

Лабораторная работа №3. Расчет Mel-Frequency коэффициентов.

Одним из информативных признаков аудиосигнала являются мел-кепстральные коэффициенты Звуки, воспроизводимые человеком, определяются формой голосового тракта, включая язык, зубы и т.д. Форма голосового тракта описывается огибающей спектра и задача расчета MFCC состоит в том, чтобы представить эту огибающую.

Шкала Мел соотносить высоту чистого тона (мел) с фактической измеренной частотой (Гц). Люди гораздо лучше различают небольшие изменения высоты звука на низких частотах, чем на высоких. Эта зависимость не совсем линейная и описывается следующей формулой (преобразование частоты в мел):

$$M(f) = 1127,01048 \ln(1 + f/700).$$

Обратное преобразование:

$$M^{-1}(m) = 700(e^{m/1127,01048} - 1).$$

Алгоритм расчета мел-частотных кепстральных коэффициентов:

1. Разделить исходный сигнал на фреймы. Размер от 20 до 40 мс.

Считается, что речевой сигнал на этом промежутке не сильно меняется.

$x(n), 0 \leq n \leq N$, где N - размер фрейма или длина окна, $x_j(n)$ - j -ый фрейм. Далее применяем все для каждого фрейма по отдельности.

2. Речевой сигнал конечен и не является периодическим, поэтому из-за разрывов на его концах при применении преобразования Фурье проявляется эффект утечки. Для того, чтобы снизить его влияние на результат, каждый кадр умножается на оконную функцию Хемминга:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1.$$

К получившемуся результату применяем дискретное преобразование

Фурье: $X_j(k) = \sum_{n=0}^{N-1} x_j(n)w(n)e^{-\frac{2\pi i}{N}kn}, 0 \leq k \leq N$, j - номер фрейма.

3. Вычисляем периодограмму для каждого фрейма:

$$P_j(k) = \frac{|X_j(k)|^2}{N}$$

4. Вычисляем блок мел-фильтров. Для этого треугольные фильтры (от 20 до 40) умножаются на периодограмму и суммируются. В

результате мы получим энергии набора фильтров. Каждый треугольный фильтр моделируются с помощью следующей функции:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \leq k < f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, & f(m) \leq k \leq f(m+1) \\ 0, & k > f(m+1) \end{cases},$$

где m – это число фильтров, которое мы хотим получить. Зная число фильтров (обычно 26) и диапазон интересующих нас частот, функции $f(\cdot)$ можно найти, используя формулы прямого и обратного мел преобразования.

- Полученные энергии логарифмируют. Это также мотивируется человеческим слухом: мы не слышим громкость в линейном масштабе. Обычно, чтобы удвоить воспринимаемую громкость звука, нам нужно затратить в 8 раз больше энергии. Это означает, что большие колебания энергии могут звучать не так уж и по-другому, если звук с самого начала громкий. Эта операция сжатия делает наши функции более близкими к тому, что на самом деле слышат люди. Мы получаем некоторый набор коэффициентов, которые еще не являются MFCC:

$$S_j(m) = \ln \sum_{k=0}^{N-1} P_j(k) H_m(k), 0 \leq m \leq M.$$

- Далее, используя дискретное косинусное преобразование, получим мел-кепстральные коэффициенты:

$$c_j(n) = \sum_{m=0}^{N-1} S_j(m) \cos(\pi n(m + 1/2)/M), 0 \leq n < M$$

Задание:

- Загрузить [датасет](#) с аудиозаписями, сделанными людьми разного пола
- Реализовать функцию расчета MFCC
- Для случайных 10 файлов из папки каждого пола рассчитать MFCC