딥러닝을 이용한 스피커 음질 개선

수학교육과 2019-18406 김창연

서론.

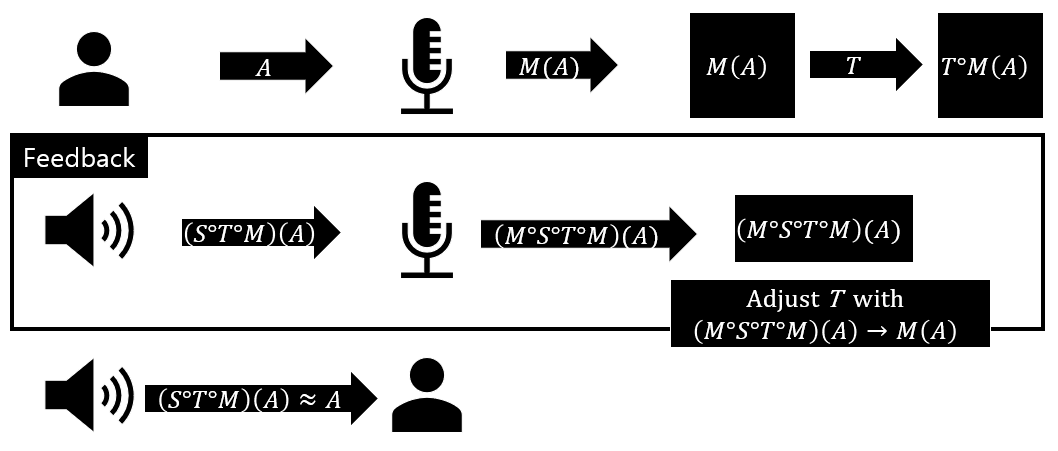
이 프로젝트는 마이크에 입력되는 원음을 스피커로 재현하려는 것을 목적으로 한다.

스피커는 컴퓨터에서 시간에 따른 전압 값을 받아, 이에 따른 압력을 주어 소리의 진동으로 바꾸어 주는 장치이다. 이 과정에서 마이크나 스피커의 진동판의 복원력 등의 차이로 변형이 일어난다. 이 때, 컴퓨터에서 적당하게 변형한 값을 스피커에 전달하게 된다면, 원음에 해당하는 소리가 나게 할 수 있다고 생각했다.

임의의 녹음파일과, 마이크와 스피커로 다시 녹음한 파일로 피드백을 거치면, 원음에 해당하는 소리가 나도록 할 수 있다 생각했고, 이를 구현해보았다.

원리.

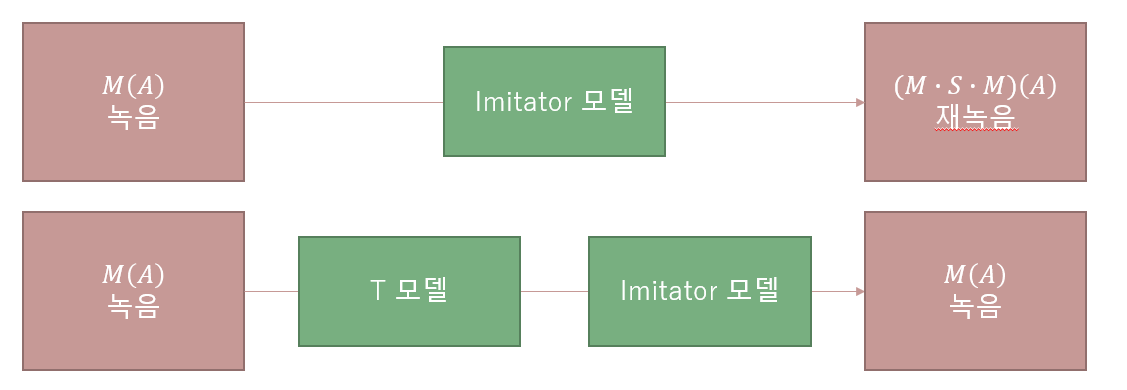
다양한 소리가 담겨있는 소리파일과, 이를 스피커로 출력한 것을 마이크로 다시 녹음해서 소리파일을 얻으면, 스피커와 마이크의 변형 상태를 담은 데이터를 얻을 수 있다. 이를 이용해 모델을 학습시키면, 마이크에 입력되는 원음을 스피커로 출력하는 모델을 얻을 수 있다.



원음을 , 원음이 마이크에 의해 측정되어 파일로 변환되면서 생기는 변환을 , 소리파일이 스피커를 거치며 출력될 때 생기는 변환을 이라 하자.

원음 는 마이크로 들어간 뒤, 스피커로 출력되며 가 된다. 목표는 마이크로 입력된 에 적절한 변형 를 적용한 뒤 스피커로 나온 결과물 가 원음 에 근접하는 것이다. 이는 를 마이크로 측정한 결과인 을 원본 파일인 에 근사하도록 함으로써 달성할 수 있다.

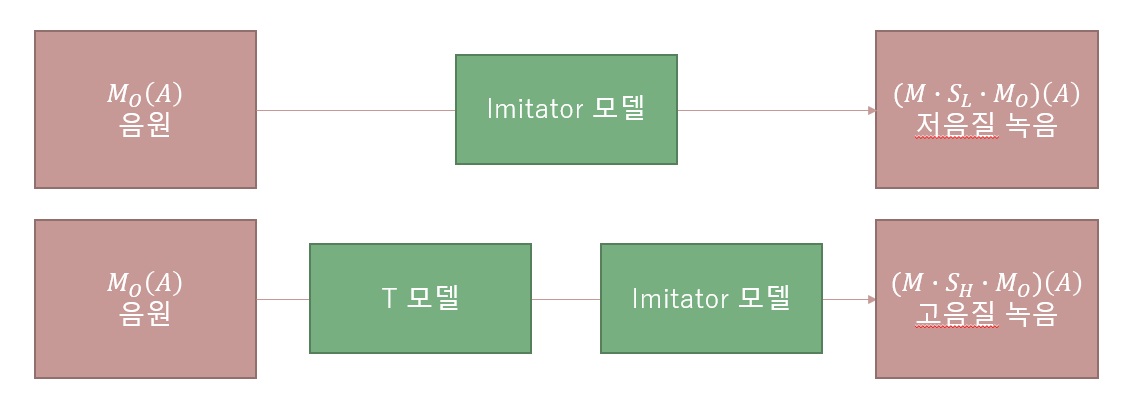
이를 딥러닝으로 구현한 방법은 다음과 같다. 우선 원본 파일 와 이를 스피커로 재생한 것을 녹음한 파일 을 데이터로 구한 후, 이를 이용해서 에 해당하는 변형을 근사하는 모델(Imitator)을 학습시킬 수 있다. 그 후, 다른 모델 를 이 모델의 앞에 붙여서 입력과 출력이 같도록 학습시키게 되면, 가 에 근사하게 된다. 고로 이 는 를 에 근사시키는 모델인 것이다.



이에 발전해서, 한 스피커가 다른 스피커의 변형을 모방하도록 만들 수도 있다. 이것은 원음이 녹음될 당시의 마이크의 변형 정도를 구할 수 없을 때 더 좋은 음질을 얻고자 하는 시도이다.

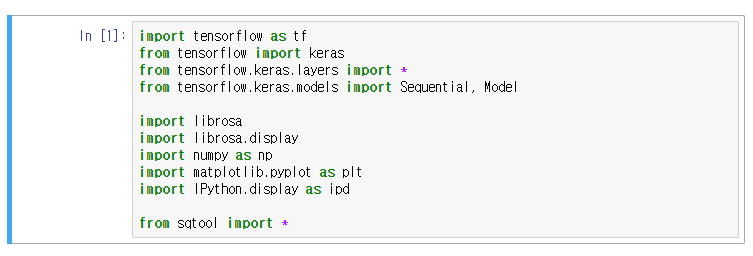
데이터셋 음원 원본을 , 저성능 스피커에 의한 변형을 , 고성능 스피커에 의한 변형을 라 하자. 이 모델을 학습시키기 위해서는 , 이를 각각의 스피커로 재생 후 마이크로 측정한 과 가 필요하다.

우선 Imitator 모델을 와 를 이용해 학습시킨다. 그 후, X는 , Y는를 주어 T 모델을 앞과 같은 방식으로 학습시키게 되면, 모델은 을 에 근사시키는 모델이 된다.

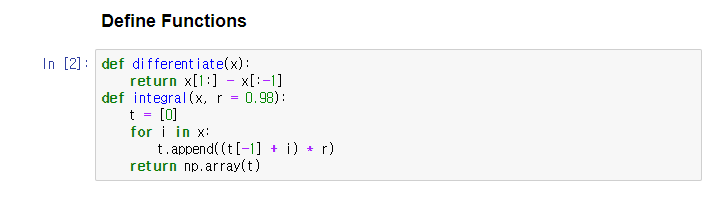
**

구현.

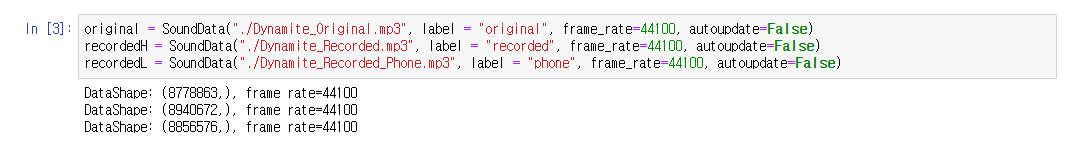
제시된 모델의 구현은 실제 코딩한 내용을 바탕으로 설명하겠다.



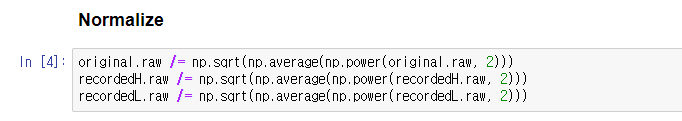
필요한 모듈들을 Import 한다.



이산적인 미분, 적분을 정의한다.



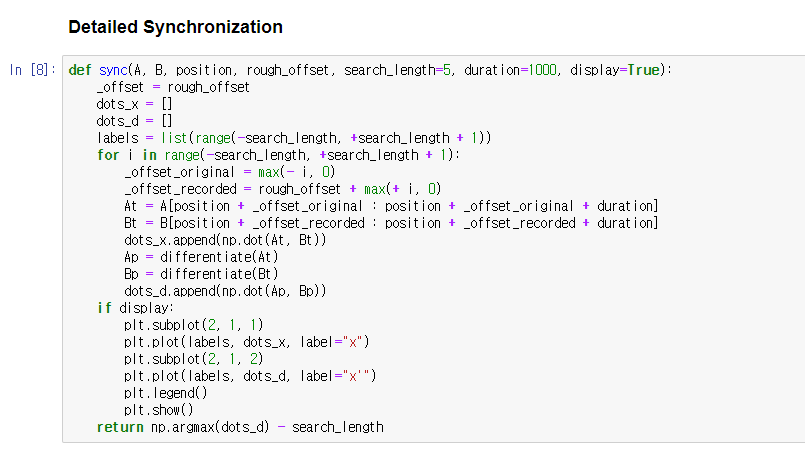
데이터를 불러온다. SoundData는 sqtool이라는 직접 만든 모듈의 클래스이다.

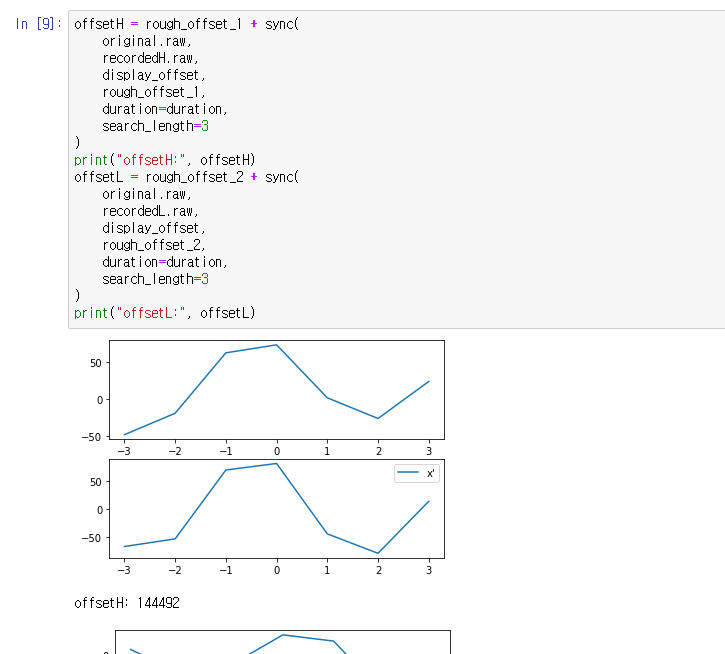


데이터의 Amplitude의 크기를 1시그마가 1이 되도록 맞춰준다.



세 개의 파일을 동기화하는 과정이다. rough\_offset\_1 은 사람에 의해 Manual 하게 정해져야 하는 값이며, 그래프를 확인하며 눈대중으로 두 파일이 같은 지점의 소리를 나타내는지 설정한다.





자동으로 세밀한 타이밍을 맞춰준다. 이는 두 파일의 일정 부분을 발췌해서 미분한 후, 내적한 값이 최대가 되게 하는 지점을 찾아주는 것이다.



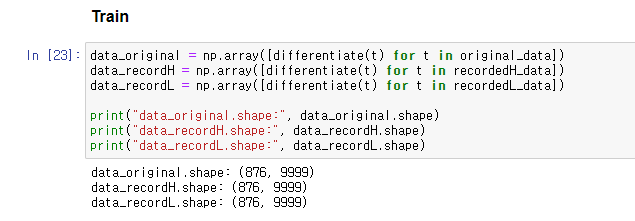
음악파일 타이밍을 맞추다보면, 갈수록 offset이 어긋나는 경향을 볼 수 있었다. 이를 부분부분 맞춰가며 데이터셋을 만드는 과정이다.



정확성을 위해 일차함수로 근사한 후에 sync를 다시 맞추어 준 후 정리하는 코드이다.



Keras로 모델을 build하는 과정이다.



데이터를 미분해 주었다. Raw 데이터보다 미분한 데이터가 fit이 잘 되는 것으로 보였다.



장기간 Train하기 위해 Imitator과 T를 반복해서 training해주는 모습이다.

결과.

잡음보다 더 작은 데시벨에서 정해지는 디테일 고음 부분이 치직 소리로 들리는 경향이 있었다. 이를 해결하기 위해 GAN을 도입해 보았지만, 성공적이지는 못했다. 하지만 저음질 스피커에서 고음질 스피커의 음색을 따라가려는 모습은 보였으며, 저음질 스피커에 의해 고주파 부분이 강조되는 경향이 부드러워지는 것을 볼 수 있었다. 전체적으로 강조해야 할 부분의 dB가 높다보니, 전체적인 음량은 줄어드는 것을 볼 수 있었다. 결과는 <https://www.youtube.com/watch?v=-Uudo4q3XKE> 에서 확인할 수 있다.

기대점.

이와 같은 방법을 잘 이용해 학습시킨다면, 마이크가 원음의 정보량을 해치지 않는다는 가정 하에 원음을 그대로 들을 수 있는 모델을 구현할 수 있다. 또한 많은 기기에서 사용자에게 조금 더 좋은 스피커의 음질로 소리를 들을 수 있도록 소프트웨어적으로 처리를 하도록 많은 기기에 적용할 수도 있을 것이다.