

СИСТЕМА ОЦЕНКИ КАЧЕСТВА ЦИФРОВОГО КАНАЛА РЕЧЕВОЙ СВЯЗИ

Вашкевич И.А., студент

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Азаров И.С. – д-р. техн. наук, профессор

В работе описывается система, позволяющая предположить какой из двух данных аудио-поток является оригиналом, а также методы их сжатия путём оценки ряда относительных характеристик, используя известные особенности популярных кодеков и полученную в процессе исследования базу данных.

Сегодня большинство аудио-поток в сети интернет передаются при помощи кодеков сжатия с потерями. Это означает, что аудио-поток могут быть сжаты несколько раз, что вызывает деградацию. Чтобы уменьшить вероятность использования записей низкого качества или определить вероятность вмешательства в сообщения речевого канала связи, например, склеек и ретрансляций, нужна система оценки качества цифрового канала без опоры на психоакустическую модель человека.

Алгоритм выполняется в несколько этапов: подготовка сигнала, анализ и работа с базой данных.

Подготовка состоит из загрузки двух аудиофайлов и их преобразования в формат WAV с использованием программы *FFmpeg*, преобразование выполняется в монофонический аудио-сигнал с частотой дискретизации 44,1 кГц.

Синхронизация сигналов выполняется методом кросс-корреляции, позволяющим определить временной сдвиг между сигналами. После этого сигналы выравниваются по времени и обрезаются до общей длины. Нормализация громкости выполняется в частотной области на основе кратковременного преобразования Фурье. Для каждого сигнала вычисляется спектральное представление, затем определяется общая частотная полоса, в которой оба сигнала обладают значимой энергией. В этой полосе рассчитывается среднеквадратичная энергия для сигналов (*RMS*), после чего вычисляется коэффициент усиления на основе отношения полученных энергий. Такой подход минимизирует влияние потерь энергии, возникающих при сжатии с потерями.

Следующим этапом является анализ относительных характеристик двух аудио-сигналов. Сначала определяется временная маска, отражающая эффект маскировки. Для этого анализируется разностный сигнал, для которого вычисляется огибающая через скользящее *RMS*-окно:

$$E(n) = \sqrt{\frac{1}{W} \sum_{k=0}^{W-1} d[n-k]^2}, \quad (1)$$

Резкие изменения огибающей выявляются через её производную и пороговое значение, используя среднее и стандартное отклонение производной. Обнаруженные сегменты соответствуют кратковременным энергетическим всплескам, за которыми могут следовать затишья, характерные для временного эффекта маскировки [1]. Эти области исключаются из дальнейшего анализа маской.

Далее выполняется спектральный анализ. Для этого вычисляется отношение полных энергий спектра, а затем распределение энергии по частотам каждого сигнала:

$$E(f) = \frac{1}{T} \sum_t |X(f, t)|^2, \quad (2)$$

Также находится характеристика на основе пересечения энергий спектров:

$$I = \frac{\sum_f \min(E_a(f), E_b(f))}{\sum_f E_a(f)}, \quad (3)$$

Помимо интегральных характеристик, дополнительно анализируется форма спектра. Для этого вычисляется отношение энергии в верхнем диапазоне частот (выше 10 кГц) и частота спектрального спада, отношение этих частот для двух сигналов позволяет оценить потерю высокочастотной составляющей. Частота спектрального спада определяется как минимальная частота, при которой накапливается заданная доля энергии сигнала:

$$\sum_0^{f_r} E(f) \geq p \cdot \sum E(f), \quad (4)$$

где p принимаем равным 0.95. Дополнительно спектр разбивается на несколько диапазонов (низкие, средние и высокие частоты), в которых вычисляются относительные изменения энергии. Локальный анализ повышает вероятность выявить характерные искажения, вносимые кодеками сжатия.

На следующем этапе вычисляются ошибки спектрального представления. Амплитудная ошибка определяется как взвешенное среднеквадратичное отклонение модулей спектра:

$$E_{amp} = \sqrt{\frac{\sum_f w(f) \cdot \frac{1}{T} \sum_{t=1}^T (|X_a(f, t)| - |X_b(f, t)|)^2}{\sum_f w(f)}} \quad (5)$$

где $w(f) = \min(E_a(f), E_b(f))$. Фазовая согласованность оценивается через когерентность:

$$C_{ab}(f) = \frac{|P_{ab}(f)|^2}{P_{aa}(f) \cdot P_{bb}(f)}, \quad (6)$$

Результат оценки когерентности сохраняется как взвешенное среднее:

$$C = \frac{\sum_f w(f) C_{ab}(f)}{\sum_f w(f)}, \quad (7)$$

В качестве дополнительной асимметричной характеристики используется отношение энергий сигналов. Данная метрика основана на предположении, что кодек с потерями не способен значительно изменить энергию сигнала, а значит она может служить индикатором потери информации или общих искажений.

После извлечения признаков формируется массив характеристик, который может быть сохранён в базе данных вместе с метаданными (тип кодека и битрейт). В противном случае производится поиск наиболее близких записей. Для этого используется взвешенное нормализованное евклидово расстояние:

$$D = \sqrt{\sum_k w_k \left(\frac{x_k - y_k}{|x_k| + \epsilon} \right)^2}, \quad (8)$$

где w_k – веса признаков. Наибольший вклад в расстояние вносят характеристики спектрального покрытия и высокочастотных искажений. Поиск осуществляется по принципу ближайших соседей, и пользователю выводятся три наиболее вероятных варианта преобразования.

Для оценки эффективности алгоритма была собрана база данных на основе 3-ёх различных образцов длительностью около одной минуты в 4-ёх распространённых битрейтах (от англ. *bit rate*) и двух кодеках, комбинируя до двух преобразований на образце (например, из WAV в MPEG-1 Layer 3, а затем в OGG Vorbis), исключая повышение битрейта в комбинации. Тестирование было проведено на отличных образцах в таких же комбинациях параметров.

В результате тестирования алгоритм смог верно определить параметры 21 из 24 тестовых образцов в тесте без комбинирования 41 из 144 с комбинациями. Данный результат говорит о наличии потенциала у подобных систем, однако ручной подбор весов и низкая эффективность некоторых признаков означают необходимость дальнейшей доработки алгоритма для получения более точных результатов.

Список использованных источников:

1. Основы психоакустики. Слуховая маскировка [Электронный ресурс] – Режим доступа: <https://web.archive.org/web/20051102163622/http://www.625-net.ru/archive/z0200/3.htm>
2. Vorbis I specification [Электронный ресурс] – Режим доступа: https://xiph.org/vorbis/doc/Vorbis_I_spec.html
3. The Theory Behind Mp3. [Электронный ресурс] – Режим доступа: https://reynal.etis-lab.fr/docs/audio-sia/tp/mp3/mp3_theory.pdf
4. Signal Processing Course. Chapter 7 [Электронный ресурс] – Режим доступа: <https://www.fil.ion.ucl.ac.uk/~wpenny/course/array.pdf>