

РАСПОЗНАВАНИЕ АКУСТИЧЕСКИХ ОТПЕЧАТКОВ В СИГНАЛЕ В РЕАЛЬНОМ ВРЕМЕНИ

И. Ю. Лебедев (Университет ИТМО, г. Санкт-Петербург)

Научный руководитель — Н. А. Осипов (Университет ИТМО, г. Санкт-Петербург)

Одной из проблем современной программной инженерии является определение и распознавание заранее определённых паттернов в звуковом сигнале. Это требуется для таких задач, как автоматизированное обнаружение защищённого авторскими правами аудио-контента, а также распознавание музыкальных произведений в реальном времени.

Помимо этого, эти технологии позволяют анализировать звуковой фон в определённой среде наблюдения. Например, происходящее вокруг пользователя мобильного приложения, при условии наличия доступа к микрофону его устройства – мы можем узнавать о «ключевых звуках» в момент их происхождения

Рассмотрим, как работает данная технология. Любой звуковой сигнал, в цифровом виде, представляет собой последовательность закодированных амплитуд, отмеренных с постоянным интервалом (частота дискретизации). Но в таком виде невозможно определить сходство двух сигналов, если они будут отличаться по громкости, или если один из них будет содержать сторонние шумы. Поэтому строится спектрограмма (с помощью преобразования Фурье), содержащая также информацию и о амплитудах отдельных частот в сигнале.

Далее, создаётся акустический отпечаток, который уже можно использовать для определения сходства нескольких сигналов. Существуют различные алгоритмы построения акустических отпечатков на основе спектрограмм, от простых и распространяемых бесплатно (например, библиотека Dejavu на Python), до алгоритмов high-end уровня, используемых крупными корпорациями.

Для применения технологии в реальном времени, необходимо постоянно буферизовать звуковой сигнал. Полученный буферизованный сегмент также имеет свой акустический отпечаток, который можно будет сравнить с имеющимися (созданными заранее) отпечатками, чтобы сделать вывод о их сходстве или различии. При этом, де-факто, распознавание будет происходить не в реальном времени, а с задержкой (равной сумме длительности буфера и времени на обработку), но для прикладного применения это является приемлемым.

Таким образом, данная технология имеет свою область применения, у неё имеется большой потенциал в сфере пользовательской аналитики, а наличие open-source реализаций делает её доступной всем желающим.