

# ИССЛЕДОВАНИЕ ПСИХОАКУСТИЧЕСКИХ МОДЕЛЕЙ КОДЕКОВ С КОМПРЕССИЕЙ ЦИФРОВЫХ АУДИОДАНЫХ

## A STUDY OF THE PSYCHOACOUSTIC MODELS OF CODECS WITH COMPRESSION OF THE DIGITAL AUDIO DATA

*Yu. Kovalgin  
D. Fadeev*

### Annotation

The results of experimental studies devoted to the assessment of the quality and effectiveness of the psychoacoustic models of the encoders compressed digital audio data. Shows the advantages of the psychoacoustic model of AAC encoder MPEG-2 ISO/IEC 13818-7 AAC.

**Keywords:** Compression of digital audio, a psychoacoustic model of the AAC codec.

*Ковалгин Юрий Алексеевич*

*Д.т.н., профессор,*

*Санкт-Петербургский государственный  
университет телекоммуникаций  
им. проф. М.А. Бонч-Бруевича*

*Фадеев Даниил Романович*

*Аспирант,*

*Санкт-Петербургский государственный  
университет телекоммуникаций  
им. проф. М.А. Бонч-Бруевича*

### Аннотация

Изложены результаты экспериментальных исследований, посвященных оценке качества и эффективности работы психоакустических моделей кодеков с компрессией цифровых аудиоданных. Показаны преимущества психоакустической модели кодера AAC стандарта MPEG-2 ISO/IEC 13818-7 AAC.

### Ключевые слова:

Компрессия цифровых аудиоданных, оценка качества, психоакустическая модель кодера AAC.

## Введение

Несмотря на довольно значительный период существования стандартов MPEG и очевидный прогресс в снижении скорости цифрового потока при кодировании высококачественных звуковых сигналов (ЗС), полноценная оценка эффективности работы психоакустических моделей до сих пор не проводилась. Существует значительное количество работ [1,2,11,12], посвященных исследованию качества компрессии того или иного кодера и возникающим при этом артефактам. В то же время особенности работы непосредственно самих психоакустических моделей (ПАМ) и получаемые при этом результаты требуют более детального изучения, особенно в части выделения и распределения бит по субполосам кодирования.

*Для исследования ПАМ кодеков с компрессией цифровых аудиоданных с потерями был реализован аппаратно-программный комплекс, позволяющий:*

- ◆ подключать исследуемые ПАМ;
- ◆ подавать на вход каждой ПАМ тональные сигналы разных частот (для проверки корректности работы) и отрывки реальных звуковых сигналов с возможностью выбора длительности анализируемого фрагмента звукового отрывка;
- ◆ вычислять отношение сигнал-маска, в дБ, а так-

же среднее, максимальное и минимальное число бит, требуемое исследуемой ПАМ для прозрачного кодирования каждого элемента выборки звукового сигнала;

- ◆ статистически обрабатывать полученные массивы числовых данных;
- ◆ представлять результаты исследований в виде графических зависимостей.

*В соответствии с этим экспериментальный программный комплекс содержал следующие основные части:*

- ◆ библиотеку звукового материала. В настоящей версии программный комплекс поддерживает формат CD Audio с частотой дискретизации 44100 Гц и разрешением 16 бит/отсчет, режимы работы моно или стерео;
- ◆ библиотеку исследуемых психоакустических моделей (ПАМ); она содержит психоакустические модели кодеков стандартов MPEG-1 [3,11], MPEG-2 [4,11] и MPEG-4 [6,12], а именно ПАМ\_1, ПАМ\_2 и ПАМ кодера AAC (ПАМ\_AAC) [5, 12], а также и ПАМ\_ВП, предложенную Г.Г.Рогозинским для кодера, работающего в вейвлетном пространстве [7, 11];
- ◆ блок обработки звукового материала, позволяющий выбрать тот или иной звуковой файл из базы (библиотеки) данных, выделить фрагмент для исследования, задавая начальную и конечную точки исследуемого фрагмента отрывка;

◆ блок статистической обработки и визуализации результатов вычислений, выполняемых в каждой исследуемой ПАМ.

Результаты вычислений, выполняемых в ПАМ, выводятся в форме массива данных, элементами которого являются значения глобального порога маскировки, в дБ, отношения сигнал–маска SNR, дБ, а также и число бит  $m$ , требуемых для прозрачного кодирования каждого элемента выборки (отсчета звукового сигнала или коэффициента МДКП) для каждой из полос психоакустического анализа.

Напомним, что различные психоакустические модели используют разное разделение на полосы психоакустического анализа. В случае с ПАМ\_1 и ПАМ\_2, полосы психоакустического анализа соответствуют полосам кодирования (табл. 1). Их ширина составляет 698 Гц. В отличие от этого в ПАМ\_ВП и ПАМ\_ААС полосы психоакустического анализа имеют различную ширину (табл. 2, 3).

После вычислений, выполняемых в ПАМ, вся информация выводится в отдельный файл электронных таблиц Microsoft Excel, что позволяет открыть его в большинстве программ, используемых для статистического анализа результатов эксперимента, а также для визуализации полученных данных.

Выбор в пользу системы компьютерного моделирования был сделан по причине наличия уже частично готовых ПАМ, реализованных в среде Matlab. В качестве программной модели ПАМ\_1 использовалась модель, созданная Fabien Petitcolas [8]. Программные модели ПАМ\_2 и ПАМ\_ААС были созданы аспирантами СПбГУТ им. проф. М.А. Бонч–Бруевича, среди которых следует отметить, прежде всего, М.В. Зырянова, [9], и Ятагаму Гамаге Д.П., [10]. Код экспериментальной вейвлетной модели (ПАМ\_ВП) заимствован из приложения к диссертации Г.Г.Рогозинского, [7].

Для проверки правильности функционирования каждой из отобранных ПАМ было проведено их предварительное тестирование на тональных сигналах с частотами 100, 500, 2500, 5000, 10000, 12000, 15000 и 18000 Гц и выполнен ручной расчет по формулам на этих же частотах. Результаты работы ПАМ и вычислений, проделанных в пакете Mathcad, подтвердили правильность выполняемых вычислений в каждой отобранной для исследования ПАМ.

Для оценки эффективности работы моделей с широкополосными звуковыми сигналами были отобраны отрывки реальных звучаний следующих жанров: органно–хоровая музыка, оркестровая музыка, рок, поп, джаз, электронная танцевальная музыка, камерная музыка,

Таблица 1.

Граничные частоты и ширина полос психоакустического анализа ПАМ\_1 и ПАМ\_2, частота дискретизации 44,1 кГц

№	$F_n$ , Гц	$F_b$ , Гц	$\Delta F$ , Гц	№	$F_n$ , Гц	$F_b$ , Гц	$\Delta F$ , Гц
1	0	689	698	17	11025	11714	698
2	689	1378	698	18	11714	12403	698
3	1378	2067	698	19	12403	13092	698
4	2067	2756	698	20	13092	13781	698
5	2756	3445	698	21	13781	14470	698
6	3445	4134	698	22	14470	15159	698
7	4134	4823	698	23	15159	15848	698
8	4823	5513	698	24	15848	16538	698
9	5513	6202	698	25	16538	17227	698
10	6202	6891	698	26	17227	17916	698
11	6891	7580	698	27	17916	18605	698
12	7580	8269	698	28	18605	19294	698
13	8269	8958	698	29	19294	19983	698
14	8958	9647	698	30	19983	20672	698
15	9647	10336	698	31	20672	21361	698
16	10336	11025	698	32	21361	22050	698

Таблица 2.

Граничные частоты и ширина полос психоакустического анализа ПАМ\_ВП,  
частота дискретизации 44,1 кГц

№	Fн, Гц	Fв, Гц	Δ F, Гц	№	Fн, Гц	Fв, Гц	Δ F, Гц
1	0	86	86	15	2067	2412	345
2	86	172	86	16	2412	2756	345
3	172	258	86	17	2756	3445	689
4	258	345	86	18	3445	4134	689
5	345	431	86	19	4134	4823	689
6	431	517	86	20	4823	5513	689
7	517	603	86	21	5513	6891	1378
8	603	689	86	22	6891	8269	1378
9	689	861	172	23	8269	9647	1378
10	861	1034	172	24	9647	11025	1378
11	1034	1206	172	25	11025	13781	2756
12	1206	1378	172	26	13781	16538	2756
13	1378	1723	345	27	16538	19294	2756
14	1723	2067	345	28	19294	22050	2756

дикторская речь; всего по пять произведений каждого жанра. Кроме того в каждом из отрывков определенного жанра при тестировании выделялось два–три разных фрагмента. Длительность каждого фрагмента отрывка при тестировании ПАМ составляла около 15 с.

Напомним, что в ПАМ для каждой полосы психоакустического анализа  $s_b$  рассчитывается отношение сигнал–маска, представляющее собой выраженное в дБ отношение энергии полезного сигнала  $E_c$  к максимально допустимому значению энергии шумов квантования  $E_{ш.кв.}$ , которое еще маскируется полезным сигналам данной полосы (т.е. лежит на пороге слышимости). При этом требуемое для прозрачного кодирования каждого элемента число бит в данной полосе определялось из условия:  $10 \lg(E_c/E_{ш.кв.})_{s_b} = 6 \cdot m$ , где  $m$  – число разрядов в кодовом слове кодируемого элемента. В качестве примера на рис. 1 представлены зависимости отношения сигнал–маска и требуемое моделью ПАМ\_ААС число бит для прозрачного кодирования.

Слишком большой объем экспериментальных исследований не позволяет в рамках одной статьи привести в полном объеме все полученные результаты. Поэтому здесь приведены лишь наиболее типичные зависимости (рис.2–5) и обобщенные результаты анализа всех экспериментальных данных. По оси ординат на рис. 2–5 отложено среднее, максимальное и минимальное число бит (без учета знакового бита), которое необходимо выделить на кодирование каждого элемента выборки для обеспе-

чения прозрачного кодирования, при котором искажения компрессии остаются еще незаметными на слух.

По оси абсцисс – номер полосы психоакустического анализа ПАМ (табл.1–3). Горизонтальные линии в центре прямоугольников соответствует медианным значениям. Верхняя и нижняя границы прямоугольников соответственно определяют 75–ый и 25–ый процентиля длин кодовых слов. Напомним, что процентиль – это процедура ранжирования полученных значений, применяемая в статистике. Например, если нижней границе прямоугольника соответствует длина кодового слова 2 разряда, то это значит, что 25% длин кодовых слов каждого кодируемого элемента не превышают это значение. В свою очередь, если верхней границе прямоугольника соответствует значение 4 разряда, то это значит, что 75 % значений длин кодовых слов также не превышают эту величину. Медиана (черта в средней части прямоугольника) – это 50–ый процентиль. При выполнении данных вычислений длина кодовых слов не округлялась до ближайшего целого большего значения, как это должно было бы иметь место в случае реального кодирования.

Заметим, что ПАМ\_1 и ПАМ\_2 имеют 32 полосы психоакустического анализа одинаковой ширины (табл.1), ПАМ\_ВП – 28 полос психоакустического анализа, приближающихся по ширине к критическим полосам слуха (табл.2), ПАМ\_ААС – 49 полос психоакустического анализа постепенно расширяющихся с повышением частоты (табл.3).

Таблица 3.

Граничные частоты и ширина полос психоакустического анализа ПАМ\_ААС,  
частота дискретизации 44,1 кГц

№	F <sub>н</sub> , Гц	F <sub>в</sub> , Гц	Δ F, Гц	№	F <sub>н</sub> , Гц	F <sub>в</sub> , Гц	Δ F, Гц
1	0	86	86	26	4651	5168	516
2	86	172	86	27	5168	5685	516
3	172	258	86	28	5685	6288	602
4	258	345	86	29	6288	6891	602
5	345	431	86	30	6891	7580	688
6	431	517	86	31	7580	8269	688
7	517	603	86	32	8269	8958	688
8	603	689	86	33	8958	9647	688
9	689	775	86	34	9647	10336	688
10	775	861	86	35	10336	11025	688
11	861	1034	172	36	11025	11714	688
12	1034	1206	172	37	11714	12403	688
13	1206	1378	172	38	12403	13092	688
14	1378	1550	172	39	13092	13781	688
15	1550	1723	172	40	13781	14470	688
16	1723	1895	172	41	14470	15159	688
17	1895	2067	172	42	15159	15848	688
18	2067	2326	258	43	15848	16538	688
19	2326	2584	258	44	16538	17227	688
20	2584	2842	258	45	17227	17916	688
21	2842	3101	258	46	17916	18605	688
22	3101	3445	344	47	18605	19294	688
23	3445	3790	344	48	19294	19983	688
24	3790	4221	430	49	19983	22050	2064
25	4221	4651	430				

По результатам всего комплекса исследований ПАМ можно сделать следующие основные выводы:

1. Все исследуемые ПАМ наибольшее число бит выделяют на кодирование информации, содержащейся на низких и частично средних частотах. Это и понятно, ибо спектры звучаний музыкальных инструментов содержат основной тон, небольшое число гармонических спектральных компонент кратных частоте основного тона (обычно не более 10...12) и несколько спектральных компонент не кратных частоте основного тона (тональные компоненты, их число не превышает обычно 5...7). Все они имеют, как правило, высокий уровень энергии и расположены на низких и средних частотах, не превышающих обычно 4000–6000 Гц. Именно эти спектральные

компоненты в значительной степени определяют индивидуальную окраску звучаний, свойственных каждому из музыкальных инструментов. А значит, и кодировать звуковые сигналы в этой области частот следует наиболее точно, но конечно с учетом одновременной маскировки, которая в этой области частот из-за существенной неравномерности спектра проявляется наиболее резко.

На высоких частотах спектры реальных звучаний имеют обычно весьма однородный вид без резких изменений амплитуды.

В самые верхние по частоте полосы анализа биты практически не выделяются всеми ПАМ;

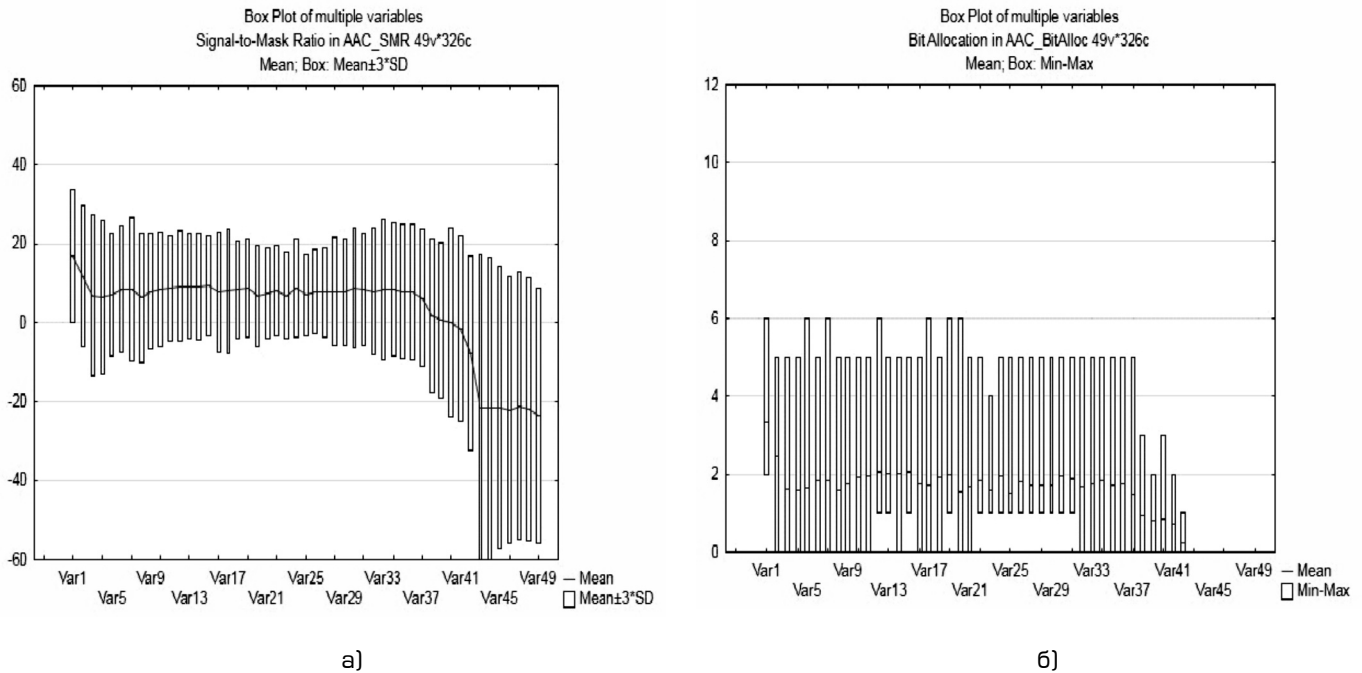


Рисунок 1. Отношение сигнал-маска (а) и требуемое ПАМ\_ААС для прозрачного кодирования число бит (б); поп-музыка (усредненные данные для четырех отрывков, из которых для исследования взяты по три фрагмента, каждый длиной 15 с). Сплошная линия – среднее значение, верхняя и нижняя границы прямоугольников соответствуют среднему квадратическому отклонению  $\pm 3\sigma$

2. Для всех звуковых отрывков ПАМ\_1 выделяет биты практически только в первые 15...17 субполос психоакустического анализа (до частоты 9000...11000 Гц), что при декодировании приводит к сокращению полосы частот звукового сигнала, а значит и к появлению артефактов компрессии, возможно, к появлению заметных на слух тембральных искажений. При этом не вызывает сомнений тот факт, что в электронной танцевальной музыке за счет использования электронных музыкальных инструментов и различных психоакустических процессоров высокочастотная область спектра оказывается, как правило, энергетически весьма насыщенной;

3. Меньшее число бит по сравнению с ПАМ\_1 требует ПАМ\_2. Однако и она выделяет малое количество битов для качественного кодирования высокочастотных компонент звуковых сигналов, т.е. недостаточно полно учитывает распределение энергии звукового сигнала по частоте, особенно в его высокочастотной части, где энергия сигнала обычно все же существенно меньше, чем на низких и средних частотах. Эта тенденция прослеживается и при кодировании звуковых сигналов других жанров. Но в отличие от ПАМ\_1 она более экономно расходует биты.

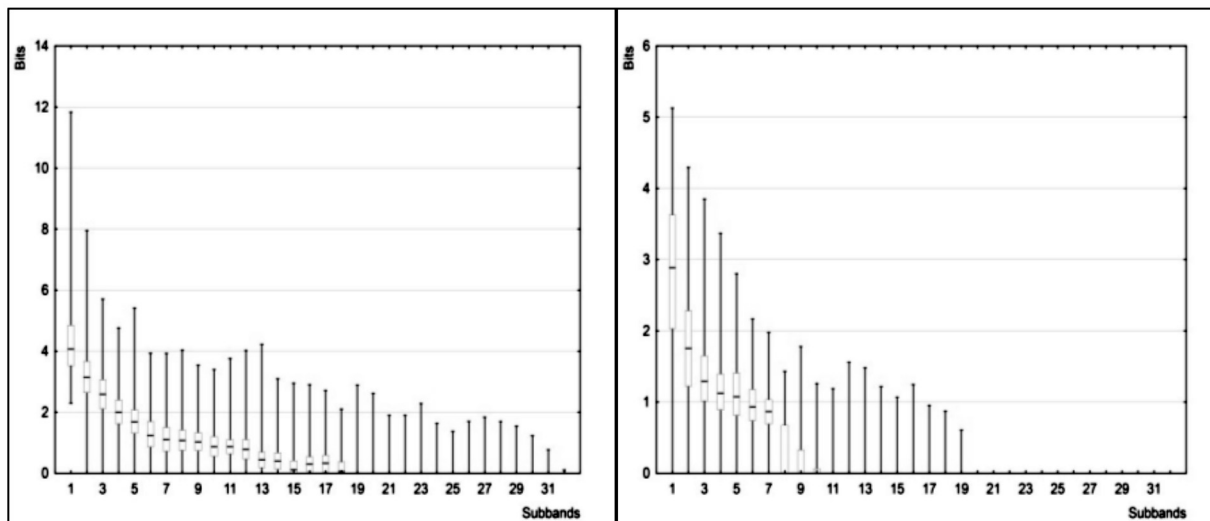
4. ПАМ\_ААС обеспечивает более равномерное распределение битов по полосам психоакустического ана-

лиза (кодирования), в том числе и в области высоких частот. Она, к тому же, имеет и более высокий разброс числа выделяемых бит при кодировании разных выборок, что говорит о более точном учете особенностей временной и спектральной структуры звукового сигнала.

5. ПАМ\_ВП выделяет биты в основном в первые 15 полос психоакустического анализа (до частоты 2500...3000 Гц). На более высоких частотах требуемое ею для прозрачного кодирования число бит крайне незначительно, что вряд ли оправданно. Эта тенденция прослеживается на звуковых сигналах разных жанров.

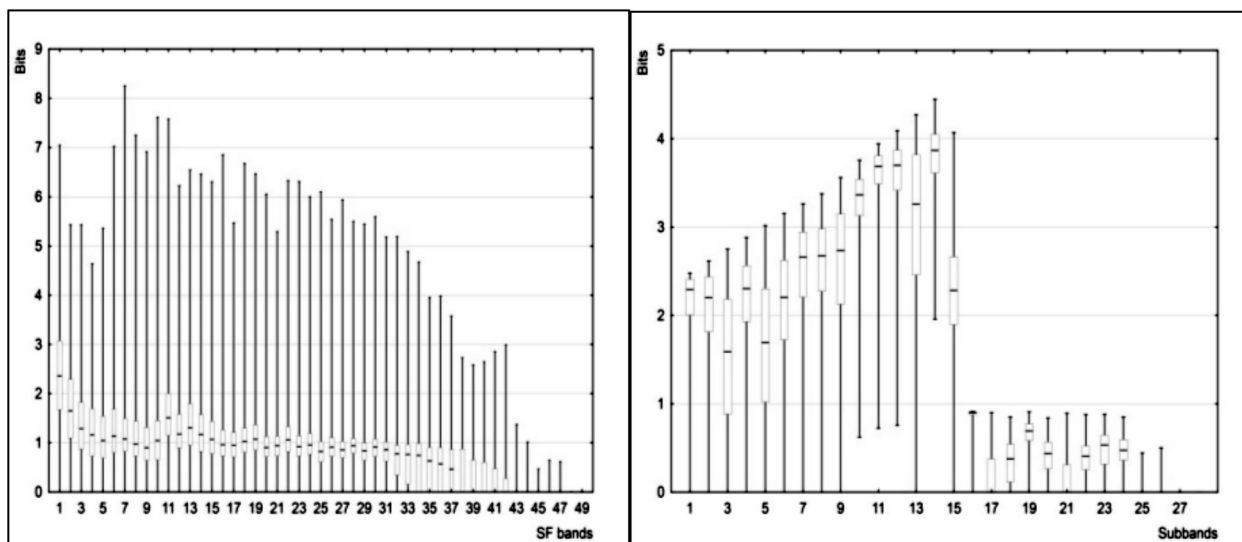
При оценке эффективности работы ПАМ важно знать и распределение бит, требуемое каждой ПАМ для прозрачного кодирования сигнала выборки, во времени. Имея данные о числе бит (длине кодового слова), требуемых для прозрачного кодирования элементов каждой выборки звукового сигнала для выбранных отрывков и зная число элементов (например, отсчетов) звукового сигнала в каждой полосе психоакустического анализа (табл.4), можно массивы этих данных представить в форме гистограмм, например, так, как это сделано на рис.6.

Здесь по оси ординат отложено число выборок звукового сигнала, а по оси абсцисс – суммарное количество бит, требуемое для прозрачного кодирования отсчетов



а)

б)



в)

г)

Рисунок 2. Среднее число бит, требуемых для прозрачного кодирования одного отсчета или коэффициента МДКП в каждой из полос психоакустического анализа (электронная танцевальная музыка): а - ПАМ\_1; б - ПАМ\_2; в - ПАМ\_AAS; г - ПАМ\_ВП (Добеши 4)

одной выборки. Заметим, что для ПАМ\_1 длина выборки составляла 1024 отсчета, в то время, как для остальных ПАМ она составляла 2048 отсчетов. Заметим, что при выполнении данных вычислений требуемое ПАМ для прозрачного кодирования число бит в каждой из полос округлялось до ближайшего большего целого значения. Кроме того к данному числу добавлялся еще один бит, необходимый для учета полярности звукового сигнала. При расчете гистограмм ширина интервалов была взята одинаковой. Их число равнялось 10.

В качестве звукового сигнала при построении этих

гистограмм для каждой ПАМ были выбраны 12 фрагментов звучаний одного жанра, каждый длительностью 15 с. На рис.б жанр фрагментов – рок музыка. Общая длительность звучания отобранных фрагментов определенного жанра составила 180 с.

Анализ полученных экспериментальных данных для звуковых сигналов разных жанров и в этом случае показал, что ПАМ\_AAS наиболее полно учитывает динамику звукового сигнала и распределение его энергии по частоте.

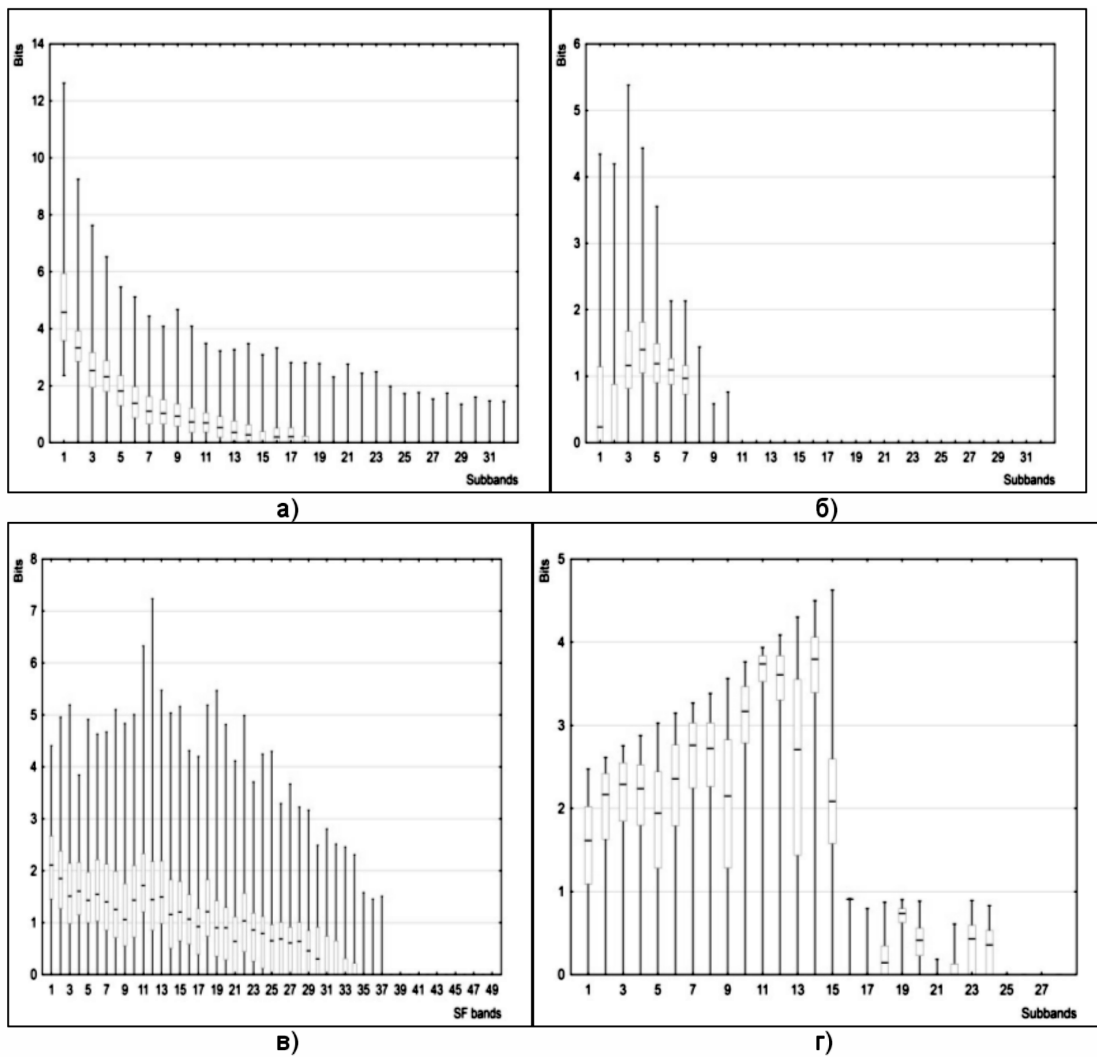


Рисунок 3. Среднее число бит, требуемых для прозрачного кодирования одного отсчета или коэффициента МДКП в каждой из полос психоакустического анализа (оркестровая музыка): а - ПАМ\_1; б - ПАМ\_2; в - ПАМ\_AAS; г - ПАМ\_ВП (Добеши 4)

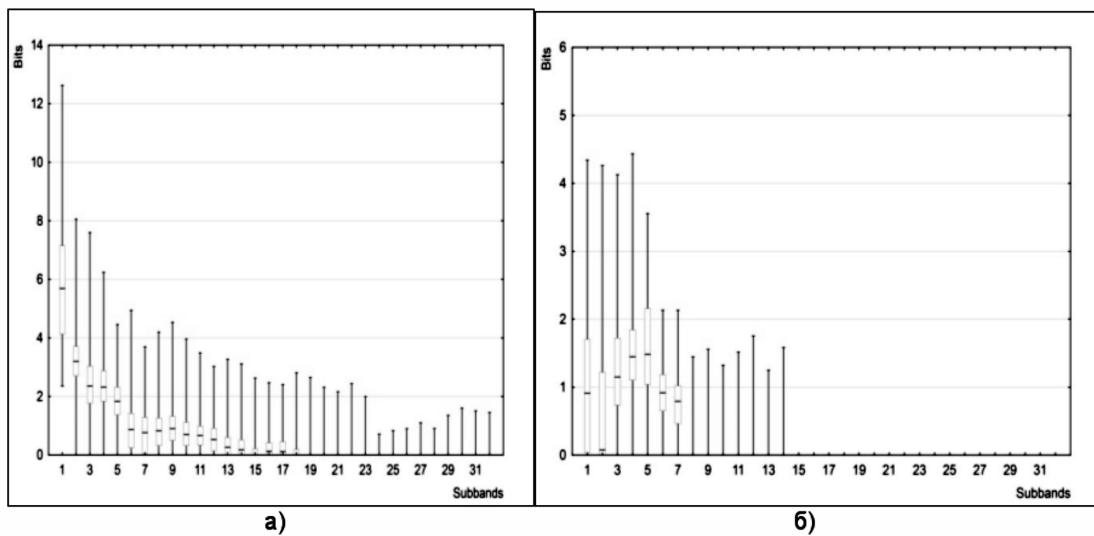


Рисунок 4/1. Среднее число бит, требуемых для прозрачного кодирования одного отсчета или коэффициента МДКП в каждой из полос психоакустического анализа (камерная классическая музыка): а - ПАМ\_1; б - ПАМ\_2; в - ПАМ\_AAS; г - ПАМ\_ВП (Добеши 4)

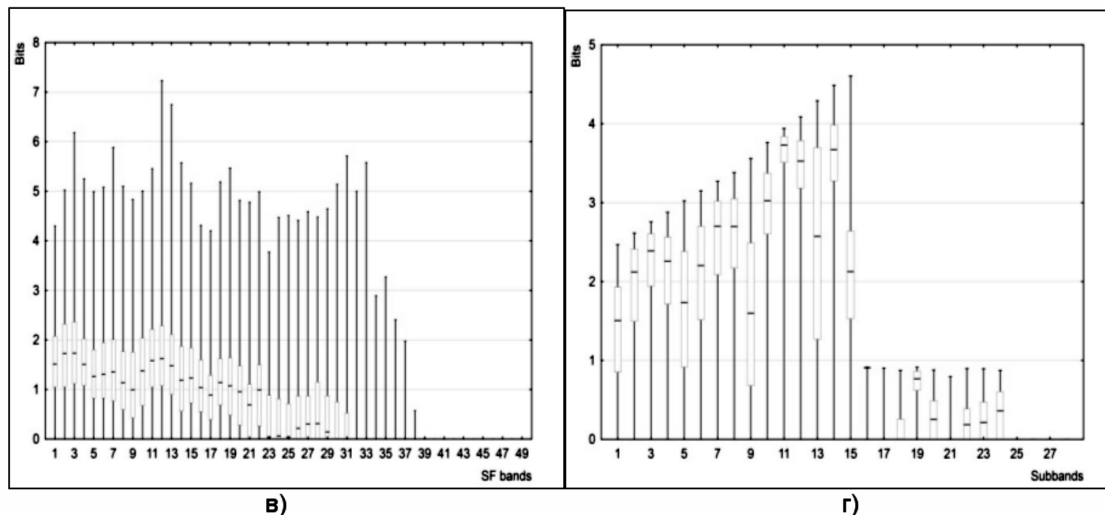


Рисунок 4/2. Среднее число бит, требуемых для прозрачного кодирования одного отсчета или коэффициента МДКП в каждой из полос психоакустического анализа (камерная классическая музыка): а - ПАМ\_1; б - ПАМ\_2; в - ПАМ\_ААС; г - ПАМ\_ВП (Добеши 4)

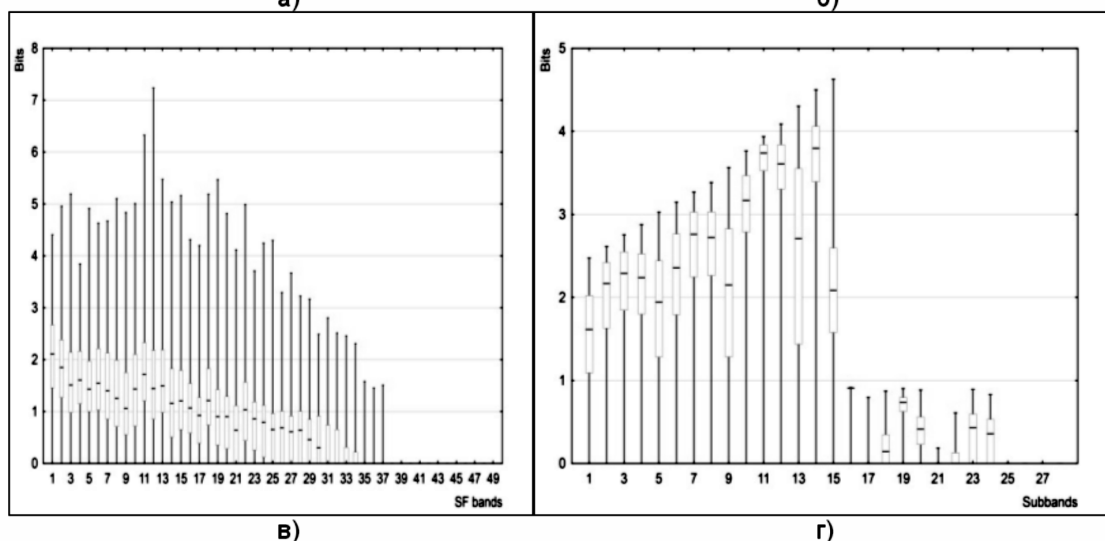
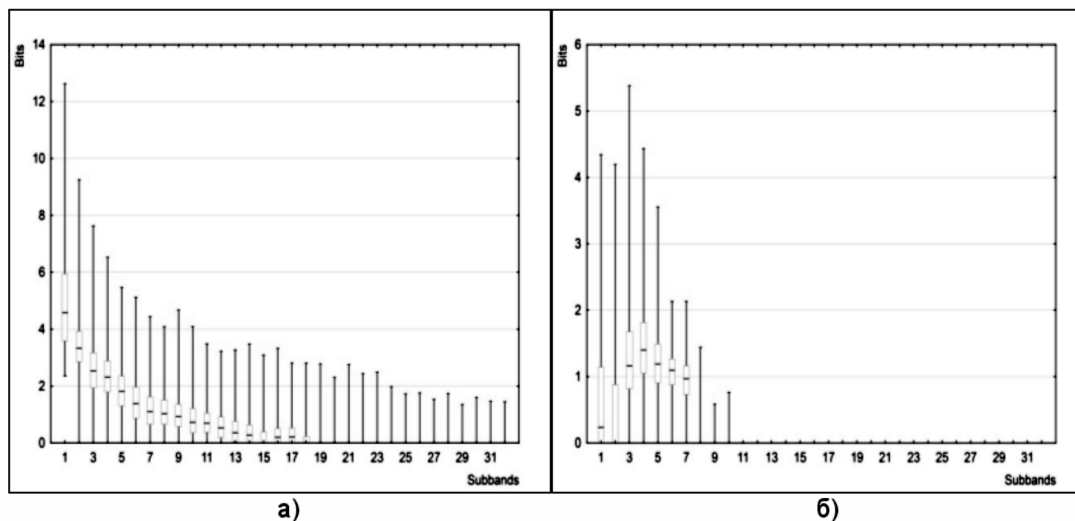


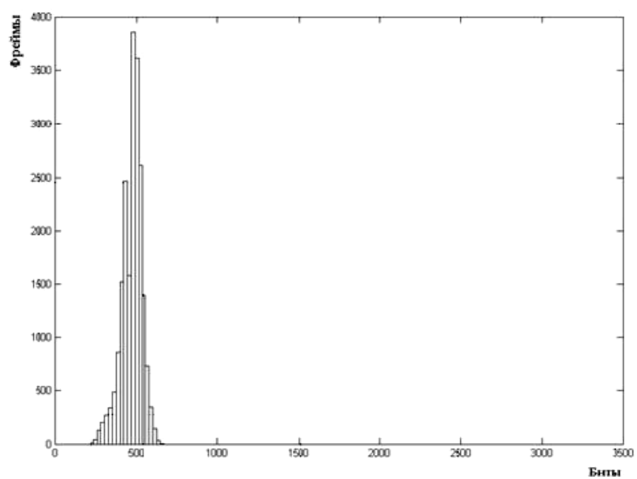
Рисунок 5. Среднее число бит, требуемых для прозрачного кодирования одного отсчета или коэффициента МДКП в каждой из полос психоакустического анализа (рок- музыка): а - ПАМ\_1; б - ПАМ\_2; в - ПАМ\_ААС; г - ПАМ\_ВП (вейвлет Добеши 4)



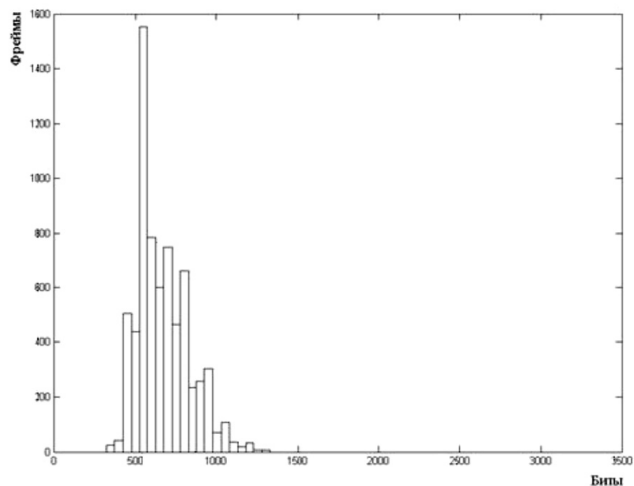
Таблица 4.

Ширина полосы частот и число отсчетов звукового сигнала для ПАМ\_ААС при длине выборки 2048 отсчетов и частоте дискретизации 44,1 кГц.

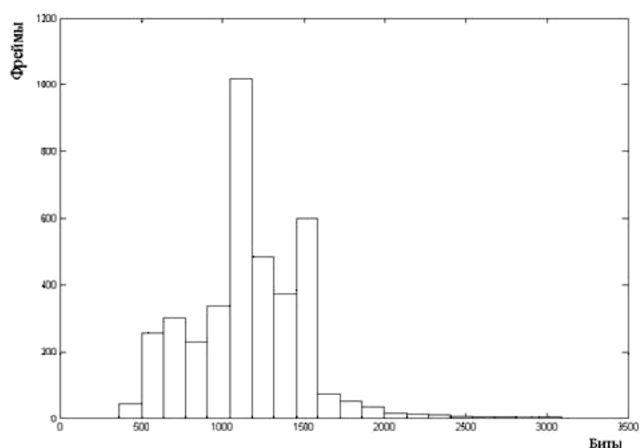
Номер полосы психоакустического анализа	Ширина полосы частот, Гц	Число отсчетов звукового сигнала в каждой полосе психоакустического анализа	Номер полосы психоакустического анализа	Ширина полосы частот, Гц	Число отсчетов звукового сигнала в каждой полосе психоакустического анализа
1...10	86	8	24...25	430	40
11...17	172	16	26...27	516	48
18...21	256	24	28...29	602	56
22...23	344	32	30...48	688	64
			49	2064	192



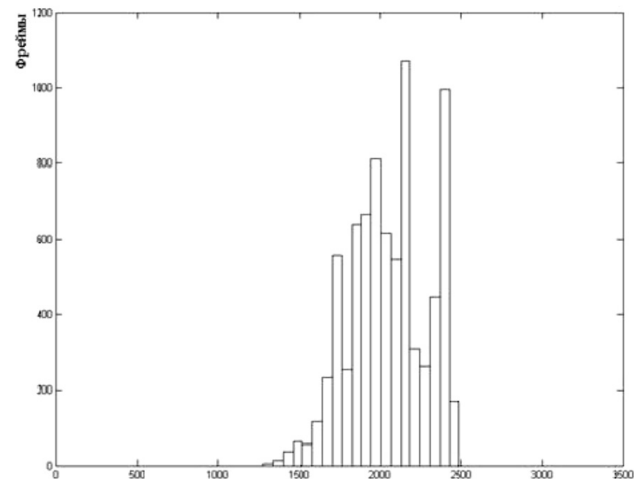
а)



б)



в)



г)

Рисунок 6. Гистограммы распределения бит, требуемых ПАМ для прозрачного кодирования выборки звукового сигнала (рок- музыка): а - ПАМ\_1; б - ПАМ\_2; в - ПАМ\_ААС; г - ПАМ\_ВП (вейвлет Добеши 4)

Выводы по результатам данного исследования:

1. Анализ публикаций, посвященных расчету и моделированию порога одновременной маскировки, выполненные в данной работе исследования, особенности распределения энергии звуковых сигналов по частоте, все это показывает, что наилучшие результаты распределения бит по субполосам анализа/кодирования обеспечивает психоакустическая модель кодера AAC (ПАМ\_ААС).

2. В модели ПАМ\_ААС спектральные компоненты выборки группируются в 49 достаточно узких по ширине полос психоакустического анализа. Ширина полос психоакустического анализа в ней постепенно возрастает с повышением частоты, оставаясь все же существенно уже критических полос слуха, что позволяет более точно учесть особенности распределения энергии звуковых

сигналов по частоте при кодировании.

3. ПАМ\_ААС обладает наилучшим частотным и временным разрешением, наиболее экономно расходует биты при кодировании звуковых сигналов разных жанров, может быть при соответствующей адаптации рекомендована для применения в кодере, работающем в вейвлетном пространстве.

4. Вследствие существенно худшей избирательности дерева анализирующих QMF-фильтров по сравнению с банками PQMF-фильтров, широко применяемых в алгоритмах компрессии группы MPEG, при разработке кодеров работающих в вейвлетном пространстве необходимо дополнительно изучить особенности распределения энергии шумов квантования по частоте и разработать дополнительный механизм для их учета при кодировании коэффициентов вейвлетного дискретного преобразования.

#### ЛИТЕРАТУРА

1. Bosi M., Goldberg R.E. Introduction to Digital Audio Coding and Standards, Boston: Kluwer Academic Publishers, 2003.
2. Pan D., A tutorial on MPEG/Audio compression, IEEE Multimedia, 1995, vol.2, no.2, pp.60–74.
3. International Standard ISO/IEC 11172–3. Information technology–Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s– Part 3: Audio, 1993–08–01.
4. International Standard ISO/IEC 13818–3. Information technology–Generic coding of moving pictures and associated audio information. Part 3:Audio, 1995–05–15.
5. International Standard ISO/IEC 13818–7. Information technology–Generic coding of pictures and associated audio information. Part 7: Advanced Audio Coding (AAC), 1997 (E).
6. ISO/IEC FCD 14496–3 Subpart 1. Information Technology–Very Low Bitrate Audio–Visual Coding. Part 3: Audio, 1998–05–10 (ISO/JTC 1/SC 29, N2203).
7. Рогозинский Г.Г. Перцепционное сжатие звука с использованием вейвлетных пакетов. Автореферат диссертации на соискание ученой степени канд.техн. наук //Санкт–Петербургский государственный университет кино и телевидения, 2010, –19 с. <http://www.petitcolas.net/fabien/software/mpeg/index.html>
8. Зырянов М.В. Повышение эффективности алгоритмов компрессии цифровых аудиоданных на основе учета временной маскировки. Автореферат диссертации на соискание ученой степени кандидата технических наук, СПб.: СПбГУТ, 2007. – 16 с.
9. Ятагама Гамаге Д.П. Повышение эффективности алгоритмов компрессии цифровых данных при кодировании сигналов стереопары. Автореферат диссертации на соискание ученой степени кандидата технических наук, СПб.: СПбГУТ 2005. – 16 с.
10. Аудиотехника: учебник/Ю.А.Ковалгин, Э.И.Вологдин; под ред. проф. Ю.А.Ковалгина. – М.: Горячая линия – Телеком, 2013. – 768 с.
11. Ковалгин Ю.А. Психоакустика и компрессия цифровых аудиоданных. Монография, СПб.: СПбГУТ, 2013.– 300 с.

© Ю.А. Ковалгин, Д.П. Фадеев, ( Daniilfadeev2015@yandex.ru ), Журнал «Современная наука: актуальные проблемы теории и практики»,

