공간을 재현하는 정도가 저음에 상당히 많이 들어 있다. 예를 들면, 교회의 크기나 연주회장의 크기는 어떻게 저역을 넓히는가에 의해 그 연주회장의 크기까지도 재현할 수 있다. 이것은 실내의 울림은 체적이 클수록 실내의 공명도 낮아지므로 시각 이외에 청각도 공간의 크기를 느낄 수 있기 때문이다. 녹음기술의 현저한 진보는 악기 이외의 저음도 확실하게 픽업하여 마루의 울림이나 공조 노이즈까지도 음악 소스에 들어가 체크에 사용하는 경우도 있다.

현재의 A/V 장치에서도 중저음은 현장감이 재생하는데 커다란 역할을 한다. 특히 화면의 크기가 커질수록 저음과의 밸런스가 중요해진다. 100인치 화면 가득히 제트가 날아가면, 그것에 알맞게 중저역이 재생되고 음압 레벨도 상승해야 오디오와 비디오가 일치된다. 중저음 스피커(Sub-woofer 또는 Super-woofer)는 일반적으로 150Hz 이하의 주파수를 재생하고, 20~30Hz의 중저음까지도 재생할 수 있다. 일반적으로 우리들은 개인 차이도 있지만 저음을 느끼는 것은 70~80Hz이며, 50Hz까지 재현하는 악기, 특히 파이프 오르간은 16Hz까지 재생이 가능하다. 중저음을 느끼는 주파수는 60~100Hz이다.

한편 중저음이 갖는 생리적인 현상으로는 다음과 같이 설명할 수 있다. 일반적으로, 저음을 들을 경우 근육의 수축으로 고막의 임피던스가 증가하여 귀의 감도가 낮아진다. 그와 동시에 청각 대응으로서의 피부 감각과 진동 감각이 움직이는 이와같은 음을 청하여 촉각 음(Tactile Sound)라고 하며, 전신 진동의 주파수 대역은 1~100Hz이다. 이와같은 진동감을 저음감의 대응으로서 이용한 것으로 음악 신호에 맞추어 트랜스듀서로 인간의 피부와 뼈에 직접 국부적 진동을 주는 음향 장치도 있다. 그러나 음은 본래 공기 진동이므로, 저음감을 기계 진동으로 대응하기 보다는 공기 전반에 의한 자연적인 저음의 전송으로 신체 전체의 진동감을 만들어 내는 것이 이상적이라고 할 수 있다.

5. Audio/Visual의 심리

최근에 레이저 디스크, Hi-Fi VTR, Video CD와 같이 Audio/Visual(A/V) 미디어를 통해서 음악을 즐기는 것이 일반화되어 가고 있다. A/V 미디어를 통한 정보 전달에 있어서는 소리와 영상이 중첩되고 시각과 청각의 상호 작용이 일어나, 시청자에게 보다 강한 인상을 주게 된다. 청각과 시각과 같이 Modality가 다른 감각들이 서로 어떠한 작용을 미치는가 하는 문제는 심리학의 분야에서도 오래전부터 연구되어 왔지만 A/V 미디어의 보급 및 고품질 TV의 등장 등으로 그에 대한 관심은 점점 높아지고 있다.

A/V 미디어를 통해서 음악을 감상할 때 밝은 음이 화면을 밝게 하는 것과 같이 시각과 청각에 공통되는 심리적 성질(공통 양상성)을 같은 방향으로 변화시키는 공명 현상, 또 음과 영상이 중첩됨으로서 시각과 청각이 상승작용을 일으켜 음이 영상의 효과를 높이고, 영상이 재생음의 효과를 높이는 협합적인 상호 작용이 생긴다. 그러나, 영상을 부가함으로써 음의 더러움을 눈에 띄지 않게 하는 효과도 있다. 예를 들면, TV화면

에 피아노가 비치면 그 소리(80Hz~5kHz 대역)도 피아노 소리같이 들리게 되고, 이 경우에 소리에 불만이 어느 정도는 없어진다. 이와같이 오디오 시스템에는 용납될 수 없는 음질이 영상을 부가함으로써 허용이 되는 것이다. 감각의 감수성의 변화란, 예를들면, 음을 자극으로서 주는 것에 의해서, 시각의 감각이 증가(잘 보이게 되는) 하는 현상을 말한다. A/V 미디어에 의한 정보전송계에 이 현상을 적용하면 음과 영상을 동시에 제시함으로써, 음향 재생계의 음질에 대한 감각계의 감수성이 저하되고, 음질의 나쁨에 대해서 용납이 되는 것이다. 또는 역으로 청각계의 감수성이 상승하여 음질에 대해서 엄한 판단을 할지도 모른다. 그리고, 이와같은 현상이 실제의 시청공간에 있어서 나타난다면, 어떠한 상황에서 어떻게 나타나는가를 명확히 할 필요가 있다.

여기에서는 음향 재생계의 음질의 차이가 A/V 미디어를 통한 정보 전달에 있어서 청각과 시각의 상호작용에 어떠한 영향을 미치는가에 대해서 소개한다. 음향 재생계의 음질을 일부러 열화시키고, 영상의 부가에 의한 음향 재생계의 인상의 변화를 측정하여, 음질을 열화시키지 않은 경우와 비교함으로써 영상이 시각계의 감수성에 미치는 영향을 검토할 수 있다. 또한, 역으로 이러한 재생계의 음질의 차이가 영상의 영향에 대한, 음향 재생계의 영향에 어떻게 반영되는가도 알 수 있다.

1) 음향재생계의 대역 제한이 음악 재생음의 인상에 미치는 영향

Hi-Fi 조건과 대역 제한 조건의 인자 특점의 차이가 대역제한에 의한 재생음의 인상의 변화에 상당한다. 즉, 꼭 죄이는 인자와 풍부함의 인자의 차이가 현저하게 나타난다. 대역제한에 의해서 재생음이 헤이해지고, 빈약한 음으로 변환된다.

2) 영상이 음악 재생음의 인상에 미치는 영향

소리만을 들은 상태와 비교하여 영상을 부가했을 때에 음악 재생음의 인상이 어떻게 변화하는가에 대한 실험결과로서, 영상이 음악 재생음의 인상에 미치는 영향, 즉, 시각계가 청각계에 미치는 영향을 알 수 있다. 여기에서도 재생음의 인상에 대한 대역제한의 영향이 현저하게 나타난다. 앞에서와 같이 긴장감 인자와 풍부함의 인자에 변화가 나타난다. 대역 제한 조건에 있어서 풍부함의 인자에 있어서 대역제한에 의해 빈약해진 음이 영상을 부가함으로써 풍부함이 증가되는 경향이 있다. 즉, 영상이 마치 음질의 열화를 보상하는 기능을 하고 있는 것이다. Hi-Fi 조건에서는 영상을 부가해도 대역 제한 조건과는 달리 계통적인 변화는 나타나고

있지 않다.

3) 음악 재생음이 영상의 인상에 미치는 영향에 대한 실험결과에서는 영상만을 제시했을 때와 음악을 부가했을 때와 영상 인자 특점에 차이가 나타나지 않았다. 즉 청각계가 시각계에 미치는 영향은 없다는 것을 의미한다.

다음에는 음의 정위가 영상에 의해 어떠한 영향을 받는가에 대한 실험 결과에 대해서 소개한다.

1) 좌우방향 정위에 대해서는 영상이 없는 경우의 음상의 방향과 영상의 방향과의 차이가 10도 이내 일 때는 음은 거의 영상에서 오는 것처럼 지각된다. 그러나 방향 차이가 20도 이상이 되면 거의 영상의 영향을 받지 않는다(그림 5). 방향 차이가 10도에서 20도의 범위에서는 영향을 받는 정도가 사람에 따라서 개인 차이가 크고 또 그때마다의 변동도 크다.

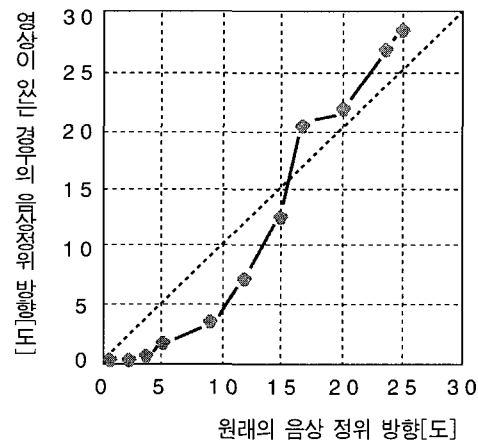


그림 5. 영상에 의한 음상의 변화

2) 대화면 디스플레이를 사용하여 실험한 결과에서는 음의 상하 방향감은 화면의 영향을 현저하게 받는다. 그림 6은 음을 같은 조건으로 재생하여도 영상의 위치가 상하 방향으로 바뀌면, 음의 상하감이 현저하게 달라진다는 것을 나타내고 있다. 즉, 음원의 위치는 변하지 않지만 영상이 제시되는 높이에 따라서 음상 정위가 변화하게 되는 것이다. 그림 6은 음과 영상의 좌우 방향이 일치한 경우이지만 좌우 방향으로부터 어긋나도 영상에 의해서 상하 방향으로 쏠리는 경향이 있다.

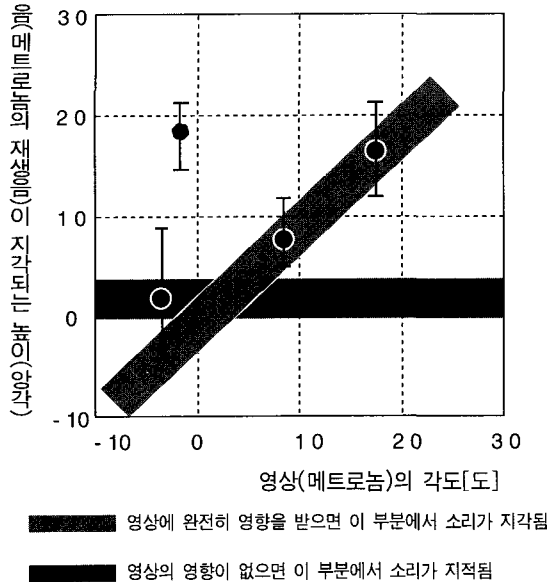


그림 6. 영상에 의한 음의 상하 방향감의 영향

음상 정위와는 다른 문제이지만, 어떤 음절을 들려 주고, 이것과 동시에 다른 음절을 발음한 영상을 보여 줄 때는 영상의 영향으로 음을 달리 지각하게 되는 현상이 있다. 예를 들면, 'BA'의 음만을 들려 주면 틀림 없이 알아듣지만, 'GA'로 발음한 입술모양을 영상으로 동시에 제시하면, 'DA'로 잘못 듣는 경향이 있다. 또 이런 종류의 예러는 취학전의 아동보다 성인에게 더 많이 일어난다. 입으로 발음하는 음성은 원래 청각에 의존하는 정보전달 수단이지만, 일상생활의 학습에 따라서 입술의 형태도 청각을 보조하는 이해의 수단이 되어 이러한 예러가 나오는 것으로 생각할 수 있다.

6. 맺는말

최근에 신문, 잡지, TV 등의 미디어에서 빈번하게 보도되고 있고, 일종의 시대의 경향이 되고 있는 것 중 에서 감성, 휴머니즘, 마음, 문화, 쾌적성 등 기술과는 약간 동떨어진 것 같은 표현이 있다. 이것은 현저한 기술개발에 의해 현대문명이 구축되고 인간의 일상생활에 도입되어 아무런 저항감없이 사용되고 있으며, 최근 디지털 기술이 탄생되어 너무나도 하드웨어 우선이 된 세상의 왜곡을 반영하고 있다고 할 수 있다. 즉, 실제로 기술개발이란 무엇인가라고 하는 것을 드디어 다시 생각해 보게 된 것이다.

아날로그적인 판단은 어느 기준에 대한 상대적 감각에 기인한다. 이것은 음질론뿐만 아니라, 행복론이나 인간의 의식에도 해당되는 철학이라고도 할 수 있다.

과거에서 현재까지 이르는 환경, 경험, 경력, 기억을 포함하여 구축된 기준에 대한 상대평가를 인간은 행하고 있는 것은 아닐까. 스피커의 재생이 아날로그이지만, 그 신호가 콤팩트, 고막, 청각 신경계를 거쳐 대뇌에 도달하고, 전체의 구성력으로 재구축될 때 비로서 완벽한 아날로그 감각량이 된다. 즉, 인간의 감각계가 아날로그인 이상(청각 신경계는 디지털이지만) 디지털 기기와의 매칭을 취하는 본래의 의미에서의 D/A 변환기가 필요하고 이것이야말로 Human Electronics일 것이다.

오디오 개발은 원래 있어야할 모습을 향하여 오늘의 상품으로부터 이상적인 상품에 도달하는 과정의 단계적으로 진행되는 오늘의 시점에서의 표현이라고 말할 수 있다. 오디오 기기는 필수품이라 보다는 취미품이기 때문에 기술자의 자유로운 발상으로부터, 기술자가 성능을 내기 쉬운 구조의 착상으로부터 태어난 형태가 많고, 보급되어 왔다. 보급도의 향상과 함께 유저와의 사이의 갭이 커지고, 오늘날 성능의 만족도가 향상된 시점에서의 유저는 오디오 기기의 형이나 조작성의 면에서 불만스럽다고 생각할 것이다. 이것이야말로 유저 지향의 개발이 필요한 것이다.

참고 문헌

- [1] 강성훈, "디지털 오디오와 심리음향", 전자공학회지 제12권 제5호, 19~29(1995. 5).
- [2] A. J. M. HOUTSMA, "Psychophysics and Modern Digital Audio," Philips J. Res. 47, 3~14(1992).
- [3] E. Zwicker, Psychoakustik, Springer-Verlag(1982).
- [4] 藤本, "디지털, ユソバAセクト. 카-세이트, 의開發に携わつて," 日本音響學會 49卷 4虎(1993).
- [5] 山本, "きこえない音," JAS JOURNAL(1995. 1).
- [6] T. Yamamoto, "Proposal of 96kHz sampling digital audio," The 97th AES Convention, San Francisco, Preprint No. 3884, Nov. (1994)
- [7] T. Holman, Psychoacoustics of Multi-channel Sound Systems for Advanced Television, 1992 NA HDTV World Conference Proceedings, PP. 13~20(1992)
- [8] G. Thelle, HDTV Sound Systems:How Many Channels?, AES 9th International Conference, pp. 217-232(1991)
- [9] D. G. Kirby, Experiences with Multichannel Sound for HDTV, Proceedings of IBC '92, pp.528-532(1992)
- [10] G. Carter, Multi-channel Sound for HDTV, AES 10th International Conference, pp.135~139(1992).
- [11] 岩宮眞一郎, 일본음향학술지 제 48권, 146~153(1992).
- [12] 岩宮眞一郎, 일본음향학술지 제 48권, 649-657(1992).

필자소개**강 성 훈**

- 1981. 광운대학교 전자과 공학사
- 1983. 연세대학교 음향전공 공학석사
- 1987. 고베대학교 음향전공 공학박사
- 1987. Yamaha 음향연구소 연구원
- 1988~1995. 한국전자통신연구소 음향연구실 실장
- 1996~현재 대전보건전문대학 방송제작기술과 음향 담당교수
- 1996~1997. 한국음향학회 학술위원장
- 1998~1999. 한국음향학회 교육연구위원장