# Методика расчёта мел-частотных кепстральных коэффициентов

1. Генерируются треугольные фильтры в соответствии со шкалой мел. Генерация фильтров происходит в несколько этапов.
   1. Сначала определяются границы по частоте, в которых будут генерироваться фильтры. Например, от 300 до 8000 Герц.
   2. Затем эти границы переводятся в шкалу мел. Для преобразования частоты в шкалу мел используется формула

Например, частотные границы 300 и 8000 Герц в шкале мел будут равны 401.97 и 2840.04 соответственно.

* 1. В указанных границах генерируются N + 2 опорных точек, где N – количество генерируемых фильтров. Например, если нужно получить 10 фильтров, необходимо сгенерировать 12 опорных точек на равном расстоянии друг от друга. В нашем примере должны получиться следующие 12 опорных точек:

[401.97, 623.61, 845.26, 1066.90, 1288.54, 1510.18, 1731.83, 1953.47, 2175.11, 2396.75, 2618.40, 2840.04]

* 1. Каждая точка переводится из шкалы мел обратно в частоту в Герцах. Для этого используется формула

В нашем примере должны получиться такие значения:

[300.0, 517.33, 781.91, 1103.98, 1496.05, 1973.33, 2554.37, 3261.65, 4122.66, 5170.79, 6446.78, 8000.02]

* 1. Значения переводятся из Герц в бины Фурье, так как фильтры впоследствии будут применяться к спектру в бинах, полученному с помощью быстрого преобразования Фурье. Для перевода герц в бины Фурье используется формула

– исходная частота

– размер блока при преобразовании Фурье. Он соответствует количеству отсчётов сигнала в одном блоке (см. пункт 2).

– частота дискретизации сигнала. В нашем случае она равна 48 кГц.

После перевода частот в бины Фурье должны получиться такие числа (для примера fftSize равен 1024, sampleRate равняется 48000):

[6, 11, 16, 23, 31, 42, 54, 69, 87, 110, 137, 170]

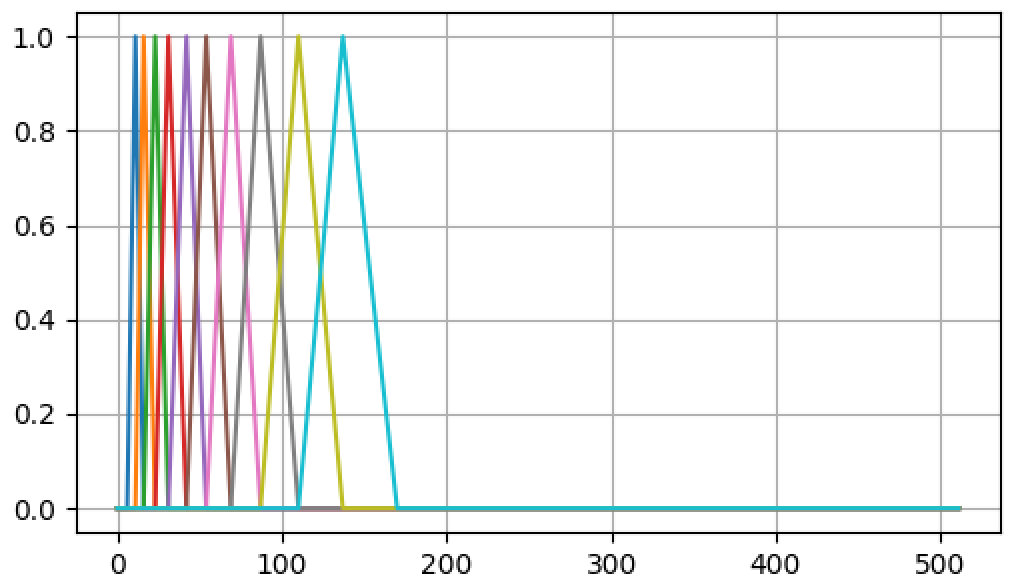
* 1. Вычислить значения фильтров по формуле

Здесь – это номер фильтра. Если генерируются 10 фильтров, то это будет число от 1 до 10.

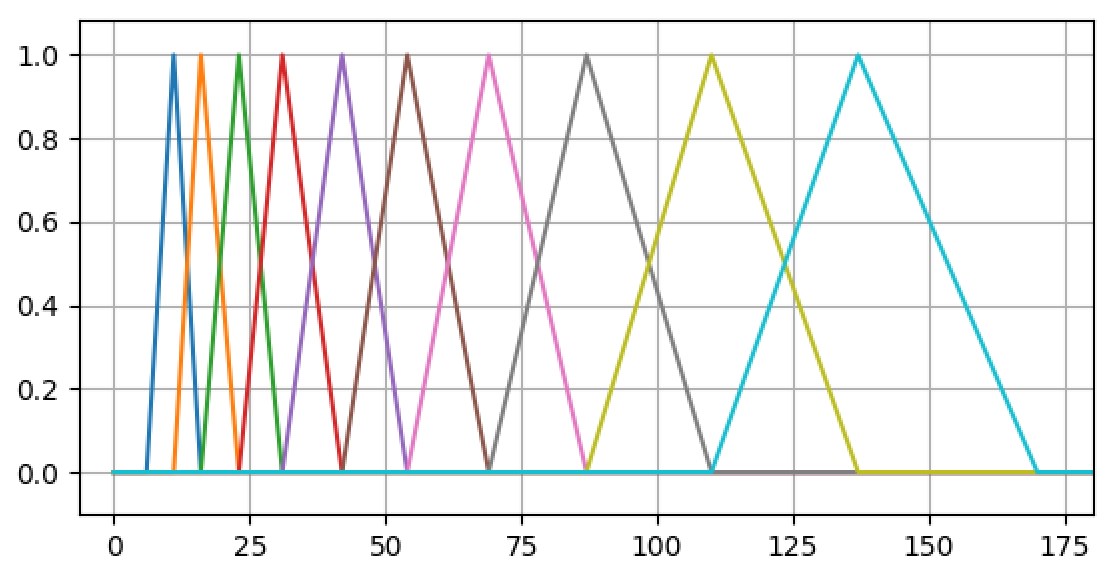
– значение m-го фильтра в точке .

– значения опорных точек фильтров в бинах Фурье, посчитанные на предыдущем этапе.

В результате должно получиться 10 треугольных фильтров, которые выглядят так, если их изобразить на графике:



Или увеличенная версия:



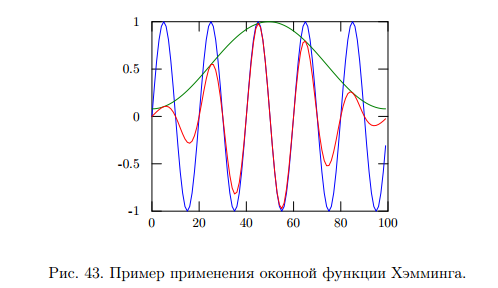
1. Исходный сигнал разбивается на блоки величиной от 20 до 40 мс. Дальнейшие вычисления выполняются с этими блоками, которые могут пересекаться.



Например, если сигнал разбивается на блоки по 25 мс, а частота дискретизации сигнала 48 кГц, то на один блок будет приходиться 1200 отсчётов сигнала:

1. Блок сигнала может быть поэлементно перемножен на оконную функцию, например, окно Хемминга, чтобы уменьшить растекание спектра. Окно Хемминга описывается формулой

Рисунок ниже демонстрирует применение окна Хемминга:



Синий график – исходный сигнал. Зелёный график – окно Хемминга. Красный график – результат применения окна Хемминга к исходному сигналу (поэлементное перемножение).

1. Вычисление амплитудного спектра блока сигнала

Для этого обычно используют быстрое преобразование Фурье (БПФ). Английская аббревиатура FFT (Fast Fourier Transform). Полученный спектр будет комплексным, и чтобы получить из него амплитудный спектр, необходимо вычислить модуль комплексных чисел, составляющих спектр. Затем полученные числа возводятся в квадрат.

На языке Python с использованием библиотеки NumPy код вычисления спектра может выглядеть следующим образом:

sp = (numpy.square(numpy.absolute(numpy.fft.rfft(block, n=fftSize))) / fftSize)[:fftSize//2]

block – блок сигнала

fftSize – размер блока БПФ (количество отсчётов в блоке сигнала).

1. Полученный амплитудный спектр также может быть умножен на оконную функцию.
2. Для вычисления коэффициентов необходимо спектр поэлементно перемножить на каждый из фильтров, которые были вычислены на этапе 1. Например, если необходимо вычислить 10 коэффициентов, и сгенерировано 10 фильтров, то нужно будет перемножить спектр на каждый из 10 фильтров. Полученные числа складываются, и от полученного значения берётся натуральный логарифм (см. формулу):

– k-е значение спектра.

– k-е значение m-го фильтра.

– количество рассчитываемых коэффициентов.

1. Применить дискретное косинусное преобразование (ДКТ) к вычисленным на предыдущем этапе коэффициентам.

Полученные коэффициенты и есть искомые MFCC. Значения должны совпадать с тем, что даёт функция mfcc\_feat из библиотеки python\_speech\_features.