

---

---

---

---

---



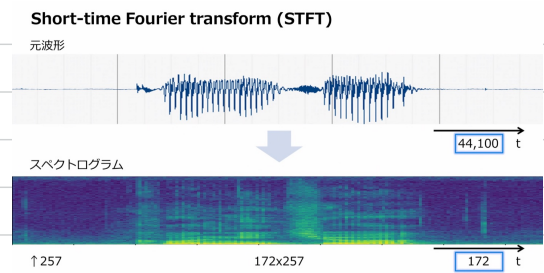
◦ ディープラーニングを行う。(時系列データを扱うリカレントニューラルネットワーク)

< 前処理 >

◦ 音声波形をスペクトログラム(画像)に変換する(前処理)



cf: 短時間フーリエ変換 (STFT)    44.1 k/s → 172/s に!



◦ 周波数変換(前処理)

◦ ダウンサンプリングを行う(隠れ層での処理, MaxPooling)

◦

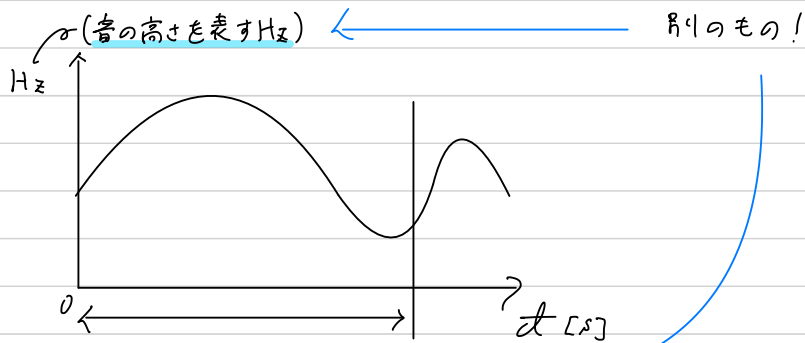
音声データの形式 = PCM (パルス符号変調)

振幅を1秒間に何回保存するか = サンプリングレート

例. 1秒間に1000回  $\Rightarrow$  1000サンプル. サンプル数はヘルツ [Hz] で表す為.

1000 Hz. または 1kHz と表す.

競技で用いられるものは. PCM 48kHz. (= 48k個/s)



$$PCM = \frac{\text{データ数}}{t [s]} = \sim \text{Hz}$$

$$\left( \begin{array}{l} = 1 \text{ 秒間に} \\ 1 \text{ つの音声を記録できるフレームが } \sim \text{ 個ある。} \\ \Rightarrow 1 \text{ フレームあたり } \frac{1}{\sim [Hz]} \text{ 秒} \end{array} \right)$$

$$16 \text{ bit} = 65536 \text{ 段階} \\ (2^{16})$$