

오디오 주파수

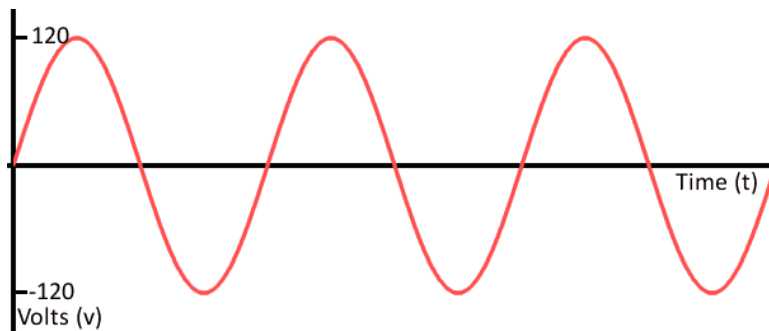
개요

- 소리(sound)

- 소리는 즉, 진동이다.
- 인간이 소리를 듣는 순서
 1. 특정 물체가 440hz로 진동을 한다.
 2. 그 물체 주변의 매질이 440hz의 진동을 전달한다. (공기를 통해)
 3. 440hz의 진동신호가 달팽이관에서 전기 신호로 바뀌어 청신경으로 전달된다.
 4. 뇌가 신호를 받아 해석한다.

- 아날로그(analog)

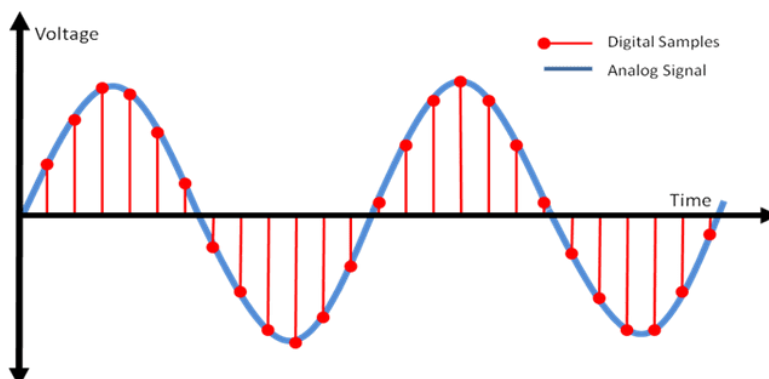
- 진동은 자연계에서 발생하는 것임으로 아날로그 형태이다.
- 아날로그는 신호나 자료를 연속적인 물리량으로 나타낸 것이다.
- 아날로그 신호는 아무리 쪼개도 끝이 없는 연속성을 지닌다.
(2를 2로 나누면 1.5, 1.5를 2로 나누면 1.25, 이렇게 $n/2$ 는 끝이 없는 연속성을 지님)



가로 축 = 시간 / 세로 축 = 전압

- 샘플링(sampling)

- 아날로그 신호를 디지털 신호로 바꾸기 위한 첫번째 단계
- 샘플링 절차 (장치의 예 : 마이크)
 1. 소리의 진동이 장치에 전달되면 이 장치가 전압을 올렸다 내렸다 변환하며 컴퓨터로 입력
 2. 컴퓨터는 특정 타이밍을 정해서 해당 타이밍마다 전압을 측정



파란색 선 = 컴퓨터로 전달되는 연속적인 신호

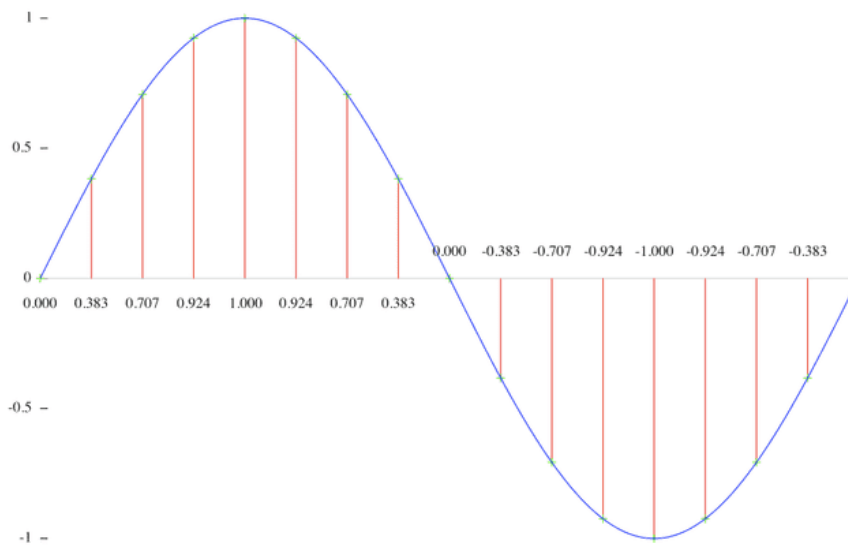
빨간색 점 = 컴퓨터가 특정 타이밍마다 찍어낸 값 (peak)

- 위 그림의 빨간색 선이 좁을수록 원래의 신호에 가까운 값을 측정할 수 있다. (컴퓨터 성능에 달림)
즉, 빨간색 선이 좁을수록 음질이 높아지는 것이다.

- **샘플 레이트(sample rate)**

- 아날로그 신호의 가로 축 해상도를 의미
- CD의 샘플 레이트 : 44.1khz
- 인간의 가청 주파수 : 20hz~20khz

- ✓ 인간의 가청 주파수는 20hz~20khz인데 왜 CD는 인간이 들을 수 없는 대역을 샘플링을 하는 것일까?
그 이유는 아래와 같다.

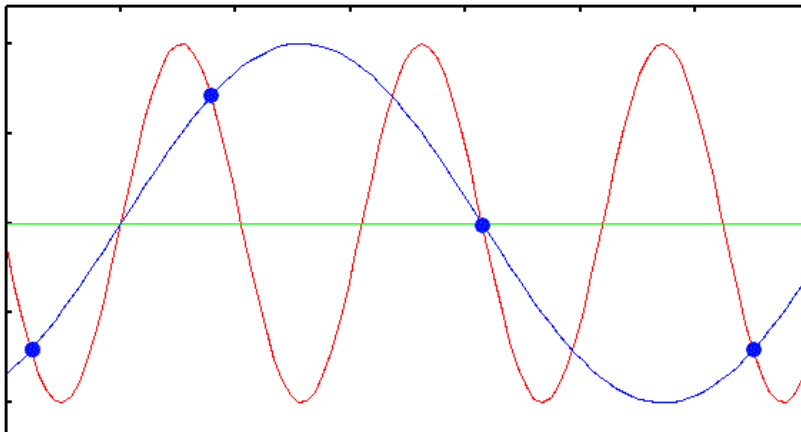


소리 진동의 한 사이클 (=오디오 주파수)

- 기본적으로 오디오 주파수는 이렇게 하나의 사이클 단위로 이루어진다.
- 위로 올라가는 +가 공기가 압축되는 부분이고 아래로 내려가는 -는 다시 공기가 팽창하는 부분이다.
즉, 소리는 공기가 압축 -> 팽창 -> 압축이 되어야 하나의 진동이 되는 것이다.
- 그래서 오디오 신호의 한 사이클을 제대로 측정하려면 +방향의 맨 위 꼭지점 하나와 - 방향의 맨 밑의 꼭지점 하나를 모두 측정해야 하기 때문에 최소 2번은 측정해야 하는 것이다. (나이퀴스트 이론)
다시 말해, 측정하고 싶은 오디오 주파수가 있다면 해당 주파수보다 두배는 더 빠르게 측정해야 한다.

- **앨리어싱(Aliasing or Foldover) 에러**

- 타겟 오디오 주파수보다 느리거나 비슷한 샘플 레이트로 샘플링을 수행할 시 발생



빨간색 선 = 타겟 오디오의 주파수

파란색 선 = 샘플링의 샘플 레이트

- 이 그림은 앨리어싱 에러의 예를 나타낸 그림이다.
- 이렇게 타겟 오디오의 샘플링이 수행되었다면 해당 오디오는 Pitch나 잡음이 담기게 된다.

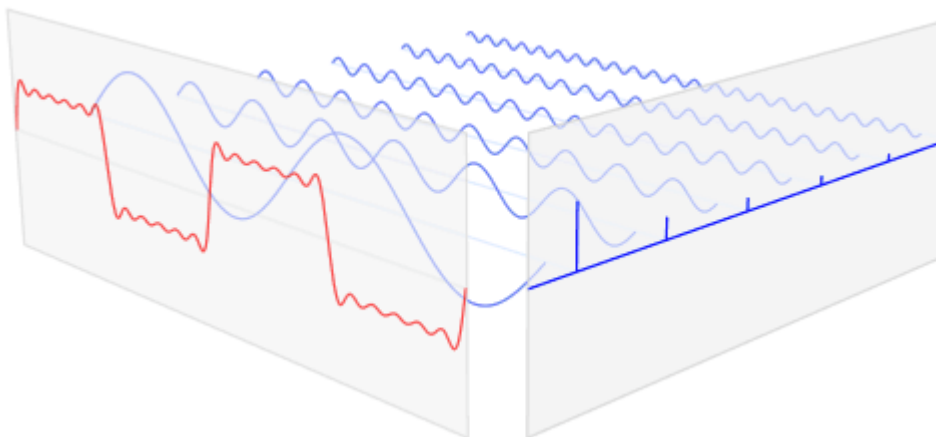
• 오디오 샘플링의 예

- 음악파일의 재생 시간 : $\text{duration} = 277.2124943310658(\text{sec})$
- 샘플링 레이트 : $\text{sampleRate} = 44100(\text{hz})$
- 샘플 개수 : $\text{sampleNum} = \text{duration} * \text{sampleRate} = 12,225,071(\text{ea})$
(1초당 44100개의 샘플을 수집함으로 이와 같은 식이다.)
- ✓ 277초의 음악파일이 44100hz로 샘플링이 되었다면 샘플 개수는 12,225,071개가 되는 것이다.
- 만약 12,225,071개의 데이터가 너무 많아 압축을 수행할 시 **나이퀴스트 이론**에 따라 소리의 한 사이클의 최대 값과 최소 값을 추출해서 차를 구하여 압축하면 된다.
- 압축을 수행하게 된다면 해상도가 다소 떨어지지만 그것은 응용 서비스에 맞춰서 개발하면 되는 것.
(여태까지 아두이노 오디오 샘플링 예제들은 압축을 수행해서 해상도를 떨어트리는 일이 었던 것이었다..)

푸리에 변환

• 푸리에 변환(Fourier Transform)

- 임의의 입력 신호를 주파수 성분으로 분해하는 작업.
- 입력 신호는 전파, 음성 신호 등과 같이 시간축(time)에 대해 정의된 신호일 수도 있고, 이미지 등과 같은 공간축에 대해 정의된 신호일 수도 있다.
- 통신분야에서는 time domain에서 frequency domain으로의 변환이라고 하고, 컴퓨터 비전 및 영상처리 쪽에서는 spatial domain에서 frequency domain으로의 변환이라고 부른다.



왼쪽 = 시간 도메인의 신호 / 오른쪽 = 주파수 도메인의 신호

- 나는 수학적 지식이 많이 부족하여 내가 이해한 얇은 지식을 최대한 풀어서 써보도록 한다.
- ✓ (오디오 신호에서) 시간 도메인은 연속적인 아날로그 신호임으로 특정 진동을 통해 발생된 것이다. 그렇다면 특정 아날로그 신호는 고유한 진동 수가 존재할 것이다. 이때 특정 아날로그 신호가 초당 몇번의 진동(hz)으로 발생된 신호인지를 알아내기 위해 푸리에 변환을 활용하는 것이다. (최대한 설명한 것..)

- **고속 푸리에 변환(Fast Fourier Transform)**

- 컴퓨터 공학에서 사용되는 이산 푸리에 변환(DFT)은 많은 시간이 소요된다.
- DFT의 Big-O는 $O(N^2)$ 이다.
- DFT의 계산시간이 너무 오래 걸려 고안된 방법이 고속 푸리에 변환(FFT)이다.
- FFT는 총 100개의 DFT신호가 있을 경우 그중 10개를 골라내 10개의 신호를 단순히 연결함으로써 제외된 신호에는 예상치를 적용하여 계산한다.
- FFT의 Big-O는 $O(N \log N)$ 이다.

- **FFT 예제 코드**

- 언어 : Python
- 링크1 : <https://techreviewtips.blogspot.com/2017/11/05-02-fft.html>
- 링크2 : <https://pinkwink.kr/708>

참고문헌

- Evan Moon(2019). <https://evan-moon.github.io/2019/07/10/javascript-audio-waveform/>
- 우주산책(2008). <https://m.blog.naver.com/PostView.nhn?blogId=ziokami&logNo=50031225044&proxyReferer=https:%2F%2Fwww.google.com%2F>
- 다크 프로그래머(2017). <https://darkpgmr.tistory.com/171>
- RosyPark(2019). <https://rosypark.tistory.com/73>
- toto R(2017). <https://techreviewtips.blogspot.com/2017/11/05-02-fft.html>
- PinkWink(2015). <https://pinkwink.kr/708>