

## КВАНТОВАНИЕ И КОДИРОВАНИЕ ПАРАМЕТРОВ МАСШТАБИРУЕМОГО АУДИОКОДЕРА

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

Герасимович В.Ю.

Петровский А.А. – д-р. техн. наук, профессор

В рамках доклада описывается схема квантования и кодирования параметров аудиокодера [1]. Приводятся методика объективной оценки восстановленных аудио образцов – PЕМО-Q. Показаны результаты оценки предлагаемого алгоритма кодирования и сравнение с популярными широко распространенными аудиокодерами MP3 и AAC.

Разрабатываемый аудиокодер, описанный в [1], построен на основе алгоритма разреженной аппроксимации (согласованная подгонка). Словарь частотно-временных функций строится для каждого фрейма входного сигнала из самого сигнала с помощью пакета дискретного вейвлет преобразования (ПДВП). Дерево ПДВП формируется динамически для каждого фрейма, при отборе атомов словаря используется психоакустическая модель слуха человека. Одной из важнейших составляющих аудиокодера, влияющих на степень сжатия сигнала, является схема квантования и кодирования передаваемых данных, которая описывается в данной работе.

Параметрами передаваемыми от кодера к декодеру являются отобранные алгоритмом согласованной подгонки атомы (вес, и позиция атома в дереве ПДВП). Необходимость передачи декодеру позиции атома объясняется тем, что реконструирование выходного сигнала производится с помощью полного дерева ПДВП. Следовательно, необходимо знать, в каком узле дерева разместить атом. Поскольку позицией атома в структуре дерева ПДВП является уровень и номер узла, то для представления уровня, на котором находится атом, в двоичном виде необходимо 3 бита (максимальный уровень дерева декомпозиции – 8), для представления номера узла атома максимальное количество бит – 4 (пределным деревом является критическое дерево ПДВП [2, 3]). Суммарное количество бит для кодирования максимально возможного дерева декомпозиции – критического дерева ПДВП – 121, что является большим числом для передачи с каждым фреймом. Эффективным вариантом уменьшения этого значения является передача структуры дерева первого фрейма сигнала, а для остальных фреймов – передача определенного кода, характеризующего отличие структуры дерева текущего фрейма от предыдущего. При этом, можно выделить три варианта изменения структуры: «без изменений», «отсечение узлов», «рост дерева». Следовательно, данный код является двухбитным. Набор, содержащий две единицы применяется в том случае, когда возникает повторяющаяся ситуация «без изменений», которая возникает достаточно часто. Максимальное количество бит, необходимое для кодирования изменения дерева, возникает в ситуации, когда текущее дерево на один уровень меньше критического – 22 бита для данного случая.

Для сжатия и передачи весового коэффициента каждого атома используется алгоритм скалярного квантования. Шаг квантования рассчитывается следующим образом:  $\Delta_{l,n} = \sqrt{12T_{l,n} / K_{l,n}}$ , где  $l$  – номер уровня,  $n$  – номер узла,  $T_{l,n}$  – порог маскирования [3],  $K_{l,n}$  – количество коэффициентов в данном узле. Заквантованные весовые коэффициенты кодируются с помощью алгоритма Хаффмана [4]. Код Хаффмана является алгоритмом префиксного кодирования с переменной длиной кодового слова. Формирование кодов Хаффмана основывается на законе распределения вероятности кодируемых данных. Кодовые книги алгоритма Хаффмана формировались для каждого уровня критического дерева ПДВП.

Среднее количество бит, необходимое для кодирования атомов, рассчитывается следующим образом. В ходе формирования и реализации энтропийного кодирования Хаффмана было рассчитано значение количества бит на отсчет для каждого уровня дерева декомпозиции ПДВП. Сумма этих значений, деленная на максимальное количество полос дерева (25 – критическое дерево ПДВП), показывает средний бюджет бит на один отсчет. Добавлением бюджета бит на кодирование структуры дерева и значения позиции атома в узле был получен средний битрейт для предлагаемого аудиокодера – 45 кбит/с для варианта с 250 атомами. Каждые дополнительные 50 атомов добавляют 8,6 кбит/с к скорости битового потока.

Для объективной оценки качества восстановленного сигнала была выбрана модель PЕМО-Q [5]. Основной подход данного метода заключается в том, что к паре оригинального (опорного) и тестируемого сигналов применяется психоакустическая обработка, которая позволяет симулировать восприятие акустического сигнала человеческим ухом. На выходе модели рассчитываются коэффициенты корреляции для определения меры перцептуального подобия сигналов (*Perceptual Similarity Measure – PSM*). Так как оригинальным (опорным) сигналом является сигнал высокого качества, то допустимо интерпретировать любое перцептуальное отклонение тестируемого сигнала от опорного как ухудшение качества. Для того, что бы соотнести данную объективную оценку качества сигнала с субъективной оценкой *SDG* (*Subjective Difference Grade*), *PSM* можно отобразить в оценку объективного различия (*Objective Difference Grade – ODG*, таблица 1), которая коррелирует с оценкой *SDG*.

Входной тестовой последовательностью служили одноканальные образцы звуковых сигналов с частотой дискретизации 44,1 кГц и разрядностью 16 бит (таблица 2). Все тестовые образцы для сравнения были сжаты с помощью алгоритмов *MPEG-1 Layer 3 (MP3)* и *Advanced Audio Coding (AAC)* с битрейтом 96 кбит/с и 100 кбит/с соответственно. Результаты представлены в таблице 3

Описание искажения	ODG
Не воспринимаемое	0.0
Воспринимаемое, но не раздражающее	-1.0
Немного раздражающее	-2.0
Раздражающее	-3.0
Очень раздражающее	-4.0

Таблица 1 – определение шкалы ODG

Тестовый образец	Описание
es01	Вокал ( <i>Suzan Vega</i> )
es02	Речь на немецком языке
es03	Речь на английском языке
sc01	Соло на трубе и оркестр
sc02	Оркестровая композиция
sc03	Современная поп-музыка
s01	Клавесин
s02	Кастаньеты
s03	Pitch pipe
sm01	Волынка
sm02	Металлофон
sm03	Plucked strings

Таблица 2 – тестовые образцы

	Предлагаемый аудиокодер					MP3	AAC
	250 атомов	300 атомов	350 атомов	400 атомов	450 атомов		
es01	-2.1271	-1.8089	-1.5285	-1.0791	<b>-0.8167</b>	-3.740	-0.218
es02	<b>-0.7938</b>	<b>-0.4067</b>	<b>-0.3510</b>	<b>-0.3280</b>	<b>-0.2968</b>	-3.731	-0.100
es03	-1.6974	-1.5252	-1.3579	<b>-0.7872</b>	<b>-0.6134</b>	-3.757	-0.132
sc01	<b>-0.1706</b>	<b>-0.1460</b>	<b>-0.1443</b>	<b>-0.1443</b>	<b>-0.1460</b>	-3.793	-0.085
sc02	<b>-0.8266</b>	<b>-0.5314</b>	<b>-0.3100</b>	<b>-0.1870</b>	<b>-0.1345</b>	-3.747	-0.154
sc03	-2.5288	-1.5465	<b>-0.7856</b>	<b>-0.3624</b>	<b>-0.2394</b>	-3.742	-0.236
s01	-2.4377	-1.8335	-1.1201	<b>-0.9266</b>	<b>-0.8594</b>	-3.781	-0.483
s02	-3.2103	-2.6996	-1.8548	-1.7548	-1.2776	-3.758	-0.918
s03	-1.1250	<b>-0.9430</b>	<b>-0.6822</b>	<b>-0.5264</b>	<b>-0.4986</b>	-3.558	-0.542
sm01	-3.3915	-2.9678	-1.0726	<b>-0.7921</b>	<b>-0.6330</b>	-3.605	-0.485
sm02	-3.2306	-2.8685	-2.3355	-1.9155	-1.8975	-3.755	-0.269
sm03	-1.4366	<b>-0.8823</b>	<b>-0.4248</b>	<b>-0.2460</b>	<b>-0.1722</b>	-3.713	-0.151

Таблица 3 – Результаты тестирования

Как видно из таблицы 3, разные по своей природе образцы могут быть сжаты с разным битрейтом. Так, например, нет необходимости повышать количество передаваемых атомов для образцов es02, sc01 и sc02, так как 250 атомов достаточно для того, что бы уровень искажений был не воспринимаем на слух. В сравнении с MP3 видно, что разрабатываемый аудиокодер показывает лучшие результаты уже при использовании 250 атомов, при этом, степень сжатия больше, чем у MP3 примерно в два раза. Сравнивая с AAC, видно, что результаты оценки предлагаемого кодера идентичны почти для всех последовательностей при использовании 450 атомов для кодирования, при этом, масштабируемость аудиокодера на основе разреженной аппроксимации позволяет уменьшать битрейт в зависимости от типа входного сигнала, что дает большую степень сжатия. Например, образец s03 можно закодировать используя 350 атомов при этом, объективная оценка качества восстановленного сигнала будет сходна с AAC, но степень сжатия будет выше (7 для AAC и 11 для разрабатываемого аудиокодера). Результат оценки хуже, чем у AAC получился на образцах s02, sm02. Это объясняется тем, что алгоритм кодирования AAC может динамически изменять длину фрейма, что способствует лучшему выявлению и обработке транзиентных составляющих в сигнале. Тем не менее, оценки для данных образцов у описываемого аудиокодера находятся в диапазоне до -2,0 (таблица 1, «немного раздражающее» искажение), что является удовлетворительной оценкой.

Анализ приведенных результатов показывает перспективность дальнейших исследований и применимость представленного подхода к задачам аудиокодирования.

Список использованных источников:

- Петровский Ал. А. Параметрический аудиокодер на основе разреженной аппроксимации с перцептуально-оптимизированным словарем частотно-временных функций / Ал. А. Петровский, В. Ю. Герасимович // Доклады БГУИР, №1(87). – Минск, 2015. – с. 5-11.
- Petrovsky Al. Auditory modeling via frequency warped transforms / Al. Petrovsky, M. Parfeniuk, A. Borowicz, A. Petrovsky // AES 124th Convention – Amsterdam, the Netherlands, 2008, May 17-20.
- Петровский Ал. А. Построение психоакустической модели в области вейвлет коэффициентов для перцептуальной обработки звуковых и речевых сигналов / Ал. А. Петровский // Научно-практический журнал «Речевые технологии», №4. – Москва, 2008. – с. 61-71.
- Huffman D. A Method for the Construction of Minimum-Redundancy Codes / D. Huffman // Proceedings of the IRE, vol. 40 – September, 1952. – pp. 1098-1101.
- Huber R. PEMO-Q – A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception