

ПРИМЕНЕНИЕ РАЗРЕЖЕННОЙ АППРОКСИМАЦИИ С ПЕРЦЕПТУАЛЬНО ОПТИМИЗИРОВАННЫМ СЛОВАРЕМ ВЕЙВЛЕТ КОЭФФИЦИЕНТОВ ДЛЯ ЗАДАЧ СЖАТИЯ АУДИСИГНАЛОВ

В.Ю. ГЕРАСИМОВИЧ¹, Ал.А. ПЕТРОВСКИЙ²

*Белорусский государственный университет информатики и радиоэлектроники
ул. П. Бровки, 6, г. Минск, 220013, Республика Беларусь
¹gerasimovich@bsuir.by, ²alexey@petrovsky.eu*

На современном этапе развития алгоритмов компрессии аудиосигналов разработано большое количество кодеров, которые используют те, или иные особенности входного сигнала. Такое большое количество разнообразных алгоритмов ставит в затруднительное положение при их выборе для использования в конкретной задаче. Оптимальным вариантом решения данной проблемы служит разработка универсального масштабируемого аудиокодера, способного эффективно работать с большинством типов входных сигналов.

Ключевые слова: частотно-временные преобразования, согласованная подгонка, аудиокодер.

Большинство окружающих нас аудиосигналов являются нестационарными. Нестационарным сигналом называют такой сигнал, спектральные характеристики которого изменяются во времени. Изменчивая во времени природа сигнала ставит некоторые трудности в его анализе и работе с ним. Классическими методами анализа сигнала являются анализ сигнала во временной и частотной областях. Анализ сигнала во временной области позволяет получить некоторую информацию о его характеристиках, однако не дает никакой информации о частотных характеристиках. Переход в частотную область с помощью преобразования Фурье, дает глобальное представление о спектральном составе сигнала, что неприменимо для анализа нестационарных сигналов. Частотно-временной анализ сигнала позволяет получить совмещенную частотно-временную картину характеристик сигнала.

В общем случае, частотно-временные преобразования можно разделить на два класса: декомпозиция сигнала и билинейное частотно-временное распределение. Частотно-временное распределение представляет собой двумерное энергетическое представление сигнала с высоким частотно-временным разрешением. В подходах, основанных на декомпозиции, происходит аппроксимация сигнала частотно-временными функциями, полученными перемещением, модуляцией и масштабированием базисных функций, имеющих определенную временную и частотную локализацию.

Декомпозиция сигнала основана на алгоритме согласованной подгонки со словарем частотно-временных функций. В данном подходе, любой сигнал $x(t)$ представляется в виде линейной комбинации частотно-временных функций (называемых атомами) $g_{T_n}(t)$, выбираемых из избыточного словаря D . Любой сигнал можно разложить с помощью алгоритма частотно-временной декомпозиции следующим образом: $x(t) = \sum_{n=0}^{\infty} a_n g_{T_n}(t)$, где a_n – масштабирующий коэффициент, который показывает вклад атома в формирование выходного сигнала.

Алгоритм согласованной подгонки является жадным алгоритмом и подразумевает поиск локально оптимальных решений с расчетом, что глобальное решение также будет оптимальным. Выбор аппроксимирующей функции из словаря заключается в поиске такой функции, которая дает максимальное значение скалярного произведения с

фреймом анализируемого сигнала. После выбора найденной функции происходит синтез сигнала по ней и вычитание из основного сигнала, что дает нам остаточный сигнал. Следующая итерация алгоритма производится над найденным остаточным сигналом. В идеале, остановка алгоритма происходит в том случае, когда остаточный сигнал равен нулю (однако, как правило, вводятся определенные пороги, которые являются признаком остановки).

При использовании алгоритма согласованной подгонки нужно учитывать два основных взаимосвязанных нюанса: выбор частотно-временных функций, формирующей словарь, и вычислительные затраты при работе алгоритма. Так как словарь должен быть избыточным, он может содержать большое множество различных базисных функций, однако, с другой стороны, чрезмерно объемный словарь увеличит время поиска подходящей функции. Следовательно, стоит вопрос о выборе оптимального словаря частотно-временных функций.

Существующие подходы в построении аудиокодеров на основе разреженной аппроксимации [1, 2] имеют основной недостаток: словарь атомов является фиксированным для каждого сигнала, что не может являться оптимально эффективным в силу того, что тип сжимаемых данных может сильно отличаться в различных сигналах. Для решения этой проблемы можно использовать большее количество элементов в словаре, однако это приведет к резкому возрастанию времени поиска по словарю и увеличению вычислительных затрат. Так же, этот вариант может не дать эффективного решения в силу нестационарности реальных аудиосигналов и их огромного разнообразия.

Анализируя вышесказанное можно выделить ряд требований, предъявляемых к разрабатываемому кодеру: масштабируемость, универсальность, низкая скорость битового потока, высокое качество восстановленного сигнала. В качестве математической модели описания сигнала решено было выбрать модель разреженной аппроксимации сигнала. Данная модель позволяет описать входной аудиосигнал минимальным количеством параметров. В отличие от существующих подходов, словарь атомов будет формироваться из самого сигнала для каждого входного фрейма индивидуально. Это позволит добиться максимальной гибкости и оптимальной эффективности алгоритма. Еще одним шагом оптимизации работы кодера будет являться тот факт, что словарь вейвлет коэффициентов будет формироваться на основе перцептуально-оптимизированного ПДВД вместо использования полного дерева декомпозиции [3]. Это позволит исключить маскируемые компоненты, не воспринимаемые при прослушивании восстановленного аудиосигнала.

Список литературы

1. *Umapathy K., Ghoraani B., Krishnan S.*, Audio Signal Processing Using Time-Frequency Approaches: Coding, Classification, Fingerprinting and Watermarking // *EURASIP Journal on Advances in Signal Processing*, 2010. Volume 2010. 28 p.
2. *Wenxin He, Tianshu Qu*, Audio Lossless Coding/Decoding Method Using Basis Pursuit Algorithm // *International Conference on Acoustics, Speech and Signal Processing*. Vancouver, 26-31 may 2013. P. 552-555.
3. *Petrovsky Al., Petrovsky A.*, Matching pursuit algorithm with frame-based auditory optimized WP-dictionary for audio transient modelling // *Electronica. Konstrukcje, technologie, zastosowania*, nr4/2008. P. 74-80.