

## 디지털 오디오와 心理音響

姜 成 勳

韓國電子通信研究所

### I. 서 론

아날로그 디스크나 테이프에서 콤팩트 디스크와 디지털 테이프 레코더(DAT), AM에서 FM과 디지털 방송으로 오디오 기술은 현저하게 발전되어 가고 있으며, 고품질의 오디오를 손쉽게 접할 수 있게 되었다. 이러한 배경에는 많은 기술적인 공헌이 있지만, 기술의 성과를 최종적으로 판단하는 것은 청취자이므로 아무리 물리적 특성이 좋다 하더라도 청취자가 좋은 소리로 느끼지 못하면, 그 오디오 기기는 좋다고 할 수 없다. 보다 좋은 음질을 추구하고, 포터블 기기는 보다 더 작게 하는 등 청취자의 욕구는 끝이 없지만, 청취자인 인간의 청각 특성과 관련지어면서 오디오 기술을 개발해야 하는 점은 말할 필요도 없다.

한편, 최근에는 영상 신호나 음성 신호의 대역을 압축하는 기술이 활발히 연구되고 있다.<sup>[1~11]</sup> 이것은 멀티미디어의 정보로서 화상이나 음악을 효율적으로 저장하거나 전송하기 위해서이다. 이러한 신호를 디지털화하여 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 기록하는데는 아날로그의 음향 신호에 비교하여 35배 이상의 막대한 주파수 대역이 필요하기 때문에, 한정된 전송대역이나 기록 용량내에서 멀티 채널화 및 콤팩트화를 실현하기 위해서는 막대한 디지털 음향신호를 대역 압축하여야 한다. 따라서 최근의 기술 개발의 관심은 청각 감각상에서 CD품질을 유지하면서 데이터를 압축하는 것이다. 그 대표적인 것이 미테랑 대통령의 제안에 의한 EUREKA의 프로젝트, 1986년 디지털 음성 전송 지상방송 시스템의 개발 프로젝트(EUREKA 147 DAB Research Project), MUSICAM(Masking-pattern Universal Sub-band Integrated Coding and Multiplexing), ISO-IEC/SC2/WG11(약칭, MPEG; Moving Picture Experts Group)의 제안에 의한

MUSICAM, ASPEC(Audio Spectral Perceptual Entrophy Coding), SB-ADPCM(Sub-band Adaptive Differential PCM), CCIR 제안에 저역 예측형 서브밴드 부호화, Dolby Stereo Digital (AC-3), Philips사의 DCC(Digital Compact Cassette)나 Sony의 MD(Mini Disk) 등이다.

이러한 오디오 신호 압축에 적극적으로 활용되고 있는 청각심리 특성은 어느 주파수 성분이 존재하고 있을 때, 그 근방의 낮은 레벨의 주파수 성분은 지각되지 않는 마스킹 현상과 최소 가청한계의 특성 등이다. 본 고에서는 최근에 디지털 오디오 개발에서 정보량을 압축하기 위하여 활용되고 있는 심리음향 현상 및 멀티채널 오디오와 중저음에 대한 심리음향에 대해서 개괄한다.

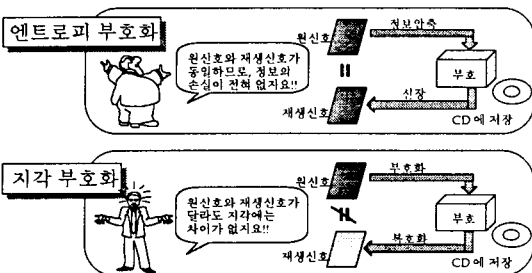
## II. 어떻게 데이터 압축이 가능한가

오디오 신호를 고능률로 압축하는데는 크게 두 가지 방법이 있다. 한가지는 신호 자체의 통계적인 특성을 이용하여 정보를 잃지 않고 효율 좋게 부호화하는 엔트로피 부호화(가변장 부호화) 방법이 있으며, 또 한가지는 청각 특성을 이용하여, 감도가 낮은 미세한 정보는 생략하여 데이터 양을 줄이는 방법으로 지각부호화(Perceptual Coding)방법이 있다. 이 두 가지 방법의 차이를 그림 1에 나타낸다. 그림 1과 같이 지각부호화에서는 재생된 신호는 원래 신호와 다르다. 그러나 인간의 청각으로

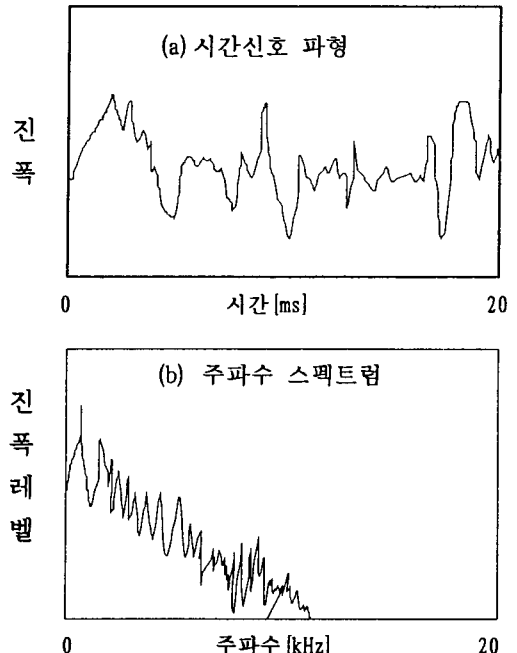
는 거의 차이를 감지할 수 없기 때문에 청각 감각 상으로는 같은 신호로 간주할 수 있다. 한편, 엔트로피 부호화의 경우에는 재생된 신호는 원래의 신호와 같으므로 정보의 손실이 전혀 없다. 보통은 이 두 가지 부호화 방법을 병용하여 사용하지만, MD나 MPEG 오디오에서는 지각 부호화가 커다란 역할을 한다.

### 1. 신호의 성질 이용

그림 2(a)는 음악 신호의 시간 신호 파형을 나타낸다. 신호 레벨이 크고, 이것을 그대로 A/D 변환하여 전송하면, 전송 용량은 768kbps(48kHz 16bit)가 필요하다. 그림 2(b)는 같은 신호를 FFT 분석기로 주파수 분석한 결과이다. 이 결과에서 보면, 신호의 주파수 성분의 대부분이 10kHz 이하에 집중되어 있는 것을 알 수 있다. 음악 신호 뿐만 아니라, 음성이나 자연계의 음향신호는 일반적으로 스펙트럼 성분이 저역에 집중되어 있는 경향이 있다. 따라서 음성신호를 주파수 분석하여 스펙트럼 성분이 존재하지 않은 주파수 대역은 전송



〈그림 1〉 엔트로피 부호화와 지각 부호화의 차이



〈그림 2〉 음악신호의 시간 신호 파형 및 주파수 스펙트럼

하지 않거나, 또는 주파수 스펙트럼의 레벨이 낮은 성분은 비트 할당을 적게하여 정보 전송량을 줄일 수 있다. 이렇게 함으로서 원래의 음이 가지고 있는 정보량을 1/2~1/3 정도까지 압축할 수 있다.

2. 청각 특성의 이용

어린 아이라도 엄마의 목소리는 확실히 식별할 수 있고, 약간의 음악 교육을 받으면 합주속에서 특정한 악기의 음을 구분할 수 있다. 또, 사람이 많이 모인 회의장에서 멀리 떨어져 있는 사람의 목소리를 주목하게 되면 명료하게 들을 수 있는 것을 경험하게 될 것이다. 이러한 현상을 칵테일 파티 효과(Cocktail Party Effect)라고 한다. 그리고, 콘서트 홀이나 오페라 하우스에서 음장의 울림을 나타내는 지표로서 잔향시간이 널리 사용되고 있는데, 잔향시간이 거의 같은 음장에서도 놀랄 정도로 울림의 차이를 느낄 수 있다. 이것은 건물의 형상이나 내장재에 의한 반사음, 특히 직접음에 계속되는 초기반사음의 시간구조와 도래 방향, 즉 공간 음향정보의 차이에 기인한 것이다.

또한, 인간은 단지 2개의 귀를 가지고 정밀하게 음원의 방향이나 거리를 지각하고 있다. 마이크로폰을 2개 사용하여도 전후나 상하 방향을 검출할 수 없다. 인간은 머리를 움직이어서 거리나 방향을 판정한다고 하지만, 머리를 움직이지 않고, 또 한 쪽 귀를 막아도 방향 판별은 어느 정도 가능하며, 이것은 인간의 청각 능력이 우수함을 나타내는 한 가지 특징이다. 컴퓨터가 급속히 발전되고 있는 상황이라고 해도, 기계가 인간과 같이 고도의 처리를 실시간으로 처리하기는 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 마스키(Maskee)의 주파수에 따라서 그 패턴이 달라진다. 이것은 최소 가청한계와 마스킹 곡선 이하의 소리는 양자화 잡음을 허용하여도 들리지 않으므로, 이러한 신호 성분은 전송하지 않아도 된다. 따라서 앞에서 기술한 음성신호의 성질에 더하여

이러한 청각 특성을 이용하면, 정보량을 1/4~1/6 정도까지 압축할 수 있다.

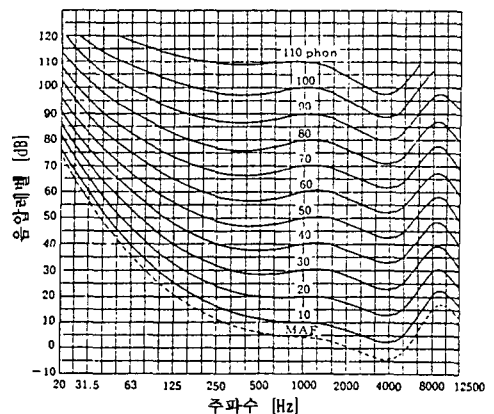
청각심리는 최소 가청한계, 마스킹 효과, 라우드니스, 임계 대역, 음색, 두 귀 효과 등 많은 테마가 있지만, 3절에서는 오디오 데이터 압축에 사용되고 있는 청각 성질만을 기술한다.<sup>[12]</sup>

III. 청각 심리

1. 최소 가청한계

청각은 20~20,000Hz의 음을 들을 수 있지만, 청각의 감도는 주파수에 대하여 일정하지 않다. 그림 3은 ISO-226에 표준화되어 있는 등감곡선을 나타낸다. 이 그림의 곡선은 각 주파수의 순음에 대하여 같은 크기(Loudness)로 들리는 음압레벨의 관계를 나타낸 것이다. 그림의 가장 아래 곡선은 최소 가청한계(Minimum Threshold)를 나타내며, 곡선이 밑에 있을수록 음압레벨이 낮아도 들리는 것을 의미한다. 일반적으로 저역의 감도는 낮고, 주파수가 높아질수록 감도가 상승하여 3~4kHz 부근에서 가장 감도가 좋은 것을 알 수 있다.

어떤 음에 대해서 정상적인 청력을 갖는 사람이



(그림 3) 등감 라우드니스 곡선(Equal loudness contours)

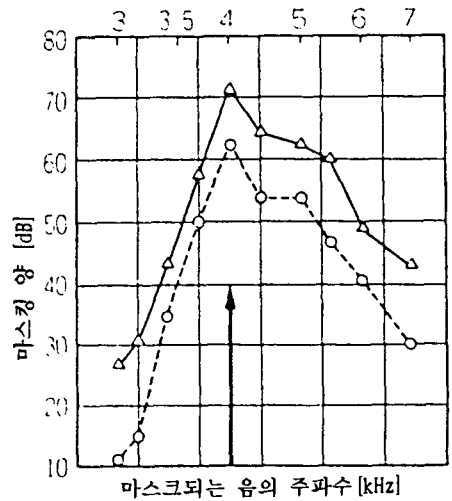
그 음과 같은 크기로 들린다고 판단한 1,000Hz의 순음의 음압레벨의 수치를 라우드니스 레벨(Loudness Level, 단위; Phon)이라 한다. 저역에 대한 청각의 감도와 고역에 대한 청각의 감도 차이는 음의 레벨이 높아질수록 적어진다. 낮은 레벨에서 스테레오를 청취하면 저역이 잘 들리지 않지만, 레벨을 올리면 저음이 잘 울리는 것은 이와 같은 청각의 주파수 특성 때문이다. 다시 말하면, 저역의 음은 고역의 음과 비교하여 음압레벨이 상승하면 급격히 커지며, 이것은 저역의 다이내믹 레인지가 좁다는 것을 의미한다. 오디오 앰프의 라우드니스 스위치는 이러한 청각 특성을 보상하기 위한 것이다.

한편, 그림 3에서 110phon 이상의 레벨에서 소리는 들리지만, 귀에 통증이 오며, 이렇게 큰 소리에 장시간 노출되면 일시적으로 역치가 상승하는 현상(Temporary Threshold Shift; TTS)이 생기며, 계속되면 영구적인 청력 손실(Permanent Threshold Shift; PTS)이 생기게 된다. 공장뿐만 아니라 디스코 클럽이나 록 콘서트 장소에서도 음압레벨이 110phon을 넘는 경우가 많으며, 더욱이 워크맨도 외부에 소리가 새어 나올 정도의 큰 소리이면 이러한 청각 장애 현상이 생긴다.<sup>[2]</sup>

라우드니스는 음의 세기에 대한 청각상의 성질로 정의되어 있으며, 다른 조건이 일정하면, 소리의 세기의 증가할수록 라우드니스도 증가한다. 라우드니스 단위는 sone으로, 1000Hz 40dB 음의 라우드니스를 1 sone으로 한 비율 척도이며, 2 sone의 음은 1 sone의 2배의 라우드니스가 된다는 것을 의미한다.

### 2. 마스크링 현상(Masking Effect)

마스크링은 ‘어떤 음에 대한 최소가청 한계가 다른 음의 존재에 의해 상승하는 현상’으로 정의되고, 그 상승치를 dB로 나타낸다. 2개의 음이 제시되는 조건에 따라서 동시 마스크링(Simultaneous Masking), 시차 마스크링(Temporal Masking), 두 귀사이의 마스크링(Contralateral Masking)으로 구분된다. 이러한 마스크링 현상의 연구는 오래전부터 연구되어 왔으나, 오디오 개발에는 MD나 DCC 개발시 정보량을 압축하기 위해서 도입되었다.



〈그림 4〉 4kHz 순음에 의한 동시 마스크링

#### (1) 동시 마스크링

마스크하는 음(Masker)과 마스크되는 음(Maskee)이 동시에 제시되는 경우를 동시 마스크링이라고 한다. 순음이 다른 주파수의 순음에 의해 마스크될 때의 마스크링량을 그림 4에 나타낸다. 마스크(여기에서는 화살표로 표시된 4kHz)에 대해서 마스크의 주파수가 높을수록 낮은 주파수에 의해 마스크 되기 쉽다.

#### (2) 시차 마스크링

2개의 음이 시간적으로 계속 제시되었을 때 생기는 마스크링 현상이다. 시간적으로 선행하는 음이 후속되는 음을 마스크하는 현상을 순방향 마스크링(Forward Masking 또는 Post-masking), 후속되는 음이 선행하는 음을 마스크하는 현상을 역방향 마스크링(Backward Masking 또는 Pre-masking)이라 한다. 전방향, 역방향 마스크링에서 마스크와 마스크가 시간적으로 떨어지면 마스크링량은 급속히 감소된다.

음악이나 음성은 시시각각 레벨과 주파수가 변화되고, 레벨이 큰 음 뒤에 레벨이 낮은 음이 계속되는 경우는 흔히 있는 일이다. 계속 마스크링은 일상생활에서도 음의 정보전달을 이해하는 데도 중요한 현상이다.

(3) 두 귀사이의 마스크

2개의 음을 좌우의 귀에 따로따로 들려 줄 경우에 생기는 마스크 현상을 말한다. 같은 쪽 귀 마스크(Ipsilateral Masking)과 비교하여 마스크량은 50dB 정도 낮다. 이것은 마스크하는 음이 骨傳導(Bone Conduction)에 의해 반대쪽 귀에 전달되고, 거기에서 동시 마스크가 발생하기 때문이다.

3. 임계 대역

청각계에 존재한다고 가정한 일종의 대역 필터이며, 라우드니스나 마스크 등에 있어서 그 양태가 급속히 변화되는 대역을 말한다. 임계대역의 개념은 1940년 Flecher에 의해 제안된 임계비의 개념으로부터 시작된다. Flecher는 순음의 대역 잡음에 의한 마스크 역치를 측정할 때 대역 잡음(중심 주파수는 신호의 순음과 같음)의 파워를 일정하게 유지하고, 대역폭을 변화시켜 순음이 가까스로 마스크될 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 Flecher는 청각계의 필터의 존재를 제시하였고, 그 후에 Zwicker는 체계적인 실험을 통하여 임계대역 개념을 명확히 하였다.

라우드니스 및 마스크에 대해서는 다음과 같은 현상이 나타난다.

(a)라우드니스 : 전체의 파워가 일정할 때 대역 잡음의 크기는 임계대역 내에서는 대역폭에 관계 없이 일정하지만, 임계대역 이상으로 넓어지면 증가한다.

(b)마스크 : 순음을 대역잡음으로 마스크할 경우에 순음의 파워와 순음의 주파수를 중심으로 임계대역내의 잡음의 파워가 같을 때, 순음은 가까스로 마스크되지만, 대역잡음외의 잡음은 마스크에는 별로 기여하지 않는다.

Zwicker는 임계대역이 중요한 역할을 하는 라우드니스 레벨의 단위인 'phon'을 도입한 Barkhausen의 이름을 사용하여 임계대역의 단위를 'Bark'라는 명칭을 사용하였다. 임계대역 폭은 150Hz에서 100Hz, 1kHz에서 160Hz, 4kHz에서 700Hz,

그리고, 10.5kHz에서 2.5kHz로서 주파수가 높아짐에 따라서 넓어진다. 이 임계대역은 저역을 제외하고 280Hz 이상에서는 1/3 옥타브 폭에 가깝다. 일반적으로 음향분석에서 1/3 옥타 필터를 사용하고 있는 이유는 이러한 임계대역의 개념 때문이다. 한편, DCC나 MPEG 오디오에서 사용되는 대역 분할 부호화는 이 임계대역을 기본으로 하는 처리 방식이다.

IV. 청각 심리의 적용 예

그림 3과 같이 인간이 들을 수 있는 주파수 대역은 20Hz~20,000Hz이고, 다이내믹 레인지는 거의 90dB 정도이다. 오디오 기기는 이 가청 범위의 음을 재생하는 것이 요구되고, 이러한 조건을 만족시키기 위해서는 1초간에 1.4Mbit의 정보량이 필요하다. 직경 64mm의 MD에서 이 요구를 만족시키면, 불과 10분 정도 밖에 녹음·재생을 할 수 없으며, CD와 같이 74분간 녹음하기 위해서는 정보량을 1/5 정도 압축해야 한다. DCC도 아날로그 카세트와 대응성을 갖기 위해서는 1/4로 정보량을 압축하여야 한다. 음질을 열화시키지 않고 이러한 조건을 만족시키기 위해서는 청각의 특성을 고려한 기법들이 활용되어야 한다.

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

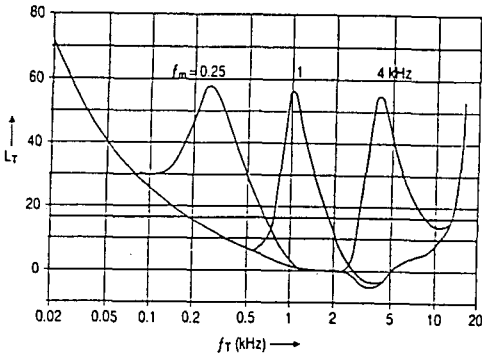
$$L_{an} = L_{sm} - 20 \log 2^n$$

$L_{an}$ : 양자화 잡음 레벨

$L_{sm}$ : 최대 레벨

n : bit 數

이 식으로부터 양자화 잡음은 16bit에서 최대 레벨보다 -96dB이고, 귀에는 들리지 않지만, 12bit에서는 -72dB, 8bit에서는 -48dB가 된다. 즉, 정보량을 압축하면, 양자화 잡음이 증가하여 그림 5와 같이 여러 군데에서 양자화 잡음 레벨이 신호



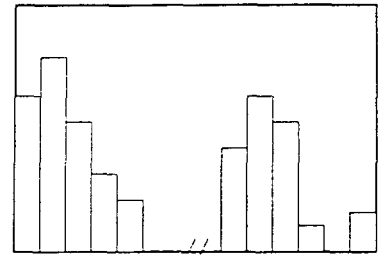
(그림 5) 협대역 잡음에 의한 마스킹 역치와 양자화 잡음 레벨(그림 속의 수평선)과의 관계

레벨보다 높아지는 곳이 생겨 양자화 잡음이 들리게 된다.

MD에서는 ATRAC(Adaptive Transform Acoustic Coding), DCC(Precision Adaptive Subband Coding)이라 불리는 방식이 각각 채용되고 있으며, 약간 방식은 다르지만, bit수를 줄여도 이러한 양자화 잡음이 들리지 않도록 하는 것과 신호 속에서도 다른 부분에 마스킹되어 들리지 않는 여분의 정보를 생략하는데 마스킹 이론을 이용하고 있다. 즉, 입력된 음원을 MD에서는 3개의 서브밴드로 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. 이러한 과정을 그림 6에 나타낸다. 또, 이러한 처리는 귀의 신호처리 속도와 같도록 해야 하며, MD에서는 11.6ms, DCC에서는 8ms마다 이루어지고 있다. 더우기 이러한 처리는 입력신호의 시간적·주파수적인 특징이 있으며, 동시 마스킹뿐만 아니라 과도적인 신호에 대해서는 시차 마스킹도 고려하여 처리한다.<sup>[13]</sup>

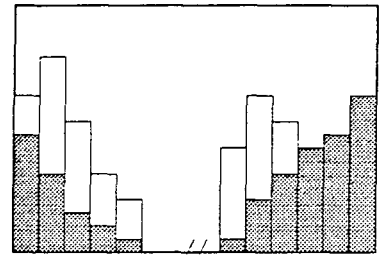
MPEG 오디오에서는 레이어I, 레이어II, 레이어III의 세 가지 모드가 있으며, 레벨이 높을수록 고품질 및 고압축률을 얻을 수 있지만, 그 반면에 하드웨어의 규모가 커진다. MPEG 오디오에서도 부호화 효율을 높이기 위해서 임계대역 등의 청각심

(1) 각 서브밴드의 음압 에너지



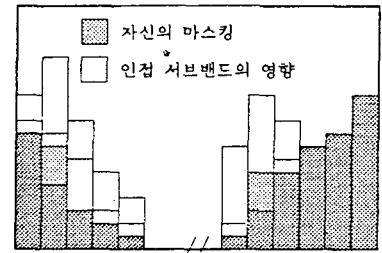
subband 0 1 2 3 4 ..... 26 27 28 29 30 31

(2) 최소 가청한계의 적용



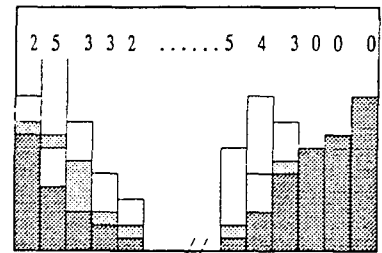
subband 0 1 2 3 4 ..... 26 27 28 29 30 31

(3) 마스킹 곡선의 적용



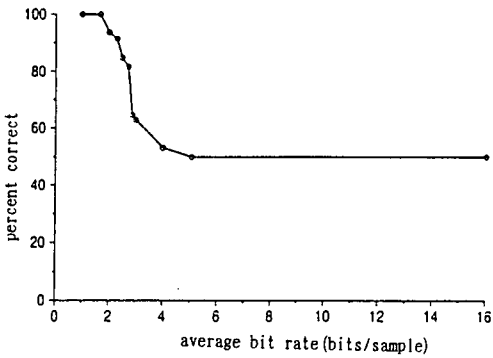
subband 0 1 2 3 4 ..... 26 27 28 29 30 31

(4) 양자화 비트의 배분



subband 0 1 2 3 4 ..... 26 27 28 29 30 31

(그림 6) 양자화 비트 결정 과정



(그림 7) DCC의 bit수와 정답 반응의 비율 관계

리를 이용하고 있다. 우선, 신호는 주파수 대역을 32개의 등간격의 밴드로 분할하여 각각의 신호를 원래의 샘플링 주파수의 1/32로 서브 샘플링하여 부호화한다. 그리고 시차 마스킹 효과의 범위에서 가장 부호화 효율이 높은 값으로서 32의 정수배인 384 샘플을 1프레임으로 하여, 각 서브밴드 12샘플, 32밴드로 변환하여 부호화한다. 한편, 2개의 채널을 조합하여 부호화할 경우에는, 부호화 효율을 한층 높이기 위해서 높은 주파수에서는 청각의 위상 감지 능력이 낮다는 사실을 이용하여, 높은 주파수에서는 신호를 모노로 전송하는 등의 결합 스테레오 코딩(Joint Stereo Coding) 방법을 사용하고 있다.<sup>[9]</sup>

이와 같이 정보량을 감소시킬 경우에 청감은 실제로 양자화 잡음을 지각하지 못하는가가 문제이다. Houtsma는 bit수를 여러가지로 변화시켜 처리한 음과 완전한 것과 비교하는 청감 실험결과를 발표하였다. 그림 7에 나타낸 바와 같이 변별한계는 평균적으로 2.5~3.0bit/sample이다.<sup>[7]</sup>

### V. 멀티채널 오디오의 심리음향

입체음향에 관한 연구는 1930년대 초 영국의 Blumlein과 미국의 Snow에 의해 시작되었으며, 상용적인 구현은 관객을 TV에 뺏기는 것을 막기

위한 방편으로 1950년대 미국 영화업자들에 의해 본격적으로 도입이 시작되었다. 입체음향은 모노포닉에 비해 크게 두 가지 중요한 효과, 즉 입체음상(Stereophonic imaging) 효과와 청각적 효과(Auditory impression)가 있다.<sup>[2]</sup> 입체 음상 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 만드는 것을 말한다. 청각적 효과란 입체 음상 효과에 비하여 복잡한 청각적인 효과로서 크게 공간적 효과(Spatial impression)와 효과음(Effect sound)으로써 구분된다. 공간적 효과에는 여러 방향으로 부터의 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 박수 소리나 웃음 소리, 빗 소리, 바람 소리 등을 현장감있게 제공하는 것을 말한다.<sup>[14~17]</sup>

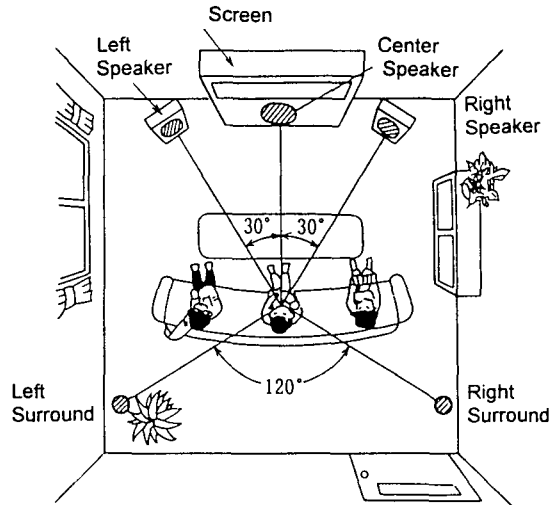
이상에서 설명한 입체음향의 효과적인 재생은 여러 음원들로 구성된 음장(Sound Field)을 어떻게 효과적으로 재생하는가에 달려 있다. 입체음향의 실제적인 구현은 다수의 스피커를 이용한 음의 재생을 통해 이루어지는데 입체 음상은 전면 스피커로 재생되고, 청각적 효과는 서라운드 스피커에 의해 재생된다. 현재까지 TV 방송이나 FM 방송 등 대부분의 분야에서는 두 채널 스테레오가 사용되어 왔지만, 고선명 TV 등에서는 여러가지 문제점이 있다. 기존 2채널 스테레오의 가장 큰 문제점은 방향감이 불안정(Directional unstable)하다는 점이다. 즉 2채널 스테레오에서는 두 스피커 사이의 중앙 위치에서 듣는 경우는 문제가 없지만, 중앙에서 조금만 벗어나도 선행 효과(Precedence effect)때문에 먼저 도달하는 스피커쪽에 음상이 정위되어 버린다. 그 동안 2채널 스테레오가 무리없이 사용되어 왔던 이유는 입체 음상 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 3~4명이상이 시청할 경우에 중앙에 위치한 사람을 제외하고는 나머지 사람들은 충분한 입체 음상을 느낄 수 없기 때문에, 고선명 TV의 원래 목적인 35mm 필름의 극장식 현장감을 실현하기 어렵다.

2채널 스테레오 재생이 가지고 있는 음상의 불안정함을 해결하기 위한 여러 가지 실험과 방법들이 연구되어 왔다.<sup>[14~17]</sup> 실험은 전면 스피커의 갯수와 위치를 변화시키면서 방향감의 오차를 측정하는 방법인데, 2채널에서 세 채널(좌, 우, 중앙)로 증가시키면 방향감이 안정되는 효과가 얻어지지만, 네 채널로 증가시킬 경우는 그다지 큰 효과가 없다는 점이 많은 실험들의 공통된 결과이다. 이들 실험을 종합해 보면 스피커의 수가 많을수록 입체 음상을 재현하는데 유리하지만, 가정에서 고선명 TV의 사용을 전제로 할 때 허용되는 방향성의 오차, 가격대 성능비, 전송 비용 등을 고려하면 대체로 세 개의 전면 스피커를 사용하는 것이 적당하다.

또한, 청각적 효과를 충분히 재현하는데 있어서도 2채널만으로는 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 더 올리기 위해서 초저음(Sub-woofer) 스피커의 사용도 고려되고 있다. 저주파수 대역에서는 청각의 특성상 상대적으로 높은 신호 레벨을 필요로 하며, 또한 무지향성이므로 전면에 별도로 한개의 서브우퍼 스피커



(그림 8) 멀티 채널 오디오의 스피커 배치

를 배치하는 것이 효과적이다. 이상을 종합하면, 고선명 TV를 멀티 채널 오디오의 스피커 구성은 그림 8과 같이 설치하는 것이 바람직하다.

## VI. 저음의 심리 효과

인간에게 쾌적한 환경은 콘서트 홀이나 자연에 둘러싸인 환경, 또는 모태안 등이다. 콘서트 홀 등 실제 음상에서의 현장감을 실현하려면, 음악신호만이 아니라, 수 Hz~100Hz 이하의 중저음 영역의 재생이 중요하다. 또한, 산속에서의 바람과 나무의 술렁거림, 파도가 밀려오는 소리, 계곡물이 졸졸 흐르는 소리 등, 모든 자연계의 음도 저음이 풍부하게 포함되어 있다. 마찬가지로 모태안의 환경도 20Hz에서 스펙트럼이 피크를 갖는 저음 우위의 환경인 것을 나타내고 있다. 그리고 이들 쾌적 환경은 모두 신체가 음으로 둘러싸인 음상이다.

오디오에 있어서도 저음은 여러 가지 의미를 가지며, 파이프 오르간이나 대편성 관현악단과 같이 저음이 많은 음악에는 필요 불가결하고, 더우기 소편성의 실내악에 있어서도 저역을 충분히 재생하



면 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.

현재의 A/V 장치에서도 중저음은 현장감을 재생하는데 커다란 역할을 한다. 특히 화면의 크기가 커질수록 저음과의 균형이 중요해진다. 100인치 화면 가득히 채트기가 날아가면, 그것에 알맞게 중저역이 재생되고 음압레벨도 상승해야 오디오와 비디오가 일치된다. 중저음 스피커(Sub-woofer 또는 Super-woofer)는 일반적으로 150Hz 이하의 주파수를 재생하고, 20~30Hz의 중저음까지도 재생할 수 있다.

한편, 중저음이 갖는 생리적인 현상은 다음과 같다. 일반적으로, 저음을 들을 경우 근육의 수축으로 고막의 임피던스가 증가하여 귀의 감도가 낮아진다. 그와 동시에 청각 대응으로서의 피부 감각과 진동 감각이 활동한다. 이와 같은 음을 촉각 음(Tactile Sound)이라고 하며, 전신 진동의 주파수 대역은 1~100Hz이다. 이와 같은 진동감을 저음감의 대응으로서 이용한 것으로 음악 신호에 맞추어 트랜스듀너로 인간의 피부와 뼈에 직접 국부적 진동을 주는 음향 장치도 상용화되어 있지만, 음은 본래 공기 진동이므로, 저음감을 기계 진동으로 대응하기 보다도 공기 전반에 의한 자연적인 저음의 전송으로 신체 전체의 진동감을 만들어 내는 것이 이상적이라고 할 수 있다.<sup>[14]</sup>

## VII. 들리지 않은 소리의 중요성<sup>[18]</sup>

이상에서 기술한 바와 같이 음향신호의 압축은 청각심리를 이용하고 있다. 즉, 음향신호의 압축은 각 시간 마다 음향신호의 스펙트럼을 분석하여 그

중에서 스펙트럼이 큰 성분만을 남기고, 이것에 의해 마스크되는 성분은 버리는 방식이다. 국제표준화 위원회에서는 시청 실험 결과를 통해서, 1/4 정도로 압축하여도 CD 음질과 구별할 수 없다는 결론을 내리고 있다. 그러나, 보다 중요한 점은 기준 음질로 하고 있는 CD 음질 자체가 실제 연주와는 상당히 다르다는 점이다. 화상의 압축 기술과 음향 신호의 압축기술을 비교하면, 화상의 경우에는 인간의 시각에 보이지 않은 곳은 주변의 밝기나 색을 얼버무리는 것이다. 이것에 대해 음성의 경우는 청각의 마스크 현상을 이용하는 것으로서, 성분 음이 적은 음을 버리는 것은 순간 순간의 음의 스펙트럼에 구멍을 뚫어버리는 것과 같다.

화상의 압축기술은 고주파 영역을 자르면 되지만, 이것을 음향신호의 압축기술로 대체하면, 음성신호의 고역 성분을 자르는 것에 해당한다. 그러나 이것은 CD에 녹음할때 대역을 22kHz로 제한하는 것으로서 이미 실행되고 있다. 따라서 이 이상으로 음향신호의 대역을 압축한다는 것은 마스크 효과와 같은 현상을 이용할 수 없는 것은 아닌가? 실제로 CD보다 1/4~1/5로 대역 압축한 음은 고주파 성분은 적고, 마치 넓은 초원에서 몇몇의 음악가가 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 마스크되어 버리므로, 이 성분들을 잘라 내버린다면 음장감이나 현장감이 재생되지 않는 것은 당연하다. 이것이 넓은 초원에서 연주하고 있는 것처럼 들리는 이유이다.

인간의 귀에는 마스크 효과가 있다고는 하지만, 인간은 레벨이 낮은 수 많은 성분음에 의해 악기의 섬세함이나 연주의 섬세한 뉘앙스 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 생각이 든다.

### VIII. 맺는말

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.<sup>[10]</sup>

한편, 오디오의 궁극적인 목적은 세계 콘서트 홀의 음을 그대로 리스닝룸에 가져오는 것이다. 건축상으로 구현할 수 없는 홀을 전기음향적으로 재생하면 어떠한 홀에서 들은 음보다도 좋은 음을 들을 수 있는 가능성도 생각할 수 있다. 이러한 연구도 인간의 청각심리를 적극적으로 활용하면 충분히 가능한 것이다.

### 참 고 문 헌

- [1] 김기수 “디지털 오디오에서의 신호처리 기술의 응용,” 한국음향학회 제9권 10호, 73 (1992)
- [2] 이준용, “MPEG-2오디오 압축 알고리즘,” 방송과 기술, 통권 제 40호, 80~88(1995)
- [3] 서양석 외, “MPEG-4 표준화 동향,” 정보통신기술, 제8권 제2호, 52~67(1994)
- [4] 홍진우, “MPEG 오디오 기술 분석,” 전자통신동향분석(1995)
- [5] 김연배 외, “인간의 청각심리를 이용한 디지털 오디오 압축 알고리즘, 대한전자공학회 추계종합학술대회논문집, 제16권 제2호, 901-904(1993)
- [6] 홍진우 외, DTV용 멀티채널 오디오 소스 코덱 개발, 제5회 통신정보학술대회(1995.5)
- [7] A.J.M. Houtsma, “Psychophysics and Modern Digital Audio,” *Philips J. Res.* 47, 3~14(1992).
- [8] D. Wiese and G. Stoll, “Bitrate Reduction of High Quality Audio Signals by Modelling the Ears Masking Thresholds,” *The 89th AES Convention*, LA, Preprint No. 2970, Sep. 21~25, (1990).
- [9] G. Stoll, G. Theile, S. Nielsen, “Extension of ISO/MPEG-Audio Layer II to Multi-Channel Coding: The Future Standard for Broadcasting, Telecommunication, and Multimedia Applications,” *The 94th AES Convention*, Berlin, Preprint No. 3550, Mar. 16~19(1993).
- [10] K. Brandenburg and G. Stoll, “The ISO/MPEG-Audio CODEC: A Generic Standard for Coding of High Quality Digital Audio,” *The 92th AES Convention*, Vienna, Preprint No. 2970, Mar. 24~27(1992).
- [11] B. Grill, J. Herre, K.H. Brandenburg, E. Eberlein, J. Koller, and J. Muller, “Improved MPEG-2 Audio Multi-Channel Encoding,” *The 96th AES Convention*, Amsterdam, Preprint No. 3865, Feb. 26~Mar. 1(1994).
- [12] E. Zwicker, *Psychoakustik*, Springer-Verlag(1982).
- [13] 藤本, “ドイツタルコンパクト.カセットの開発に携わって,” *日本音響學會誌* 49卷 4号 (1993).
- [14] T.Holman, *Psychoacoustics of Multi-channel Sound Systems for Advanced Television, 1992 NA HDTV World Conference Proceedings*, PP.13~20(1992).
- [15] G. Thelle, “HDTV Sound Systems: How Many Channels?,” *AES 9th International Conference*, pp.217~232(1991).
- [16] D. G. Kirby, “Experiences with Multichannel Sound for HDTV,” *Proceedings of IBC '92*, pp.528~532(1992).
- [17] G.Carter, “Multi-channel Sound for HDTV,” *AES 10th International Conference*, pp.135~139(1992).

- [18] 山本, “きこえない音,” *JAS Journal*(1995. 1).
- [19] T. Yamamoto, “Proposal of 96kHz sampling digital audio,” *The 97th AES Convention*, San Francisco, Preprint No. 3884, Nov. (1994).

저자 소개

姜 成 勳

- |             |                      |
|-------------|----------------------|
| 1981年 2月    | 광운대학교 전자과(B.S.)      |
| 1983年 2月    | 연세대학교 대학원 전자과(M.S.)  |
| 1987年 3月    | 고오베대학 대학원 전자과(Ph.D.) |
| 1987年 4月    | YAMAHA 음향연구실         |
| 1988年 2月~현재 | 한국전자통신연구소 음향통신연구실    |