

Vocal Conversion

Audio Input Augmentation

Using 8-bit mu-law encoding

8-bit는 제가 켈 처음에 드렸던 자료에도 있듯 Amplitude를 256 단계로 쪼개는 것을 말합니다. (음질 지못미...) 그럼에도 불구하고 mu-law encoding을 사용하여 성능을 높였습니다.

Mu-law Encoding이란? 자주 나오는 구간은 더 잘게 쪼개고 자주 나오지 않는 구간은 더 넓게 쪼개서 전체적으로 음질은 보존하되, bit수는 유지하는 방법. 실제로 효과가 있었대요.

Music translation network와 다른점

Music Translation network에서는 특정 Audio clip내에서 input의 pitch(음정)를 locally하게 변화시켰는데, 여기서는 이게 안좋아서 사용하지 않았다구함!

(아무래도 노래를 부르게 하는 거니까 음정변화를 locally하게 하는게 안 좋아보이긴 하네요 중요해 보이지는 않습니다.)

Main Obstacle

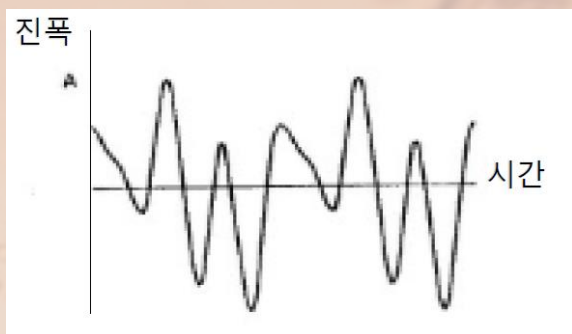
한 사람당 4~9개의 짧은 Short song으로만 되어있어서, data가 부족함. 그래서 더 많은 데이터를 생성(Data Augmentation) 하는 것이 Main Obstacle이다!

Augmentation을 하기 위한 방법1

1. 거꾸로 재생해도 Energy Spectrum은 변하지 않는다. 따라서 거꾸로 재생한다.

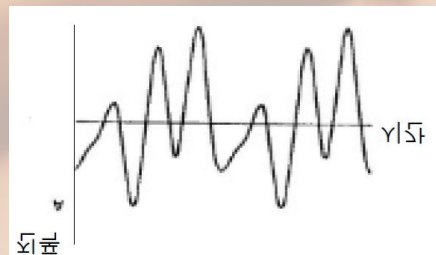
Augmentation을 하기 위한 방법2

2. 소리데이터는 다음과 같음



<- 여기서 진폭은 특정 점에서의 압력변화인
사람들은 음수나 양수나 중요하지 않음.

따라서 뒤집어서 한 번 더 쓰자



Augmentation

데이터가 그냥 재생해서 한번 거꾸로 재생해서 한 번, 뒤집어서(-1을 곱해서) 한 번 거꾸로 재생한 다음 뒤집어서 한 번해서 총 4배가 되었다. 짹짹짹짹짜까짜까짜까짜까