

우루루 읽기 모임 <음향적 조건등색을 찾아서> 1주차 (2024.3.3)

Andy Farnell(2010). *Designing Sound*. MIT Press.

Ch.6 Psychoacoustic. Section 6.1 (p.77~93)

Section 6.1 소리의 지각 – p.77

심리 음향학은 진폭, 주파수 등 측정 가능한 파동의 물리적 특성을 소리와 라우드니스, 음높이 등 주관적인 현상에 대한 지각과 연관시킨다. 이는 소리의 심리학이라고 할 수 있다. 이번 장에서는 기술적인 부분도 있지만 인간적이고 감정적이며 문화적인 부분을 많이 다룰 것이다. 이러한 모든 측면을 사운드 디자인에 결합해야 큰 그림을 그릴 수 있다. 음향심리학 및 스트리밍, 범주화, 차폐와 같은 프로세스를 이해하면 소리의 특성을 효율적으로 정의하기가 더 쉬워진다. 소리 인지는 청각에서 의식적인 이해로 이동하는 계층으로 볼 수 있다. 이 체계에서 가장 낮은 수준인 물리적인 것부터 시작해 보자. 심리음향학의 하위 집합인 *물리음향학*이 바로 그것이다.

파동은 종이나 스피커처럼 실제 물리적으로 진동하는 물체에서 나오지만 아직은 소리가 아니다. 엄밀히 말하면 소리는 지각적 경험이며, 이를 경험하는 마음이 없으면 진동하는 파동만 존재할 뿐이다. 파동은 고막을 진동시키고 뇌가 감각을 인식하게 한다. 이 감각은 실제로 귀에만 국한되지 않는다. 우리는 신체의 여러 부위에서 1Hz에서 20Hz 사이의 저주파 소리를 느낄 수 있다. 우리는 때때로 이러한 주파수를 '소리'라고 부르지 않지만, 이는 소리를 발생시키는 동일한 물리적 현상과 경험의 일부이다. 때때로 우리는 소리의 시각적 표현을 보기도 한다. 창문을 흔들거나 드럼통의 모래를 흔들면 소리는 우리의 눈에 보인다. 폭발과 같이 높은 강도로 발생하는 소리는 물체를 쓰러뜨리거나 산산조각 낼 수 있다.

귀

이 시점에서 귀의 다이어그램을 보여주는 것이 일반적일 것이다. 생물학 수업으로부터 귀가 어떻게 작동하는지 기억하고 있을지, 전통을 깨고 설명해 보겠다. 귀의 익숙한 모양인 바깥 부분, 즉 *귓바퀴(pinna)*, *외이도(meatus)* 및 *엽(葉, lobe, 고실)*이 있다. 사람의 경우 귓바퀴는 소리의 위치를 찾기 위해 부분적으로는 증폭기로, 또 부분적으로는 필터 역할을 한다. 외이도는 주로 외이와 내이를 연결하고 분리하는 역할을 하지만, 약 2kHz의 공명 공간으로 음성을 증폭하는 데 유용한 역할을 하기도 한다. 엽이 정확히 어떤 역할을 하는지는 밝혀지지 않았지만, 특정 주파수에서 귓바퀴의 댐퍼 역할을 하는 것으로 보이며 균형 감각에 중요한

역할을 할 수 있을 수 있다. 고막(*eardrum, tympanic membrane*)은 중이에서 기압파를 진동으로 변환하는 주요 변환기 역할을 한다. 그 다음 달팽이관(*cochlea*)으로 소리를 전달하는 내이(*malleus*), 내이골(*incus*), 등골(*stapes*)의 내부 작동, 즉 이소골(*ossicles*)이 있다. 이소골은 고막의 임피던스를 달팽이관에 포함된 액체와 일치시키고 지렛대를 통해 추가 증폭을 유도한다. 달팽이관은 점점 가늘어지는 구멍이다. 점점 가늘어지기 때문에 서로 다른 주파수가 서로 다른 위치에서 공명을 일으킨다. 그 내부는 진동을 신경 신호로 바꾸는 기저막(*基底膜, basilar membrane*)이라고 하는 작은 털로 이루어진 융단이 덮여 있다. 또한 내부에 작은 털이 자리하고 있는 바깥쪽 털이 있다. 이 털은 달팽이관을 조율하기 위해 위아래로 움직이며 주의력과 소리 선택에 중요한 역할을 한다.

인간의 가청 주파수 범위 – p.78

약 20Hz 이상의 소리에 대해 귀는 연령에 따라 10kHz에서 20kHz까지 들을 수 있다. 인간의 청각 범위를 아는 것이 중요한 이유는 합성적으로 분석하거나 처리해야 하는 주파수의 범위가 제한되기 때문이다. 말이나 음악과 같은 자연스러운 소리의 대부분은 이 범위의 중간인 300Hz에서 3kHz 사이의 주파수를 가지고 있다. 하지만 실제 소리의 배음은 이 범위의 너머까지 확장된다. 자연의 소리에는 대역 제한이 없다. 진흙이 튀는 소리나 나무 벽돌이 부딪히는 소리처럼 걸보기에는 둔한 소리도 청각 범위를 통과하는 약한 요소가 있다.

비선형성

기저막은 실제로 선형적인 시스템이 아니다. 귀 자체에 의해 발생하는 왜곡은 라우드니스 지각의 비선형성을 초래한다. 기저막은 나중에 주의력의 문제와 관련하여 살펴볼 것처럼, 의도적으로 비선형적으로 만드는 피드백이 있는 복잡한 사이버네틱 시스템이다. 외유모세포(*Outer hair cell*)는 입력에 반응하여 특정 주파수를 강화하거나 약화시킨다. 외유모 세포의 제어에 의한 비선형적 청각을 갖는 한 가지 이유는 입력 에너지를 평균화하여 동적 해상도를 향상시킬 수 있기 때문이다. 다른 주파수의 기존 톤에 두 번째 톤이 추가되면 첫 번째 톤의 볼륨이 명백하게 감소하는데, 이를 *두 톤 억제(two tone suppression)*라고 한다. 전체 시스템의

¹ [옮긴이주] 게임 제작에서 이전 프레임의 GPU 작업 부하량 따라 일시적으로 해상도를 떨어뜨려 목표 프레임 속도를 유지하는 해상도 조절 메커니즘. 예를 들어 화면에 오브젝트가 너무 많을 경우, 또는 프레임에 갑자기 비싼 이펙트가 들어오는 경우 일시적으로 해상도를 낮출 수 있다.

비선형 주파수 응답은 배음 왜곡의 한 형태인 *큐빅 차이 톤(cubic difference tones)*²을 유발한다.

청각의 역치

우리가 들을 수 있는 가장 조용한 소리는 무엇일까? 곧 알게 되겠지만, 가장 조용하게 들을 수 있는 소리는 주파수에 따라 달라진다. 실제로 진폭, 주파수, 스펙트럼의 효과는 독립적인 것이 아니라 서로 결합하여 복잡한 지각의 풍경을 만들어낸다. 1kHz를 기준으로 삼았을 때, $1 \times 10^{-12} \text{W/m}^2$ 가 인지할 수 있는 가장 조용한 소리로 정의된다. 이는 매우 적은 양의 에너지이다. 실제로 우리는 열을 백색 소음으로 들을 수 있다. 사람의 귀는 매우 민감하고 적응력이 뛰어난 변환기이기 때문에, 특정 조건에서는 고막에 대한 공기 분자의 브라운 운동이 들릴 수 있다. 절대 침묵(생명이 없는 화산섬이나 특수 제작된 방에서만 경험할 수 있는 당황스러운 상황)에 장시간 노출되면 사람들은 희미한 쉿하는 소리가 들린다고 보고한다. 이것은 신경학적 인공물이 아니라 무작위적인 열 소음이다. 흥미롭게도 눈에 보이는 가장 작은 강도는 빛의 한 광자로 여겨지기 때문에, 우리는 기본적으로 진화에 의해 물리적 세계의 한계까지 인식하도록 조정되어 있다.

딱 알아차릴 만큼의 차이 – p.79

무언가를 지각하려면 우리는 그것을 *알아차려야* 한다. 우리가 측정하고자 하는 소리의 측면이 무엇이든 “딱 알아차릴 만큼의 차이”(just noticeable difference, *JND*)는 지각적 효과를 일으키는 가장 작은 변화이다. 이 측정은 심리음향 실험에서 사용되며, 항상 다른 값에 상대적이다. 그렇다면 100Hz 톤에서 우리가 알아차릴 수 있는 가장 작은 주파수 편차는 어느 정도일까? *JND*는 진폭 변화와 관련하여 가장 자주 보여진다.

국소화

두 눈이 입체 시각을 통해 피사계 심도를 인식할 수 있듯이, 한 쌍의 귀는 수신된 신호의 차이에 따라 음원의 위치를 파악할 수 있게 한다. 간단히 말해, 네 가지 요소가 고려된다. 두 개의 각도(일반적으로 도 단위로 표시)는 소리의 방향을 지정하며, 방위각은 소리의 위치를 우리 주변의 원에 배치하고 앙각(仰角)은 소리의 위 또는 아래 위치를 측정한다. 또한 물체의 거리와 크기도 이 단계의 일부로 인식된다. 작은 물체는 포인트에서 방출되는 것처럼 보이고, 큰 물체는 체적 범위에서 방출되는 것처럼 보인다. 일반적으로 날카로운 어택을

² 외부 유모세포가 코르티 기관의 감도와 조율을 향상시키는 피드백 루프에서 발생하는 것으로 생각되는 이음향방사. 이음향방사(otoacoustic emission, OAE)는 달팽이관의 외유모세포 활동에 의해 발생하는 에너지로, 외부의 소리가 대뇌청각중추까지 전도되는 방향과는 반대로 달팽이관에서 발생하여 외이도 쪽으로 누설되는 에너지(소리)를 뜻한다.

가진 고주파 소리가 부드러운 어택을 가진 저주파 소리보다 더 잘 국소화된다는 것이 일반적인 규칙이다. 소리의 시작 시간, 즉 일시적인 어택의 첫 몇 밀리초 동안의 소리가 가장 잘 국소화된다. 소리가 지속되면 그 위치를 찾기가 더 어려워진다. 또 다른 규칙은 반사가 많은 좁은 실내보다는 실외의 자유로운 공간에서 소리를 더 잘 찾아낸다는 것이다. 마지막으로 한 가지 일반적인 규칙은 고개를 이리저리 움직여 소리를 반복하여 들을 수 있을 때 위치를 더 잘 인식한다는 것이다. 고도를 더 잘 인지하기 위해 고개를 기울이거나 소리를 향하도록 고개를 돌리는 것은 정확한 위치 파악을 위해 중요한 행동이다. 여기에는 두 가지 현상이 중요한 역할을 하는 것으로 알려져 있다: 양이간 시차(*interaural time difference*)와 양이간 강도차(*interaural intensity difference*). 양이간이라는 것은 "두 귀 사이"를 의미하므로 첫 번째는 각 귀에 도달하는 상대적인 진폭을 측정하는 것이다(예: 스테레오 시스템에서 *패닝* 동작). 두 번째는 전파에 대해 우리가 이미 알고 있는 사실, 즉 소리가 공간에서 거리를 이동하는 데는 유한한 시간이 걸린다는 사실에 기반한다. 두 가지 모두 머리 전달 함수³로 알려진 모델에서 결합된다.

양이간 시차(Interaural Time Difference)

양이간 시간차(Interaural Time Difference, ITD)는 동일한 소리가 각 귀에 도착하는 시간차를 말한다. 간단히 말해, 소리가 왼쪽 귀보다 오른쪽 귀에 먼저 도착하면 소리의 근원이 오른쪽 필드에 있을 가능성이 높다. ITD는 관련된 파장 때문에 700Hz 이하의 주파수에서 가장 효과적이다. 뇌는 음원에 대해 교차 상관 관계를 수행하는 것처럼 보이기 때문에, 전체 파장 이상이 컷속 거리(약 15~25cm)에 들어가면 모호함이 발생한다. 이는 ITD가 작동을 멈추는 1.5kHz 이상에서 발생한다. 머리의 반경이 r 이고 θ 가 중간(코의 방향)으로부터의 각도이고, c 가 음속일 경우

$$ITD = r \frac{(\theta + \sin(\theta))}{c} \quad (6.1)$$

이미 ITD가 있고 각도를 알고 싶은 경우, 수식을 단순화하여 재구성하면 다음과 같다.

$$\sin^{-1}\theta = \frac{c \times ITD}{2r} \quad (6.2)$$

³ [옮긴이주] 머리 전달 함수(Head transfer function)란 동일한 소리를 전방위에서 발생시켜 방향에 따른 주파수 반응을 측정, 3차원 함수로 정리한 것을 말한다.

양이간 강도 차이(Interaural Intensity Difference) – p.80

간단히 말해, 소리가 왼쪽 귀보다 오른쪽 귀에서 더 크게 들리면 음원이 오른쪽 필드에 있을 가능성이 높다. 양쪽 귀만으로는 양이간 강도 차이(Interaural Intensity Difference, IID)를 사용한 정확한 음원의 위치 파악이 어렵다. 동일한 방위각을 가진 곡선의 한 지점에서 나오는 음원은 동일한 양이간 강도 차이를 산출한다. 따라서 이 정보는 가능한 위치를 제한하지만 전체 그림의 한 부분일 뿐이다. 절대적인 강도 차이는 실제로는 다소 작기 때문에 머리가 제공하는 폐색(occlusion), 즉 머리가 드리우는 *음향 음영(acoustic shadow)*이 특히 고주파수에서 중요한 역할을 한다. 국소화를 위해서는 두 가지 정보를 모두 사용하고 사용 가능한 데이터에서 최선의 추측을 해야 한다. IID는 700Hz 이상에서 유효하지만, ITD 범위와 상호 보완적인 이유로 1.5kHz 이상에서 제대로 작동한다. IID 신호가 ITD 신호와 모순되는 경우, 주파수가 주로 1.5kHz 이상이면 ITD의 정보가 우세하다.

머리 반응 전달 함수

IID와 ITD 모두 모호한 정보를 생성하기 때문에, 추가 정보가 없으면 우리 뇌는 고도⁴를 인식하고 앞뒤 혼동⁵을 피하기 위해 추가적인 정보를 사용한다. 이때 코가 중요한 역할을 한다. 머리의 한쪽에 장애물이 있으면 주변의 소리 경로가 대칭이 아니게 되어 소리의 앞뒤 출처를 구분할 수 없다. 각 귀의 컷바퀴는 우리가 보고 있는 방향의 신호에 유리한 증폭기 역할을 하며, 뒤에서 나는 소리는 그 방향을 통과해야 하므로 앞뒤 구분이 더 어려워진다. 코와 광대뼈는 앞쪽에서 귀 쪽으로 소리를 쓸어내는 역할을 하며, 부드럽고 살이 많기 때문에 일부 소리를 흡수하여 앞쪽에서 나는 소리임을 알 수 있도록 도와준다. 또한 머리카락의 감쇠 효과와 목과 어깨의 위치도 소리의 방향 감각에 영향을 미친다. 이러한 현상은 회절 및 흡수 효과에 의존하기 때문에 주파수에 민감하다. 고주파는 저주파만큼 머리 주변으로 이동하지 않으므로 반대쪽에서 덜 들린다. 2kHz 전후의 소리는 방향이 혼동될 가능성이 더 높다.

거리

⁴ [옮긴이주] 방위각(Azimuthal Angle)은 기준 방향에서 좌우 방향의 각도를 뜻한다. 반면 양각/고도각 (Elevation Angle)은 지면에서 위아래 방향으로의 각도, 일반적으로는 올려보는 각도를 칭한다.

⁵ [옮긴이주] 앞뒤 혼동(front-back confusion)이란 어떤 소리가 앞쪽에 위치하는지 뒤쪽에 위치하는지 판단하기 쉽지 않은 현상을 뜻한다.

우리는 출처를 알지 못하면 거리를 측정하는 데 능숙하지 않다. 목소리나 자동차 소리와 같은 익숙한 소리는 그 크기에 대한 우리의 내부적 측정이 있기 때문에 비교적 잘 판단할 수 있다. 하지만 낯선 소리는 가까이서 들었을 때 조용한 소리인지, 멀리서 들었을 때 큰 소리인지 구분하기가 어렵다. 우리가 하는 일은 소리의 스펙트럼에서 소리에 대한 몇 가지 가정을 하는 것이다. 고주파는 거리가 멀어질수록 감소되기 때문에, 만약 고주파 성분이 있어야 하는 어떤 소리의 프로파일에 그 부분이 없거나 조용하다면 우리는 그 소리를 멀리 떨어져 있는 것으로 인식할 것이다. 또 다른 심리음향학적 과정은 환경을 사용하여 거리를 판단하는 것이다. 방과 같은 특정 환경에 적응되어 있는 경우, 직접 소리와 반사 소리의 비율을 기준으로 삼는다.

음원 식별 – p.81

서로 연관된 음원이 주어진 경우, 가능한 가장 가까운 음원을 가장 큰 소리로 지정한다. 같은 방에 두 대의 라디오가 같은 방송국에 튜닝되어 있다면 다른 라디오보다 조용하더라도 가장 가까운 라디오에서 소리가 나는 것처럼 보인다. 이 효과는 가상 모델링에서 중복된 상관관계가 있는 음원을 배제하고, 가장 가까운 하나의 방출 음원에 출처를 포함시키는 방법으로 사용할 수 있다. 두 개의 라우드스피커로 스테레오 필드를 구성할 때도 동일한 심리 음향 프로세스가 작동한다. 같은 거리에 있는 두 소스 사이에 앉으면, 왼쪽과 오른쪽의 소리가 따로 들리지 않고 가운데에서 나는 것 같은 소리가 들린다.

라우드니스에 대한 인식

소리의 강도는 귀가 받아들이는 에너지의 객관적인 척도이다. 그것은 절대 음압 레벨 또는 단위 면적당 출력의 음 강도 레벨 측면에서 진폭에 비례한다. 소리의 *라우드니스(loudness)*는 주파수 및 기타 요인에 따라 달라지는 주관적인 값이다. 라우드니스에 대한 우리의 인식은 음압 레벨이나 음 강도 레벨과 동일하지 않아 혼란스러울 수 있다.⁶ 라우드니스는 아래에 설명될 *폰(phons)*과 관계가 있는 또 다른 상대 비율인 *손(sones)* 단위로 측정된다. 1 손은 40dB SPL⁷에서 1kHz 사인파의 라우드니스로 정의된다. 10dB SPL이 증가하면 1음은 2배, 20dB SPL은 4배, 40dB SPL은 16배 더 큰 것으로 인식된다.

⁶ [원주 1] 오디오 엔지니어링과 심리음향학의 인터페이스에서 스케일과 단위의 확산은 노련한 전문가와 학자들에게도 혼란스러운 것이다.

⁷ [옮긴이주] 음압 레벨(Sound Pressure Level, SPL)은 소리를 생성하는 기압의 변화를 나타내는 데시벨 dB로 표시되는 음량을 뜻한다.

라우드니스 스케일 및 가중치

어떤 주파수는 다른 주파수보다 우리 귀에 더 크게 들린다. 중간 음역대의 소리는 같은 압력 수준의 저주파 및 고주파 소리보다 더 크게 느껴지는데, 이는 음성을 가장 중요한 음역대로 선택하는 진화적 영향을 고려할 때 놀라운 일이 아니다. 과거에는 플레처-먼슨 곡선과 로빈슨-대드슨 곡선 등 여러 가지 “등청감” 곡선이 개발되었다. 가장 최근의 DIN/ISO 225 곡선은 이전 곡선의 장점을 결합하고 일부 오류를 제거했다. 물론 절대적으로 객관적인 등청감 곡선이라는 것은 없으며, 대부분의 사람들에게 잘 맞는 곡선에 대한 최상의 근사치일 뿐이다. DIN/ISO 226 곡선은 1kHz에서 1폰이 1dB SPL과 같도록 폰을 지정한다. 이는 단위가 아니라 특정 주파수의 순음⁸에 대해 주관적으로 측정된 음량에 대한 윤곽선 또는 데이터 포인트 집합이다. 그림 6.1에서 80폰-dB(ISO226)에 대한 곡선을 볼 수 있다.

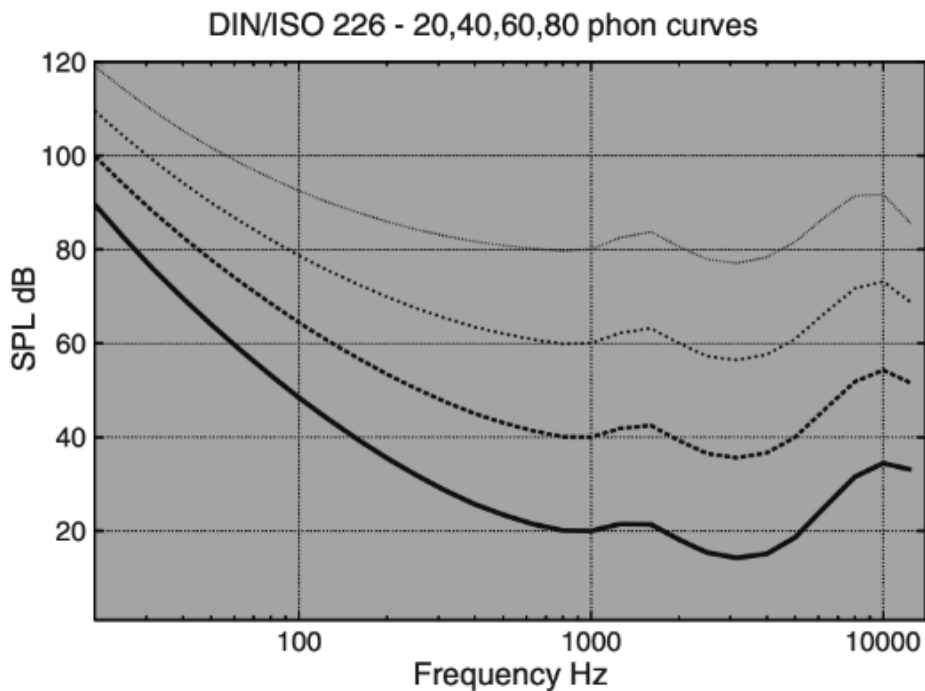


그림 6.1

20폰, 40폰, 60폰, 80폰에 대한 등청감 곡선.

1kHz에서는 정확히 80dB SPL이지만 양쪽 모두 달라진다. 20dB(ISO226)에서 80dB(ISO226) 사이의 다른 곡선은 모양에 약간의 변화를 보이는데, 신호가 클수록 약간 더 평평해진다.

⁸ [옮긴이주] 순음(純音, pure tone)은 배음이 포함되지 않은 순수한 음이다.

소음 측정기는 "가중치 필터"를 사용하여 주파수 의존성을 보정하고 실제로 듣는 소리에 더 적합한 수치를 생성한다. 모든 표준과 마찬가지로 선택의 폭이 상당히 넓으며 A, B, C, ... 유형이 있다. 환경 소음 측정에 사용되는 A형 가중치 척도, 그리고 BBC에서 제공하는 ITU-R 468 척도(B형 척도)가 가장 오래 지속되어 온 두 가지 유형이다. B형은 노이즈 성분이 포함된 음악과 음성에 더 효과적이기 때문에 방송 및 마스터링에 사용된다. 또한 B형은 피크 에너지가 크지만 주관적으로 더 조용한 일시적인 사건에 대한 인식을 설명하는 *소포메트릭 가중치(psophometric weighting)*를 사용하는 반면, A형 가중치는 일반 RMS 평균¹⁰을 사용한다. 소리 크기 측정값이 제공될 때 괄호 안에 가중치를 지정하는 단위가 표시될 수 있으므로, 100dB(A) 또는 100dB(ITU)로 표시될 수 있다.

지속시간 및 라우드니스 – p.82

앞서 말했듯 짧은 소리는 최대 강도가 크더라도 귀를 오랫동안 자극하지 않기 때문에 더 조용하게 느껴질 수 있다. 짧은 소리의 음량에 대한 인식은 지속 시간에 따라 증가하므로, 시간이 지남에 따라 전달되는 총 에너지가 증가하면서 힘들어 통합된다. *시간적 통합*이라고 하는 이 효과는 사운드 디자인에서 총소리나 짧은 충격을 몇 밀리초 늘려서 더 뚜렷한 볼륨을 만드는 데 사용할 수 있다. 이 효과는 최대 약 200ms의 지속 시간에 효과적이다.

피로 – p.83

순음으로 이루어진 일정한 소리는 우리가 거기에 얼마간 노출된 이후 더 조용하게 들리기 시작한다. 이는 절대 강도(Bekesey)¹¹와 중단 여부에 따라 다르지만, 일정한 톤에 약 1분가량 노출되면 주관적으로 절반 정도

⁹ [옮긴이주] 국제 전신 전화 자문 위원회(CCITT)의 전신인 국제 전화 자문 위원회(CCIF)가 소음 측정 세트 또는 흡포계에 사용하기 위해 제정한 소음 가중치.

¹⁰ [옮긴이주] 제곱평균제곱근(root mean square; rms) 혹은 이차평균(quadratic mean)은 변화하는 값의 크기에 대한 통계적 척도로, 각 샘플들의 레벨값을 제곱해서(Square) 평균을 구하고(Mean) 그렇게 얻어진 평균에 제곱한 것을 상쇄시키기 위해 루트를 씌워서(Root) 얻어진 값이다. 전자적으로 구성되는 샘플의 볼륨은 +와 -를 교대로 갖는 교류 성분으로 구성되어있기 때문에, 일반적인 평균을 구하는 식으로는 값이 0에 수렴하게 되는 문제를 갖는다. 따라서 제곱을 해서 모두 양수로 만들어 평균을 구한다.

¹¹ [옮긴이주] 베케시 청력도 측정(Bekesy audiometry)은 특정한 자극을 귀나 청각에 관계하는 곳에 가해서 청력을 측정하는 방법이다.

더 조용하게 느껴진다. 이것이 달팽이관에서 발생하는 것인지 아니면 주의력의 기능인지는 아직 명확히 밝혀지지 않았는데, 짧은 중단으로도 초기의 라우드니스를 회복 할 수 있고, 저주파에서 변조되는 "흥미로운"소리는 우리를 덜 피로하게하여 주관적으로 큰 소리로 유지되기 때문이다.

라우드니스의 변화

지각할 수 있는 가장 작은 강도의 변화는 순음의 경우 주파수의 함수이며, 실제 음량의 함수이기도 하다. 음량 변화에 대한 민감도는 수백 헤르츠의 저주파수 범위에서 가장 낮다. 주파수가 1kHz 이상으로 증가함에 따라 진폭 변화를 구분하는 능력이 훨씬 좋아지고, 동일한 음량과 유사한 곡선을 따라 보컬 범위에서 가장 잘 구분할 수 있으며 약 4kHz 이상에서 그 능력은 다시 떨어진다. 또한 소리가 크지만 너무 크지 않을 때 음량의 변화를 더 잘 들을 수 있으므로, 적절한 볼륨으로 믹싱해야 할 것이다. 음량을 가장 잘 구분할 수 있는 범위는 일반적인 대화 소리 바로 위의 범위인 약 60dB SPL~70dB SPL이다. 동일한 데이터를 해석하는 또 다른 방법은 동일한 주파수의 기존 톤이 있을 때, 해당 톤이 별도의 소리로 들리기 위해 필요한 최소 레벨이 얼마인가라는 질문을 던지는 것이다. 이를 통해 *강도 차폐(intensity masking)*을 위한 JND를 얻을 수 있다.

주파수의 지각

*강도*와 *라우드니스*가 절대적이고 지각적인 용어인 것처럼, *주파수*에 대한 심리적 대응 개념은 *음고(pitch)*다. 음고의 인식에 대해서는 여러 가지 모델이 있지만, 일반적인 소리 합성의 경우 음악적 용도로는 그다지 중요하지 않다. 음악적 음고는 생각만큼 간단하지 않은데 이것이 음악을 흥미롭게 만든다. 소리에 관한 일반적인 목적에서는 절대 주파수가 더 중요한 역할을 하는, 덜 정교한 음고 모델에 관심을 갖는 경향이 있다.

순음의 식별

우리의 *분해능(resolving power, 해상력)*은 가깝지만 다른 주파수를 지닌, 동시에 들려오는 두 개의 음색을 분리해서 들을 수 있는 능력이다. 두 음이 너무 가까우면 하나의 음으로 융합된다. 두 음이 번갈아 재생될 때는 다르다. 시간 경과에 따른 주파수 변화 사이의 *차분 임계값(differential threshold)*은 동시 주파수보다 작다. 이것은 차폐(masking)의 개념을 소개하며 곧 살펴볼 것이다.

임계 대역

임계 대역(critical band)¹² 모델은 달팽이관을 대역 간에 유한한 해상도를 지닌 스펙트럼 분석기¹³로 간주하여 만들어졌다. 달팽이관에는 수천 개의 작은 털이 있다. 각각의 털은 달팽이관이 조율되는 영역에서 작은 범위의 주파수에 의해 활성화된다. 이 위치 정보는 자극이 발생하는 기저막의 위치와 관련이 있으므로 이를 *위치 정보*라고 부르며, 머리 외부의 음원에 관한 국소화와 혼동해서는 안된다. 이것은 그래픽 이퀄라이저처럼 각각 1/3 옥타브씩 약 24개의 임계 대역으로 그룹화된다. 음악적 음정과 다른 인식 주파수에 대한 또 다른 척도를 바크 척도¹⁴라고 한다. 바크 주파수를 헤르츠 단위로 구하기 위해 다음을 사용할 수 있다:

$$Bark = 13 \tan^{-1} 0.00076f + 3.5 \tan^{-1} 1/(f/7500)^2 \quad (6.3)$$

이렇게 하면 1에서 24 사이의 숫자가 생성되므로, 인간의 평균 청력 범위인 20Hz~15kHz에 걸쳐 24개의 임계 대역을 제공한다.

범위 – p.84

(모든 임계 대역을 포함하여) 많은 주파수 범위는 소리와 음악의 다양한 분야에서 보다 일반적인 언어로 표현된다. 그림 6.2의 상단에 표시된 가장 오래되고 단순한 구분은 기본 톤 컨트롤에서 볼 수 있는 *저음(base)*, *중음(mid-range)*, 그리고 *고음 대역(treble)*으로 나뉜다. 사운드 프로듀서들은 오랫동안 각 중요한 청각 대역을 식별하는 데 그치지 않고 보컬과 악기에 대해 이야기할 때 유용한 방식으로 오디오 스펙트럼의 영역을 묶는 보다 정교한 이름 체계를 사용해 왔다. *서브(sub)*의 범위를 실제 악기가 차지하는 경우는 거의 없으며, 영화관이나 댄스 클럽 PA 시스템에서 천둥이나 기타 큰 효과음에 깊이를 더하는 데 사용되는 중앙 채널 또는

¹² [옮긴이주] 1933년 Harvey Fletcher 에 의해 소개되고 1940년에 개선된 심리음향학의 개념으로, 달팽이관에 의해 생성된 "청각 필터"의 주파수 대역폭을 설명한다. 두 번째 톤이 청각적 차폐에 의해 첫 번째 톤의 인식을 방해하는 주파수 대역을 뜻한다. 주파수의 차이가 임계대역보다 작을 때 청각이 주파수를 잘 분석하지 못하는 현상과 관련하여 맥놀이(beating)과 거칠기(roughness)와 같은 개념들이 사용된다.

¹³ [원주 2] 달팽이관이 '스펙트럼 분석기'라는 것은 사실과 다르다. 스파이크 뉴런 과정에 대한 보다 현대적인 관점에서는 음조가 물리적으로 국소화된 자극뿐만 아니라 시간별로 구분된 신경 자극으로 인코딩된다는 것을 보여주기 때문이다.

¹⁴ [옮긴이주] 바크 척도(Bark Scale)는 1961년 독일의 음향 과학자 에버하르트 초비커가 제안한 음향심리학 척도이다. 인간은 소리의 크기와 높이, 길이, 음색등의 특징을 청각기관을 이용하여 구분할 수 있는데, 소리의 특징을 구체적으로 나타내기 위해서는 서로 다른 소리를 구별하기 위한 척도가 필요하다. 소리의 크리를 측정하는 기법을 최초로 제안한 하인리히 바크하우스젠(Heinrich Barkhausen)의 이름에서 유래한다.

서브우퍼 라우드스피커를 의미한다. 저음과 높은 저음(*upper base*)는 킥 드럼, 베이스 기타 기본음, 피아노와 기타의 낮은 옥타브를 담당하며, 중음역대는 보컬, 스트링, 브라스 스윗스팟과 관련된 영역으로 나뉜다. 고음역대에는 보컬과 악기에 선명함을 더하는 *프리센스(presence)*, 심벌즈와 시끄러운 소리를 위한 *탑(top)*, 시스템의 한계까지 모든 가장 높은 배음을 포함하는 *에어(air)*¹⁵가 포함된다. 일부 연구자(예: Leipp 1977)는 분석용으로 이를 정량화하여 소위 *감각 대역(sensible bands)*으로 분류했다. 그림 6.2의 아래쪽 눈금은 32밴드 그래픽 이퀄라이저의 분할을 보여준다.

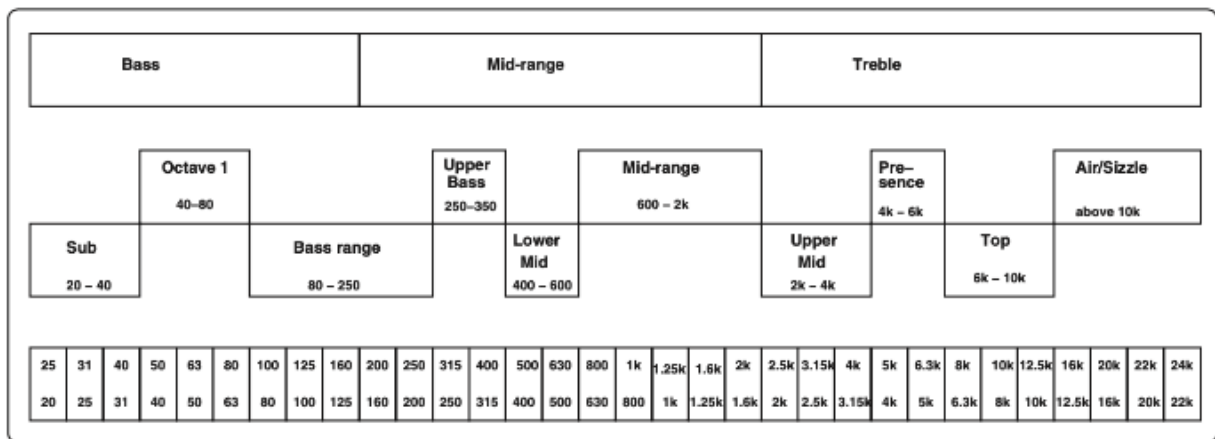


그림 6.2

사운드 스펙트럼 내의 범위에 대한 일반적인 설명.

해상도 - p.85

달팽이관은 한정된 수의 유모세포로 구성되어 있기 때문에 주파수의 해상도에는 한계가 있다. 그러나 유모세포와 우리가 들을 수 있는 주파수 사이에는 일대일 대응이 없는데, 만약 그랬다면 우리는 약 18,000개의 개별 주파수만 들을 수 있었을 것이다. 대신 우리는 두 가지 종류의 정보를 결합한다: 파형의 각 주기마다 뉴런이 발화하는 타이밍 정보, 그리고 달팽이관속 위치로 인해 임계 대역에 있는 여러 뉴런 그룹이 공명하는 위치 정보가 그것이다. 뉴런은 특정 속도(일부에서는 최대 1kHz로 추정)까지만 발화할 수 있다. 개별 뉴런은 특정 속도로만 발화할 수 있지만, 나중에 여러 뉴런의 출력이 결합되어 더 나은 해상도를 제공할 수 있다. 이 과정은 결코 선형적이지 않다. 저음역대에서는 100Hz와 105Hz의 차이가 5Hz밖에 나지 않는데도 우리는 그 차이를 듣는 데 아무런 문제가 없다. 이러한 주파수는 쉽게 구분할 수 있다. 그러나 최고음에서는

¹⁵ [옮긴이주] 초고역대와 고역대의 소리에서 얻을 수 있는 공간적 음향.

16000Hz와 16005Hz의 차이를 구분할 수 없다. 그래서 우리는 이러한 주파수를 *분해할 수 없다(unresolvable)*고 말한다. 이 두 개의 고주파를 함께 들을 때 우리는 뉴런 타이밍에 의해 쉽게 인코딩되는 5Hz의 맥놀이 톤(beat tone)을 감지하므로 이 범위에는 두 개 이상의 주파수가 있다는 것을 알 수 있다. 하지만 16000Hz와 16005Hz인지, 아니면 16000Hz와 15995Hz인지는 명확하지 않으며, 두 주파수 모두 5Hz의 맥놀이 패턴을 나타내므로 지각에 모호함이 있다. 건강한 청소년이 들을 수 있는 20kHz까지 도달하면 임계 대역폭이 너무 커서 20kHz와 21kHz가 같은 영역을 자극하지만 비트 차이가 너무 빨라서 타이밍으로 인코딩하기 어렵다. 이 시점에서는 모든 주파수가 감각을 유발할 수는 있어도 식별할 수는 없다. 따라서 청각은 여러 정보 채널을 사용한다. 여러 뉴런의 타이밍 정보를 위치 정보와 결합하고 중요 대역 사이의 맥놀이를 사용하여 주파수를 구분하는 능력을 사용하면, 귀의 해상도는 매우 뛰어나게 된다.

근접 구성 요소의 평균

많은 실제 파동은 두 개 이상의 근접 주파수를 포함하지만 여전히 주기적(periodic)이다. 감지된 주파수가 반드시 전체 파동 패턴의 주기가 되는 것은 아니다. 피크와 영점은 신경 타이밍에 의해 인코딩되며, 개별 주파수는 임계 대역의 위치를 자극한다. 임계 대역에 속하는 음고에 대한 인식은 이 두 가지 지표를 통해 형성되므로 둘 사이의 어딘가에 있는 것처럼 보인다. 예를 들어 20Hz와 30Hz가 혼합된 음은 10Hz에서 주기적이지만, 약 25Hz로 들린다.

진폭의 급격한 변화

진폭 변화는 음고에 작은 영향을 미칠 수 있다. 정현파의 진폭이 기하급수적으로 변화하면 명백한 음고가 수정된다(Hartmann 1977, 2004). 이는 지각적 특징으로 간주하는 것이 가장 좋으며 변조로 인해 측정 가능한 주파수가 발생하는 측파대와는 실제로 동일하지 않지만, 여기에는 분명히 근본적인 연관성이 있다. 예를 들어, 총소리나 천둥소리의 빠른 역동성으로 인해 어택 단계에서 실제보다 더 높은 주파수로 느껴질 수 있다.

팬텀 기음 – p.86

어떤 소리는 실제로 존재하는 어떤 주파수보다도 낮은 음고를 가진 것으로 들린다(Seebeck 1841; Schouten 1940). 특정 스펙트럼은 현실에 존재하지 않는 기본 주파수를 *감지*한다. 오보에와 일부 다른 목관악기 소리가 이에 해당한다. 이 현상은 오디오 스펙트럼의 하단에 "공간을 만드는" 데 사용될 수 있으며, 신호를 왜곡하여

실제 기본 주파수보다 낮은 팬텀 주파수를 도입하는 일부 서브 베이스 효과 프로세서에 의해 사용된다. 엄격하게 적용하기는 어렵지만 일반적인 규칙(Goldstein 1973)은 일련의 배음이 그것이 존재 *해야 하는 것처럼* 배열되어 있으면, 그것이 들린다는 것이다. 이것이 전화기에서 남성의 목소리가 신호가 기본 범위 이상으로 강하게 하이 패스 필터를 통과하였더라도 여전히 선명하게 들리는 이유다.

허긴스 양이간 음고(Huggins Binural Pitch)

음고는 물리적, 주기적 요소가 없는 곳에서도 명백하게 나타날 수 있다. 양쪽 귀에 모두 들리지만 한쪽 귀에서 작은 주기 T만큼 이동된 백색 소음이 중간에서 1/T의 음으로 인식될 수 있다(Cramer and Huggins 1958). 이 현상은 각 귀의 입력을 상호 연관시켜 소리의 신호 대 잡음비를 최소화하려는 뇌의 기능인 IID 처리와 관련이 있다(Bilsen 1977). 이는 스테레오 페이징 및 플랜지 효과의 딥 컬러와 관련이 있으며, 음향적 특성만으로는 공명을 설명할 수 없음에도 불구하고 하강 및 상승 음을 내는 것처럼 보이는 제트 비행기와 같은 것을 고려할 때 사운드 디자인에 유용하다.

빌센 밴드 에지 음고(Bilsen Band Edge Pitch)

광대역 노이즈에 매우 가파른("벽돌 벽") 필터¹⁶를 적용하면 필터가 최대한으로 평평하더라도 차단 주파수¹⁷에서 음고가 있는 것처럼 인식될 수 있다. 이는 사운드 디자인에서 중요하며, 특히 컬러를 피해야 하고 특정 노이즈 분포가 효과에 중요한, 비내리는 것과 같은 미묘하고 자연스러운 효과를 만들 때 유용하다.

스펙트럼의 지각

측정 가능한 양의 주관적이고 지각적인 품질, 즉 스펙트럼에 대해 다시 한 번 *음색(timbre)*이라는 새로운 용어를 소개하겠다. 물론 스펙트럼은 한 차원을 가진 스칼라 양이 아니라 값의 행렬이다. 따라서 음색은 단순한 순서로 순위를 매길 수 없으며, 특정 음색은 서로 가깝고 다른 음색은 멀리 떨어져 있는 공간(Wessel 1976; Grey 1975; Wessel and Grey 1978)을 형성한다. 음색은 기본 주파수와 일련의 배음이 있는 정적 스펙트럼을 의미한다는 점을 명심하라. 이것은 주파수 영역의 한 단면일 뿐이며 소리의 변화를 포착하지

¹⁶ [옮긴이주] 통과 대역의 전체 전송, 정지 대역에서 완전한 감쇠 및 급격한 전환을 갖춘 이상적인 전자 필터를 구어적으로 표현한 것.

¹⁷ [옮긴이주] 차단 주파수(cutoff frequency)란 필터에서 통과 대역과 차단 대역의 경계가 되는 주파수. 통상, 필터의 출력 진폭이 입력 진폭보다 3dB만큼 감소되었을 때의 주파수를 뜻한다.

못한다. 우리가 음색에 대해 이야기할 때는 일정한 스펙트럼이 만들어내는 즉각적인 지각적 감각에 대해 이야기하는 것이다.

고조파와 비고조파 스펙트럼의 인식 – p.87

이전 장에서 진동하는 물체는 모든 주파수가 기본 주파수와 단순한 정수 관계를 갖는 고조파 스펙트럼을 생성하거나, 주파수가 적분이 아닌 더 복잡한 방식으로 간격을 갖는 비고조파 스펙트럼을 생성할 수 있다는 것을 배웠다. 고조파 분포는 "순수한" 또는 협화 음색으로 인식되는 반면, 비고조파는 "거친" 또는 불협화음(dissonant)이라고 할 수 있다. 흥미롭게도 순수한 음색은 고조파가 절대 정수 값과 완벽하게 일치하는 것이 아니라, 일정한(단조적인) 진행을 형성해야 한다. 예를 들어 100Hz, 199Hz, 396.01Hz, 788.06Hz 계열은 완벽하게 순수한 소리다. 각 배음은 이전 음의 1.99배인 거의 1옥타브이다. 비정수 배음으로 구성된 음은 불협화음이 없기 때문에, 절대 음계라는 것은 우리의 뇌에 고정된 것이라기보다는 학습된 것이라고 보는 것이 타당해 보인다. 반면에 100Hz, 200Hz, 400Hz, 411Hz는 처음 세 음질이 완벽한 옥타브임에도 불구하고 거칠고 불협화음으로 들린다. 첫 번째 예의 경우, 음계로 연주하려고 하기 전까지는 완벽하게 순수하게 들리지만 잘못된 소리가 나는 반면, 후자는 본질적으로 "거칠게" 보이지만 하모닉 음계로 연주할 때 균일하고 일관되다.

협화음, 조화음, 그리고 거칠기

위의 현상은 스펙트럼과 그것의 이동된 복사본이 맺는 관계인 국소적 협화음(local consonance)의 원리를 보여준다. 특정 악기가 특정 음계에서 더 잘 들리는 이유도 바로 이 때문이다. 또한 이 현상은 작곡에서 오케스트레이션과 편곡이라는 매혹적인 예술을 탄생시키는 음색과 화성 사이의 복잡한 관계도 설명한다. 음악적 영역으로 넘어가는 동안 화음과 불협화음과 같은 화성 규칙도 이 원칙에서 벗어난다는 점도 주목할 가치가 있다. 음표의 배음이 음계에서 다른 음표의 배음과 일치하면, 그 음색은 그 음계의 화음이다. 헬름홀츠는 맥놀이(beating)의 관점에서 협화음과 불협화음을 설명했다. 맥놀이 주파수가 몇 헤르츠 이상일 때, 느린 위상 이동을 일으키는 가까운 주파수는 귀에 성가신 느낌을 준다. 정확하게 정렬된 배음은 맥놀이를 일으키지 않는다. 최근의 연구(Plomp와 Levelt 1965)에 따르면 달팽이관 임계 대역의 유모세포가 모호하게 자극될 때 불쾌한 거칠기¹⁸가 유발되는 것으로 보이며, 두 사인파 사이에서 인지된 부조화는 그래프로 나타낼

¹⁸ [옮긴이주] 거칠기(roughness)는 소리의 빠른(15-300Hz) 진폭 변조에 대한 주관적 인식을 정량화한 심리 음향

수 있다. 이 불협화음 곡선은 약 반음(11번째 간격), 즉 임계 대역폭의 1/4인 $0.25 * 1\text{Bark}$ 의 간격에서 가장 큰 거칠기가 발생한다는 것을 보여준다. 이 이상에서는 옥타브로 갈수록 간격이 증가함에 따라 불협화음이 감소(협화음이 증가)하지만, 다섯 번째와 같은 "특별한" 간격에서는 마술처럼 변하지 않는다. 다시 말하지만, 이는 서양의 음악적 개념이 일부 텍스트에서 주장한 것처럼 지각적으로 타고난 것이 아니라는 점을 강조한다. 비음악적 업계에 종사하는 사운드 디자이너로서 우리는 여전히 이 지혜를 활용할 수 있다. 불쾌한 효과로서의 거칠기는 dB SPL 으로 더 잘 이해될 수 있다. 이는 동물의 발걸음이나 고속 바람과 같은 위험한 고에너지 현상으로 인해 발생하는 4Hz~8Hz 범위의 주기로 구성된 간섭 리듬이다. 더 높은 주파수 구성 요소를 변조하면 결과적으로 임계 대역 모호성이 발생하여 불안한 효과가 발생한다.

밝음, 둔탁함, 그리고 미트 스펙트럼의 중심점 – p.88

음악에서 많이 사용되는 주관적인 음계는 소리의 밝기(Brightness) 또는 둔탁함(dullness)이다. 밝은 소리는 단순히 둔한 소리보다 높은 배음이 더 많다. 하지만 단지 높은 주파수를 몇 개 더 가지고 있다고 해서 소리가 밝다고 할 수는 없다. 소리가 둔한지 밝은지를 측정하는 가장 좋은 방법은 스펙트럼의 '무게 중심' 또는 스펙트럼 중심(spectral centroid)을 찾는 것이다. 이는 가중 평균이므로, 우리는 (대역수를 곱한) 각 임계 대역의 주파수 강도를 더한 후 모든 대역의 총합으로 나누어 구한다. 에너지 x_n 을 포함하는 주파수 f_n 주변의 임계 대역 N 개의 경우,

$$\text{Centre of spectrum} = \frac{\sum_{n=0}^{N-1} f_n \times x_n}{\sum_{n=0}^{N-1} x_n} \quad (6.4)$$

또한 밝은 음색을 내기 위해서는 스펙트럼에 충분한 연속성이 있어야 한다. 고립된 고주파 성분이 포함된 저주파의 스펙트럼은 그 위에 별도의 '아티팩트'¹⁹가 있는 둔한 소리로 들릴 것이다. 밝기는 많은 음색 공간

변수이다. 거칠기의 측정 단위는 asper이며, 1 asper는 70 Hz에서 100% 진폭 변조된 60 dB의 1 kHz 톤에 의해 생성된 거칠기로 정의된다.

¹⁹ [옮긴이주] 음향적 아티팩트(sonic artifact)는 소리를 녹음하거나 합성하는 과정에서의 기술적 제한으로 인해 발생하는 음향 오류이다. 한편 일렉트릭 기타 또는 베이스 신호를 오버드라이브하여 잘리고 왜곡된 톤을 만들거나, 빈티지 바이닐 레코드처럼 들리게 하기 위해 녹음에 거친 팡파 클릭을 추가하는 것과 같이 의도적으로 아티팩트를 추가하는 경우도 있다.

연구에서 스케일링을 위한 차원으로 사용되며(Wessel 1973, Grey 1975, Moorer, Grey, Snell 1977-78) 마찰 및 비선형 프로세스(활 현악기와 호른)를 특징짓는 척도로 사용된다. 소리의 *어택* 부분(아래 참조)은 밝기 지각에서 매우 중요하다(Helmholtz, Wessel 1973; Grey 1975, Schaeffer 1977). 소리가 밝은 음으로 시작하면 나머지 소리는 더 밝은 것으로 가중치가 부여된다. 둔탁한 음색은 스펙트럼 중심이 낮거나 음성 범위 내에서 해상 불가능한 배음이 없다. 이러한 소리는 댐핑되거나 음소거된 것처럼 보인다.

공명, 평탄도 및 포먼트

공진 음색(Resonant timbres)은 대부분의 주파수가 하나 이상의 고정점 주변에 모여 있는 *날카롭고* 좁은 스펙트럼을 가진다. 또는 스펙트럼이 고르게 분포되어 있는 평탄하거나 넓은 음색이라고도 한다. 심벌즈와 같은 잡음이 많은 소스는 넓고 평평한 음색을 내는 반면, 나무 피리는 날카롭고 공명하는 음색을 낸다. 기타와 바이올린과 같은 많은 어쿠스틱 악기에는 여러 개의 고정된 공명이 있다. 어떤 음이 여기자(현)²⁰에서 나오는 것인지, 어떤 음이 공명기(몸체)의 일부인지 구분할 수 없기 때문에 단일 음으로는 이를 알 수 없지만, 두 개 이상의 다른 음이 연주되면 공통 공명이 분명해진다. 지각 용어로는 이 고정된 공명 특징을 *포먼트(음형대, formant)*라고 한다. 300Hz~1kHz 범위의 날카로운 공명과 포먼트가 결합된 소리는 목소리처럼 들린다.

시간 구조의 지각

다음 항목은 소리를 시간으로 나누는 방법을 다룬다. 크게 세 가지 척도를 사용하여 소리를 배치할 수 있다. 첫 번째 구분은 미시 구조와 거시 구조이다. 미시 구조는 1초 이하의 모든 것을 말한다. 플루트의 트릴, 단어의 음절 또는 불의 개별적인 딱딱한 소리 등이 이에 해당한다. 거시적 구조는 대화를 구성하는 문장부터 교향곡 전체의 주제 전개에 이르기까지 그보다 긴 모든 것을 말한다. 이 규모에서는 수십 초, 분, 시간을 고려한다. 미세 구조는 더 나아가 짧은 소리, 아주 짧은 소리, '나노 스케일' 소리로 나눌 필요가 있다. 이 범위 내에서 중요한 지각적 경계가 발생하는데, 이 경계는 다음에 살펴볼 것이다.

²⁰ “여기점(exiting point)은 물체에 힘이 전달되는 지점이다. 여기는 바이올린 활처럼 긁어내는 움직임의 경우 마찰이 될 수도 있고 충돌과 같은 충격이 될 수도 있다. 물체는 주변의 난기류에 의해 여기되거나 줄을 잡아당기는 것과 같이 갑작스러운 물리적 방출로 인해 움직임을 유발할 수 있다. [...]대부분의 여기는 진동하는 다른 연결된 물체로부터 직접적인 힘을 받는 단순한 결합을 통해 발생한다. 시스템에 에너지를 가져오는 것을 여기자(exitor)라고 하고 진동하는 것을 여기(excitation)라고 한다.” Designing Sound Cp.1 Section 3.1 에서 발췌.

입자도(粒度, Granularity) – p.89

이 용어는 소리의 미세 구조에 실제로 적용된다. 약 20ms 미만의 소리는 그보다 긴 소리와 다르게 들린다. 가보르(1946)와 슈톡하우젠은 이 개념에 기여했으며, 세분화된 합성은 이 개념에서 나온 실용적인 방법 중 하나이다. 하한에서는 사건에 대한 개념이 음고에 대한 인식과 융합된다. 매우 팽팽한 드럼 롤의 속도가 빨라지면, 약 18Hz에서 더 이상 개별적인 비트의 연속이 아니라 하나의 음고가 된다. 또한 50밀리초에서 150밀리초 범위의 밀러/반 노든 임계값과 1초 장벽을 넘어서는 구간에서도 입자도가 적용된다. 이렇게 하면 소리의 미세 구조가 세 가지의 입자도의 영역으로 효과적으로 분할되어 약간 다른 규칙이 적용된다.

사건과 흐름

우리는 학교에서 수학을 배울 때 정수를 먼저 배운다. 1의 배수는 '3시'나 '티타임'과 같은 사건과 마찬가지로 어린 마음에도 이해가 된다. 나중에 우리는 이것들이 허구이며 모든 것이 항상 움직이고 있다는 것을 알게 되고, 어른들의 마음을 더 편안하게 해주는 또 다른 허구, 즉 *실수*와 연속적인 흐름에 대한 개념을 도입한다. 정신적 도구로서 두 가지 모두 똑같이 유효하며, 우리가 사건을 다루든 흐름을 다루든 그것은 단지 유용성의 문제일 뿐이다. 거시 구조적 수준에서 일어나는 일, 즉 음악의 특정 마디, 영화의 특정 장면이 사건을 구성한다고 말하는 것은 상당히 안전하다. 그리고 미시적 구조 수준에서 일어나는 일, 즉 소리를 구성하는 샘플과 파형은 연속적인 흐름이라고 해도 무방하다.

포락선

사건과 흐름을 의미 있는 프레임으로 나누는 좋은 방법은 시스템 내에서 발생하는 에너지의 변화를 고려하는 것이다. 증가하고 있는가? 감소하고 있는가? 안정적인가? 사운드 디자인에서는 일반적으로 *어택*(*attck*), *디케이*(*decay*), *서스테인*(*sustain*), *릴리스*(*release*)라는 네 가지 단어를 사용하여 소리의 포락선을 설명한다. 이 단어들은 진폭, 주파수 또는 음색과 같은 모든 가변적인 질에 적용될 수 있다. 신세스에 사용되는 제어 포락선은 이러한 단계를 반영하기 위해 ADSR로 약칭하는 경우가 많다.

어택

소리가 시작될 때 에너지가 0에서 최대 에너지로 이동하는 기간을 어택이라고 한다. 타악기 소리의 경우 어택은 매우 빠르다. 10ms 미만의 어택 시간은 일반적으로 딸깍 소리로 들리지만, 대부분의 실제 소리는 매우

빠르더라도 일부 에너지가 소리로 방출되기 전에 신체에 흡수되기 때문에 순간적인 어택이 없다. 물체가 작을수록 이 초기 에너지 흡수가 더 빨라진다. 어택력이 거의 없다고 생각할 수 있는 소리는 작은 유리 구슬을 타일 바닥에 떨어뜨리는 소리일 수 있다. 어택을 *트랜지언트(transient)*와 *상승기(rise)*의 두 단계로 더 세분화하는 것이 유용하다.

트랜지언트 및 상승 시간 –p.90

트랜지언트(transient)²¹는 여기 단계(excitation state)에 해당한다. 트랜지언트는 종종 나머지 사운드보다 상당히 크며 항상 짧다. 일반적인 트랜지언트의 길이는 5ms에서 50ms 사이다. 여기에는 여기된 물체와 여기자(예: 드럼 스틱 또는 벨 해머)의 소리가 모두 포함될 수 있다. 진동 에너지가 시스템에 계속 축적되는 동안 상승이 발생한다. 일반적으로 기본 모드의 한 주기 또는 물체의 길이를 재료의 음속으로 나눈 값보다 짧은 경우가 많다. 트랜지언트 상태가 끝나기 전에 최대치에 도달하는 경우가 많으므로, 이 둘은 분리할 수 없는 것처럼 보인다. 그렇지 않은 좋은 예가 징이다. 징을 칠 때 금속을 때리는 채에서 발생하는 트랜지언트 현상은 초기 여기파가 몸체를 통해 퍼지는 짧은 기간에 선행한다. 채의 충격은 국소적으로 짧은 수명의 고주파를 생성하지만, 변위의 확산은 상당히 낮은 주파수 사건이며, 그 후에 주요 진동 모드가 나타나고 징이 살아난다.

느린 어택

어택이 매우 길어질 수도 있다. 완만하게 구부러진 현에서 진폭은 현의 복원력과 활에 가해지는 힘이 마찰에 의해 제한된 최대 동적 평형에 도달할 때까지 계속 증가하며, 이 과정에서 몇 초가 걸릴 수 있다. 어택만 있는 것으로 간주될 수 있는 소리는 물체가 경사진 표면에서 공중으로 미끄러지는 소리다. 이 경우 마찰 여기로 인해 계속 가속되어 표면 가장자리에서 떨어질 때까지 점점 더 큰 소리를 낸다. 또 다른 예는 간단한 불꽃놀이 로켓이다. 일반적으로 멀리 사라지는 것을 무시하고 연료의 연소 표면적이 증가하기 때문에 타거나 폭발할 때까지 소리 출력의 강도가 높아진다.

디케이

²¹ [옮긴이주] 음파가 시작되는 짧은 순간 발생하는 폭발적인 진폭 변화. 일반적으로 소리의 어택 부분 또는 디케이부분에서 발생하며, 음에서 가장 많고 복잡한 배음 정보를 담고 있다.

감쇠는 트랜지언트 단계 이후에도 에너지가 계속 공급되는 시스템에 적용된다. 이러한 소리는 일반적으로 자루를 땅 위로 끌거나 트럼펫을 부는 것과 같은 마찰음 또는 난류음이다. 감쇠는 시스템에 초기 입력된 에너지가 지속 가능한 수준을 초과한 후에 발생한다. 그 후 시스템은 일종의 안정된 동적 평형 상태에 도달하는데, 이를 *지속(sustain)*이라고 한다.

서스테인

이 기간 동안 시스템에 입력되는 에너지는 에너지 출력과 같으므로(열로 인한 내부 손실을 뺀 값) 소리를 구성하는 에너지의 비율은 일정하게 유지된다. 좋은 예로 현악기를 들 수 있다. 셰퍼(1977)는 이를 음향 분류학에서 *에너지틱 메인टे넌스(energetic maintenance)*라고 부른다. 서스테인 단계는 일반적으로 일정한 사운드 출력을 생성하지만 항상 그런 것은 아니다. 주기적으로 방향을 바꿔야 하기 때문에 전체적인 에너지 전달 수준은 상당히 균일하지만, 에너지가 약해지거나 약해지는 지점이 있기 때문에 활이 좋은 예가 될 수 있다. 흥미로운 반대 예는 길고 균일한 경사면을 굴러가는 대리석이다. 저장된 에너지가 일정한 속도로 방출되고 있기 때문에 엄밀히 말하면 시스템은 *방출* 단계에 있다. 그러나 생성되는 소리 에너지는 일정하고 지속되는 것처럼 보인다. 폭포 위로 흐르는 물은 지속적인 과정이다. 상류의 강은 일정한 에너지를 제공하고, 하류의 강은 물이 사라지는 배수구를 제공하며, 폭포 자체는 움직이는 물의 일부 에너지를 소리와 열로 바꾸는 흐름의 비효율성에 불과하다.

릴리스 – p. 91

릴리스는 시스템에 에너지 투입을 중단하면 발생하지만, 여전히 저장된 에너지가 남아 있어 한동안 계속 소리를 내고있는 상태다. 움직이는 물체의 경우 이 값은 운동량 또는 유체의 저장 용량에 해당한다. 릴리스는 항상 0이 되는 경향이 있으며 소리의 마지막 단계다.

노력과 움직임

피에르 셰퍼(Pierre Schaeffer, 1977), 미셸 시옹(Michel Chion, 1994), 루돌프 라반(Rudolf Laban, 1988)은 소리 역학에 대한 엄밀하거나 정량화하기는 어렵지만 가치 있는 해석을 내놓았다. 셰퍼의 1966년 솔프게(Solfège, 계명창)는 시옹에 의해 종, 치수, 게이지, 스트레치, 무게, 릴리프, 충격, 조음, 유지, 형태, 유지, 양식 및 충동과 같은 동적 설명자로 발전되었다. 라반의 무용 작업은 인간 형태의 질량과 움직임을

통합하여 사운드에 적용될 수 있는 제스처의 느낌을 설명한다. 노력, 공간, 시간, 무게, 흐름이라는 개념에서 시작하여 누르기, 미끄러지기, 찌기, 뜨기, 두드리기, 베기, 튕기기, 밀기 등 일련의 공리적인 제스처의 클래스가 형성되었다.

선행음과 귀속

소리를 구조적 요소로 분해할 때, *하스 효과(Haas effect)* 또는 *선행음 효과(precedence effect)*는 우리가 소스와 환경을 분리하는 지점을 결정한다. 약 30ms 후에 수신된 반향은 환경에 할당되어 별도의 에코로 들린다. 한 소리에서 30ms 이내에 발생하는 소리는 그 소리로 융합되어 동일한 물체에 귀속된다. 따라서 클래식 기타의 몸체는 그 내부에서 여러 에코를 생성하는 음향 공명기이지만, 이 에코는 동일한 음원으로 혼합된다. 이 소리들은 방이 아닌 기타에 속한 것처럼 보인다. 큰 홀에서 손뼉을 치면 개별적인 에코가 만들어진다. 그 에코는 손이 아닌 방에 속한 것처럼 보인다.

지속 시간에 대한 가버 임계

가버 임계(Gabor Limit)는 소리가 음고를 갖는다고 할 수 있는 최소 지속 시간을 나타낸다. 약 10밀리초에서 20밀리초보다 짧으면 음높이 감각을 자극하기에 충분한 파동 주기가 포함되지 않아 딸깍거리는 소리로 들린다. 클릭의 명백한 주파수(또는 *톤*이라는 모호한 단어를 사용하기에 좋은 곳일 수도 있다)는 지속 시간에만 의존한다. 짧은 클릭은 더 높고 날카롭게 들리며, 긴 클릭은 더 낮고 둔하게 들린다.

순서에 대한 허쉬 임계(Hirsh Limit for Ordering)

과학자들은 무작위 오프셋 타이밍으로 청취자에게 짧은 소리를 들려주었다(허쉬와 셰릭 1961). 때로는 A가 B보다 짧은 간격으로 앞설 때도 있고, 그 반대일 때도 있었다. 때로는 우연히 두 소리가 동시에 시작되는 경우도 있었다. 간격이 약 20밀리초 이하일 때는 두 소리의 순서가 우연에 의해 정해지지 않는 것으로 나타났다.

스트리밍²² – p.92

소리의 인지된 순서와 밀접하게 연관된 것이 *스트리밍*으로, 소리의 구성 요소가 같은 것에 속한다는 믿음으로 함께 그룹화하는 것이다. 어떤 전화 벨소리와 새소리는 짧은 고음과 저음을 오가며 빠른 연속을 구성하지만,

²² [옮긴이주] 본문에서 stream은 '스트림'으로, flow는 '흐름'으로 옮겼다.

우리는 이를 각각의 일정한 음으로 구성하지 않고 하나의 트릴로 듣는다. 비슷한 소리가 여러 개가 빠르게 연속적으로 발생하면 *융합(fusion)*되어 하나의 소리가 된다고 말한다. 빗방울의 낙하량이 적을 때는 빗방울 하나하나가 '자박 자박' 소리를 내지만 들이붓는 폭우가 쏟아지면 그것은 하나의 소음의 벽이 된다. 겉보기에 일관된 소리의 밀도가 특정 임계값 아래로 떨어지고 개별적으로 식별 가능한 부분으로 분리되기 시작하면 그 반대가 된다. 이를 *분열점(fission point)* 또는 *시간적 일관성의 경계(temporal coherence boundary)*라고 한다. 스펙트럼과 주파수의 작은 차이로 청취자는 따라갈 수 있는 윤곽을 가진 통합된 스트림을 인식한다. 스펙트럼이나 주파수가 분기하면 청취자는 고음과 저음의 두 개의 분리된 스트림을 듣기 시작한다. 이 시점에서 주의는 두 스트림 중 하나에 개별적으로 집중할 수 있어도 전체 패턴에는 집중하지 못하게 된다.

음고에 의한 스트리밍

고음과 저음을 번갈아 가며 실험한 결과(Miller and Heise 1950), 음의 주파수 간격(반음으로 측정)에 따라 50ms에서 150ms의 영역에서 분열과 융합이 일어나는 것으로 나타났다. 주파수 사이의 간격이 더 큰 경우, 음고는 교대 속도가 더 높아질 때까지 융합되지 않았다. 분열과 융합에 필요한 정확한 비율은 개별 청취자에 따라 다르다. 순음의 스트리밍에서 나타나는 현상은 음고와 시간적 식별이 서로 경쟁한다는 것을 나타낸다.

반 노어든 모호성

반 노어든(1975)이 발견한 모호성은 분열과 융합 상태(100ms에서 최대 1500ms) 사이의 영역으로, 청취자가 두 가지 방식(권유 또는 초기 조건의 영향을 받은) 중 하나를 선택하여 소리를 들을 수 있는 것이 마치 원근감이 전환되는 계슈탈트의 큐브 착시 현상과 유사하다. 반 노어든은 통합된 트릴이 두 개의 분리된 소리의 흐름이 될 때까지 주파수와 교대 주기를 변경한 다음, 흐름이 다시 통합될 때까지 하나 또는 다른 변수를 뒤로 이동하는 실험을 수행했다. 그는 분열과 융합의 경계가 어느 쪽에서 접근하느냐에 따라, 그리고 주의집중 설정에 따라 달라지면서 *히스테리시스*가 발생한다는 사실을 발견했다.

스펙트럼적/공간적 스트림

주파수와 타이밍에 따라 소리를 그룹화하는 우리의 경향은 스펙트럼의 영향을 받기도 한다. 이산 *음색 공간(discrete timbre space, Wessel 1979)*이라는 개념이 유용하다. 특정 한계까지 우리는 플루트라고 간주할 수 있는 많은 소리가 있으며, 이 한계 내에서 악기가 연주하는 멜로디를 들으면 스펙트럼이 변하더라도 동일한

음원에서 나온 것으로 간주한다. 그 한계를 넘어 오보에 쪽으로 갔다면, 이제 다른 악기가 변경된 스펙트럼으로 음을 연주하는 소리가 들리고 우리는 두 명의 음악가가 함께 연주한다고 생각한다. 매우 다른 화성 구조를 가진 두 소리는 공통 배음을 공유하더라도 명확한 경계를 만들어 두 소리를 구분한다. 국소화도 중요한 역할을 하므로 소리가 같은 위치에서 발생한 것처럼 보이면 공통된 원천으로 간주하는 경향이 있다.