

# КОДИРОВАНИЕ АУДИОСИГНАЛОВ НА ОСНОВЕ РАЗРЕЖЕННОЙ АППРОКСИМАЦИИ ДЛЯ ВСТРАИВАЕМЫХ СИСТЕМ

А.А. Петровский, В.Ю. Герасимович  
Кафедра электронных вычислительных средств,  
Белорусский государственный университет информатики и радиоэлектроники  
Минск, Республика Беларусь  
E-mail: alexey@petrovsky.eu, gerasimovich@bsuir.by

Данная работа посвящена исследованию возможности применения алгоритмов разреженной аппроксимации для компрессии аудиосигналов. Дается краткое описание алгоритма согласованной подгонки для декомпозиции сигнала на атомы. Проводится обзор работ, посвященных разработке алгоритмов компрессии аудиосигна на основе разреженной аппроксимации.

## ВВЕДЕНИЕ

Компрессия аудиосигналов является актуальной проблемой обработки мультимедиа информации. Эффективное сжатия информации требуется в таких задачах, как передача сигналов по коммуникационным каналам в реальном масштабе времени, хранении аудиоданных на носителях информации, работа с портативными мультимедиа устройствами. На данном этапе развития алгоритмов кодирования аудиоданных существует большое количество подходов, использующих определённые особенности обрабатываемых сигналов, и имеющих свои плюсы и минусы. Анализируя этот факт можно увидеть, что естественной задачей является разработка универсального аудиокодера, способного эффективно работать с различными типами входной информации.

На сегодняшний день, существуют разработки подобных универсальных аудиокодеров. Их ядром служит алгоритм разреженной аппроксимации для параметризации сигнала. Он позволяет произвести декомпозицию сигнала и представить его малым количеством параметров. В данной работе даётся краткое описание математической модели и делается сравнительный обзор существующих подходов в построении аудиокодеров на основе разреженной аппроксимации. В проведённом исследовании были выделены общие характерные черты, а также особенности и недостатки рассмотренных кодеров.

### I. АЛГОРИТМ СОГЛАСОВАННОЙ ПОДГОНКИ

Декомпозиция сигнала основана на алгоритме согласованной подгонки (*Matching pursuit – MP*) со словарем частотно-временных функций [1]. В данном подходе, любой сигнал  $x(t)$  представляется в виде линейной комбинации частотно-временных функций (называемых атомами)  $g_{\gamma_n}(t)$ , выбираемых из избыточного словаря  $D$  (содержит намного больше элементов, нежели минимальное необходимое количество базисных функций, покрывающих данное пространство). Любой сигнал мож-

но разложить с помощью алгоритма частотно-временной декомпозиции следующим образом:  $x(t) = \sum_{n=0}^{\infty} a_n \cdot g_{\gamma_n}(t)$ , где  $a_n$  – масштабирующий коэффициент, который показывает вклад атома в формирование выходного сигнала.

Согласованная подгонка – жадный итеративный алгоритм. На каждой итерации происходит выбор атома, который наилучшим образом аппроксимирует сигнал, а затем вычитается его вклад в формирование исходного сигнала:  $r_{i+1}[n] = r_i[n] - \alpha g_{\gamma_n}$ , где  $r_i[n]$  – остаточный сигнал на  $i$ -й итерации (для первой итерации алгоритма остатком является входной сигнал).

При использовании алгоритма *MP* возникает задача выбора оптимального словаря частотно-временных функций. Одним из самых эффективных вариантов является формирование словаря на основе анализируемого сигнала. Для этого выполняется разложение входного сигнала по вейвлет функциям. На основе полученных функций формируется словарь атомов. Избыточный словарь можно сформировать, применив полное дерево пакета дискретного вейвлет преобразования (ПДВП). Улучшить производительность алгоритма согласованной подгонки может перцептуальная оценка узлов дерева декомпозиции для каждого фрейма индивидуально. Это позволяет получить оптимальное дерево разложения ПДВП на каждом фрейме анализируемого сигнала, что, в свою очередь позволит добиться минимизации словаря  $D$  [2].

### II. АУДИОКОДЕРЫ НА ОСНОВЕ РАЗРЕЖЕННОЙ АППРОКСИМАЦИИ

В подходе, представленном в [3] происходит декомпозиция входного сигнала на частотно-временные атомы, взятые из заранее предопределенного избыточного словаря. Критерием остановки работы алгоритма служит анализ энергии сигнала, захваченной выбранной из словаря частотно-временной функцией. Таким образом, декомпозиция входного сигнала завершается либо при достижении предельного количества итераций работы алгоритма, либо при захвате 99,5%

энергии входного сигнала отобранными атомами. После декомпозиции сигнала на атомы, каждый из них оценивается с точки зрения перцептуальной важности для реконструкции сигнала.

В работе [4] проводится разреженная аппроксимация сигнала  $s$  следующим образом: строится избыточный словарь на основе атомов Габора. Следующий шаг – декомпозиция сигнала с помощью алгоритма *MP*. Выходом согласованной подгонки является набор параметров. Второй этап алгоритма кодирования – перцептуальная обработка найденных атомов. Для этого, они упорядочиваются по убыванию амплитуды, и к ним применяется модель маскирования. Те из них, которые ниже порога маскирования удаляются из набора.

В отличие от вышеприведённых работ, в [5] представлен аудиокодер на основе разреженной аппроксимации, но выделяющий из сигнала элементы различной природы, и работающий с ними раздельно. В данном случае, под элементами разной природы понимаются гармонические (синусоидальные) составляющие сигнала, транзиенты (кратковременные высокоэнергетические всплески), микротранзиенты (острые всплески низкоэнергетических транзиент), шумовая составляющая. Первые три элемента сигнала параметризуются на основе алгоритма согласованной подгонки с различными словарями для каждой из компонент. Моделирование шумовой составляющей реализовано на основе определения спектральной огибающей сигнала алгоритмом линейного предсказания (*LP*).

### III. СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ

В работе [3] в качестве параметров сравнения авторы приводят коэффициент сжатия (*compression rate* – *CR*) и субъективную оценку качества восстановленного сигнала. Оригинальный сигнал имеет следующие характеристики: стерео, 16 бит *PCM*, 44,1 кГц. В качестве субъективной оценки качества авторами была взята *SDG* (субъективное отличие – *Subjective Difference Grade*) со шкалой от -4 до 0: -4 – очень плохо, -3 – плохо, -2 – немного раздражительно, -1 – хорошо, 0 – отлично. В оценке приняло участие 15 случайных слушателей. Результаты представлены в таблице 1 (*ATFT* – авторский алгоритм).

Таблица 1 – Оценка качества кодеров [3]

Кодер	Значение	Рок	Поп	Классика
<i>MP3</i>	<i>CR</i>	7,5	9	13,6
	<i>SDG</i>	0,067	0	0,067
<i>AAC</i>	<i>CR</i>	9,3	9,6	9,6
	<i>SDG</i>	-0,067	-0,133	-0,2
<i>ATFT</i>	<i>CR</i>	8.4	20.6	40
	<i>SDG</i>	-0.93	-0.8	-0.8

В работе [5] для оценки качества выходного сигнала авторы использовали шкалу *MUSHRA* (*MUltiple Stimuli with Hidden Reference and Anchor*). Шкала имеет следующий вид (*ITU-R BS.1534-2*): 0-20 – очень плохо, 20-40 – плохо, 40-60 – удовлетворительно, 60-80 – хорошо, 80-100 – отлично. Для сравнения были взяты два кодера: *AAC* и *PPC* (*Philips Parametric Coder*). Исходный сигнал – одноканальный 16 бит *PCM* сигнал, 44,1 кГц. Для *AAC* и *TWN* (авторский алгоритм – *Tones Wavelets Noise*) сигнал сжимался с битрейтом 16 кбит/с; для *PPC* образцы были сжаты со скоростью 24 кбит/с. Средние значения результатов тестов следующие: *AAC* – 66, *PPC* – 77, *TWN* – 70.

### ЗАКЛЮЧЕНИЕ

Проведенный анализ показывает, что аудиокодеры построенные на основе алгоритма разреженной аппроксимации успешно справляются с задачей компрессии звуковых сигналов. В работе [2] представлен вариант алгоритма *MP* отличный от рассмотренных выше. Главная особенность подхода заключается в том, что для каждого фрейма входного сигнала синтезируется оптимальный словарь атомов на основе ПДВП и применяется психоакустическая модель для перцептуальной оценки важности каждого полученного атома. Благодаря этим особенностям можно получить более высокое качество реконструированного сигнала и уменьшить вычислительные затраты алгоритма. При этом на тестовых сигналах были получены следующие результаты (для рок, поп и классической музыки соответственно): *CR* – 18,3, 20,2, 14,7. При этом, субъективное отличие от оригинала: -0,3, -0,2, -0,4. Как видно по результатам, данная модель дает высокую степень сжатия сигнала при практически максимальном качестве реконструированных образцов.

- Mallat S. G. Matching pursuits with time-frequency dictionaries / S. G. Mallat, Z. Zhang // IEEE Transactions on signal processing – 1993. – Vol. 41, № 12. – P. 3397–3415.
- Petrovsky Al. Matching pursuit algorithm with frame-based auditory optimized WP-dictionary for audio transient modelling / Al. Petrovsky, A Petrovsky // Electronica. Konstrukcje, technologie, zastosowania – 2008. – №4. – P. 74–80.
- Umaphaty K. Audio signal processing using time-frequency approaches: coding, classification, fingerprinting, and watermarking / K. Umaphaty, B. Ghoraani, S. Krishnan // EURASIP Journal on Advances in Signal Processing – 2010. – Vol. 2010 – P. 1–28.
- Chardon G. Perceptual matching pursuit with Gabor dictionaries and time-frequency masking / G. Chardon, T. Necciari, P. Balazs // IEEE International conference on acoustic, speech, and signal processing (ICASSP) – 2014. – P. 3126–3130.
- Ruiz Reyes N. Adaptive signal modeling based on sparse approximations for scalable parametric audio coding / N. Ruiz Reyes, P. Vera Candead // IEEE Transactions on audio, speech, and language processing – 2010. –Vol. 18, № 3. –P. 447–460.