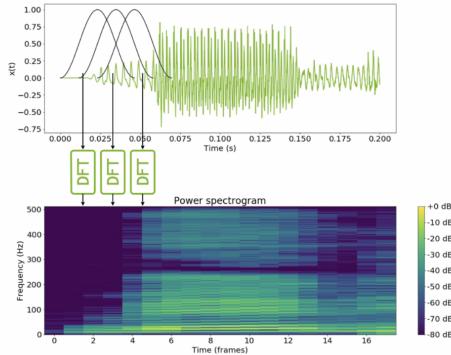


- Что такое мел-спектrogramма? Как она получается из wav-файла?  
У нас есть сигнал в зависимости от времени  
Идем окном по этому ряду  
Применяем преобразование Фурье  
Преобразовываем в мелы частоты  
Строим мел-спектограмму

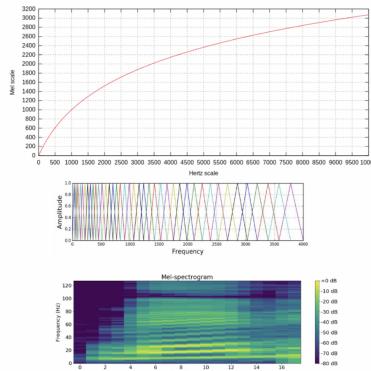
### Спектрограмма (та самая)

- Нарезаем сигнал на **окна с пересечением**
- Применяем **оконную функцию** к вырезанному окну
- Применяем дискретное преобразование **Фурье**
- Считаем **квадрат комплексной нормы**
- Берем **половину вектора** в силу его симметричности



### МелСпектrogramма

- Человеческое ухо хорошо слышит **низкие** частоты и плохо **высокие**
- Учитываем в спектrogramме **низкие** частоты больше, а **высокие** меньше
- $m = 2595 \log_{10} \left(1 + \frac{f}{700}\right) = 1127 \ln \left(1 + \frac{f}{700}\right)$   
 $f = 700 \left(10^{\frac{m}{2595}} - 1\right) = 700 \left(e^{\frac{m}{1127}} - 1\right)$
- В конце берем **логарифм** от мелспектrogramмы



- Какие метрики качества ASR вы знаете? Как они вычисляются?

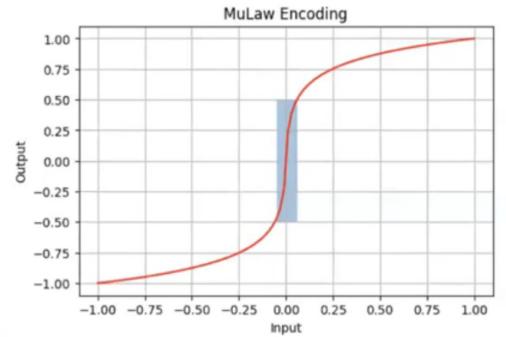
### Первым делом метрики!

- Word Error Rate (WER) -- доля "неправильных" слов
- $WER = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}$
- S** – кол-во замен  
**D** – кол-во удалений  
**I** – кол-во вставок  
**C** – кол-во совпадений  
**N** – **S** + **D** + **C** – кол-во слов всего
- True: quick **brown** fox jumped over **a** lazy dog  
Pred: quick **brow** **an** fox jumped over **red** lazy dog
- Character Error Rate (CER) -- тоже самое, но для символов (букв, например)
- Проще оптимизировать

- Что такое MuLaw-кодирование, для чего оно нужно и как оно вычисляется?

# Mu Law Encoding

- $f(x_t) = \text{sign}(x_t) \frac{\ln(1 + \mu|x_t|)}{\ln(1 + \mu)}$
- 16-bit WAV contain  $2^{16}$  values
- Softmax will die :(
- Human hearing on a **logarithmic** scale
- **Low-amplitude** sounds in **high** resolution
- **High-amplitude** sounds in **low** resolution



4. Как устроен WaveNet? Почему он медленно работает на этапе применения?

## WaveNet

$$p(\mathbf{x}) = \prod_{t=1}^T p(x_t | x_1, \dots, x_{t-1})$$

$$p(x_t | x_1, \dots, x_{t-1}) \sim \text{Cat}(\pi_\theta)$$

