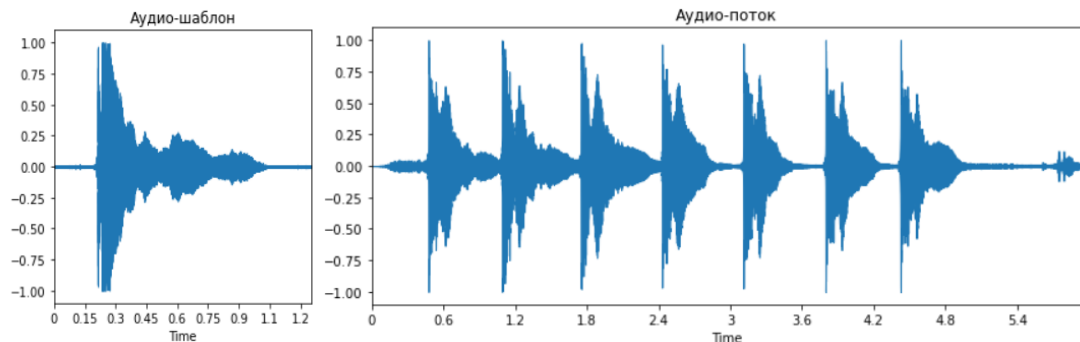
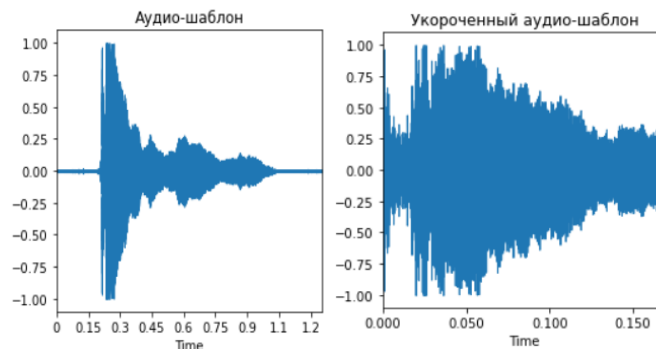


Описание метода

Обнаружение участков аудио-потока схожего по звучанию с шаблоном производится за счет использования скользящего окна вдоль временной оси.

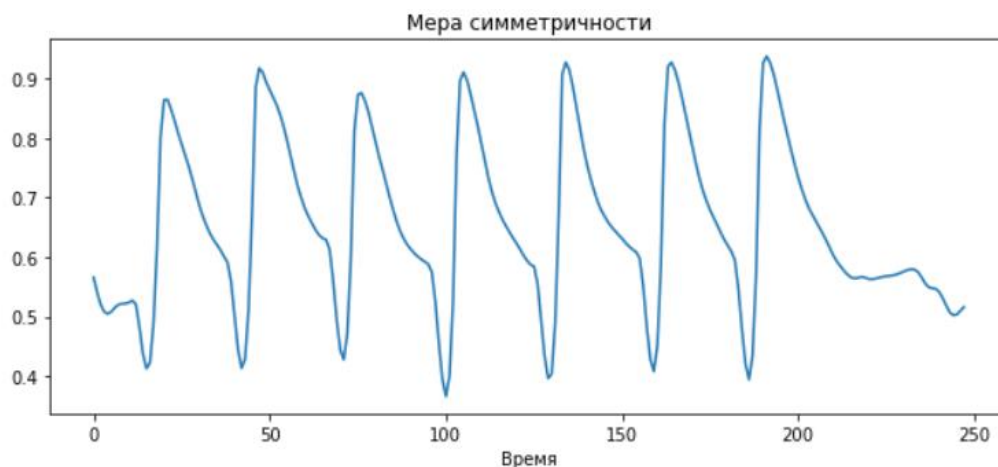


Сначала аудио-шаблон обрезается, так как помимо основного звука он также содержит и участки тишины в начале и в конце.

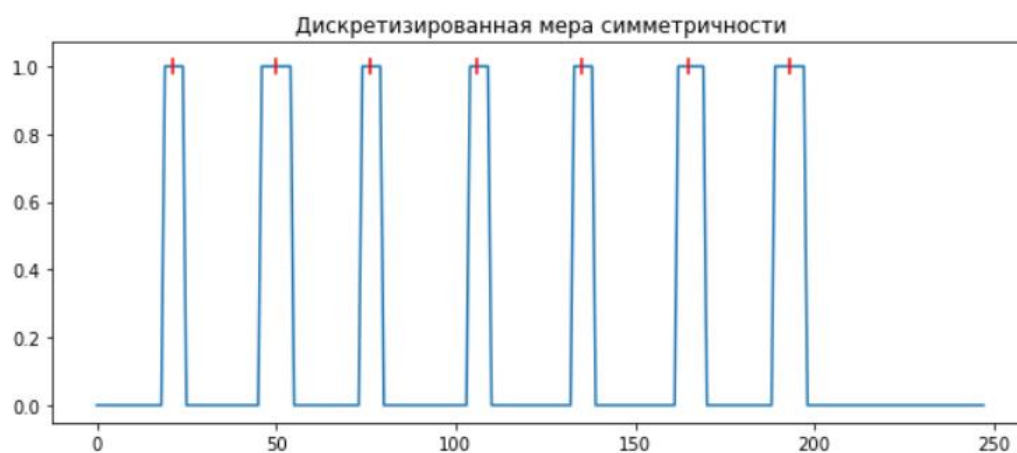


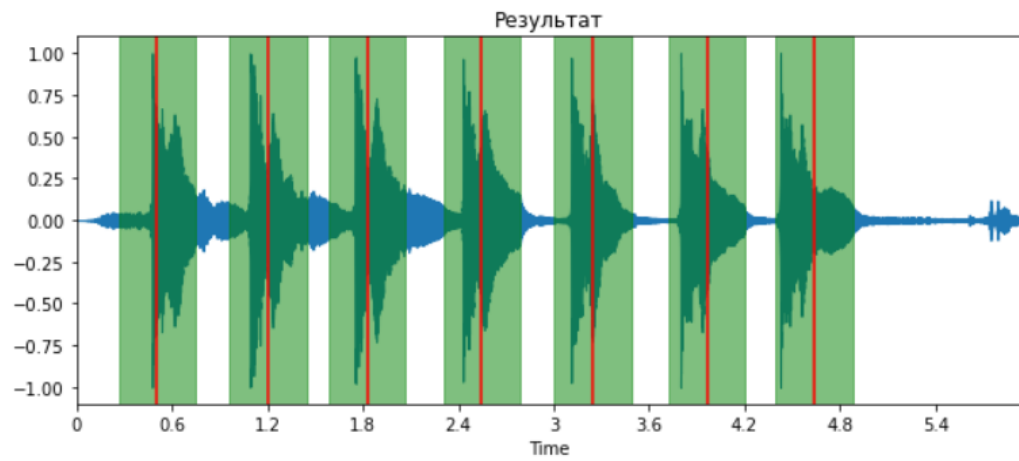
После этого шаблон и аудио-поток преобразовываются в признаки с помощью функции `librosa.feature.mfcc`, которая возвращает двумерный массив ($n_mfcc \times time$). Массив признаков шаблона проходит вдоль массива аудио-потока как окно. Мера схожести между признаками шаблона и потока вычисляется как поэлементное произведение двух матриц с последующим вычислением суммы всех произведений. Непосредственно до перемножения обе матрицы нормируются (делятся на норму вектора, получаемого из соответствующей матрицы выписыванием подряд всех строк). Таким образом, мера схожести есть не что иное, как косинус угла между двумя признаковыми векторами и чем ближе это значение к 1, тем более схожи эти вектора в признаковом пространстве, а соответственно и их матрицы тоже. Однако, чтобы не включить в результат ложные шумовые участки схожие по форме, до нормировки проводится сравнение максимального значения участка потока внутри окна и медианного значения шаблона.

Участки аудио-потока, которые имеют сильную схожесть с шаблоном, сохраняют ее и при небольших сдвигах вправо и влево. Отсюда вытекает тот факт, что график меры схожести получается гладким, как на рисунке ниже.



Чтобы получить участки локальных максимумов, мера схожести дискретизируется с помощью сравнения с пороговым значением. В участках полученной функции, где она имеет значение 1, осуществляется поиск середины, которая далее выступает как точка совпадения аудио-потока и шаблона. Недостатком такого подхода может быть смещенность найденной точки, так как она по сути является грубым приближением локального максимума.





Данный реализация не применима, в случае если аудио-шаблон сильно отличается от звуков в потоке или полностью отличается, так как возможны ложные срабатывания.