

УДК 621.396.7

ААС – ПРОГРЕССИВНЫЙ МЕТОД ЦИФРОВОГО КОДИРОВАНИЯ ЗВУКА

Иртюга В.А.

Введение

Стандарт звукового сжатия звукового сигнала ААС (Advance Audio Coding – усовершенствованное звуковое кодирование) был разработан группой компаний Fraunhofer IIS, AT&T, Sony и Nokia. По результатам этой работы в апреле 1997 года стандарт кодека был официально опубликован в составе международного стандарта MPEG-2 (ISO/IEC 13818, часть 7 «ААС») [1].

С самого начала ААС позиционировался, как эффективная замена распространенному стандарту звуковой обработки MP3, разработанному в рамках стандарта MPEG-1 (в части «Аудио», уровень 3), что и было с успехом достигнуто благодаря нововведенным алгоритмам и усовершенствованиям, использованным при разработке стандарта кодека. Исходя из анализа результатов сравнительных испытаний, можно утверждать, что для всех скоростей битового потока до 192 кбит/с (на канал) декодер ААС дает более качественный звуковой сигнал, чем декодер MP3. Отметим, что сравнивать качественные характеристики кодеков выше указанной границы скорости становится существенно сложнее, так как при потоке более 192 кбит/с на канал отличия от оригинального сигнала трудно различимы для обоих кодеков и требуют уже тонкого подбора тестового аудиоматериала и наличия большого количества высокопрофессиональных музыкальных экспертов.

Как и во многих стандартах подобного рода, собственно стандартизации подвергнута только декодирующая часть кодека, что оставляет разработчикам и производителям поле для конкуренции и совершенствования выпускаемых ими кодеров. В настоящее время существует несколько реализаций стандарта ААС. В частности, можно упомянуть разработанный компанией Apple ААС-кодер, входящий в состав пакета iTunes, бесплатно распространяемый кодек FAAC/FAAD2, а также ААС-кодек, разработанный в ЗАО «НИИР-КОМ». По результатам испытаний последнего из указанных кодеков можно сказать, что ААС-поток скоростью от 24 кбит/с соответствует качеству FM-стереотрансляции, поток от 32 кбит/с - качеству Audio-CD, а поток 128 кбит/с - сравним по качеству восстановленного сигнала с шестиканальным звуком стандарта "Dolby Digital Surround EX".

В настоящее время ААС получил свое развитие в стандарте MPEG-4 (ISO/IEC 14496, часть 3, подчасть 4 «Общее кодирование звука») [2]. При этом новый стандарт является только дополнением к уже существующим и описанным в MPEG-2 алгоритмам звуковой обработки, а не отменяет его. В частности, профиль ААС-LC (low complexity - упрощенный профиль) описан только в стан-

Рассматривается один из лучших на сегодняшний день метод цифрового сжатия звукового сигнала – ААС (Advance Audio Coding – усовершенствованное звуковое кодирование). Приводятся структурные схемы кодеров и декодеров, соответствующих различным версиям стандарта кодека, при этом особое внимание уделяется кодирующей части кодека. На примере разработанной в ЗАО «НИИР-КОМ» усовершенствованной версии кодека ААС – HE-AAC (High Efficiency) рассматривается вопрос позиционирования данного продукта на рынке мультимедийных услуг.

дарте MPEG-2, являясь при этом базовым для нового высокоэффективного варианта кодека - HE-AAC (High Efficiency).

На настоящий момент существуют две версии расширения кодека ААС: «HE-AAC v.1» и «HE-AAC v.2». Первая версия (v.1) использует технологию спектрального «дублирования» (replication) нижней части спектра звукового сигнала в верхнюю: данный алгоритм носит название SBR – Spectral Band Replication. Вторая версия (v.2) полностью включает в себя всю функциональность первой и расширяет ее с помощью алгоритма параметрического кодирования стереосигнала, этот алгоритм носит название Parametric Stereo (PS).

Структура кодирующего устройства. На рис. 1 приведена структурная схема кодера ААС. Отметим, что здесь и далее под понятиями «звуковой сигнал» и «спектр звукового сигнала» подразумеваются соответственно дискретизированный звуковой сигнал и дискретный спектр дискретизированного звукового сигнала.

Регулировка усиления (Gain control). Данный модуль не является обязательным в структуре кодека и предназначен для регулировки уровней звукового сигнала. На стороне кодера происходит поблочная регулировка уровня звукового сигнала и передача декодеру информации, необходимой для восстановления исходного соотношения. В состав данного модуля входит 4 полифазных квадратурных фильтров (PQF), обеспечивающих четырехполосное разделение спектра сигнала с последующим контролем уровня в каждой из четырех полос во временной области.

Психоакустическая модель вычисляет максимальные значения ошибок, маскируемых (то есть незаметных для человеческого уха) основным сигналом. Это так называемые пороги ошибок, превышение которых будут заметны при воспроизведении декодированного звукового сигнала. В [1] приведены таблицы (табл. С.1-С.24 в Информационном Приложении С), в соответствии с которыми производятся вычисления указанных порогов. Кроме этого в блоке психоакустической модели осуществляется выбор типа и длины взвешивающего окна, используемого в дальнейшем непосредственно перед МДКП-преобразованием (см. ниже).

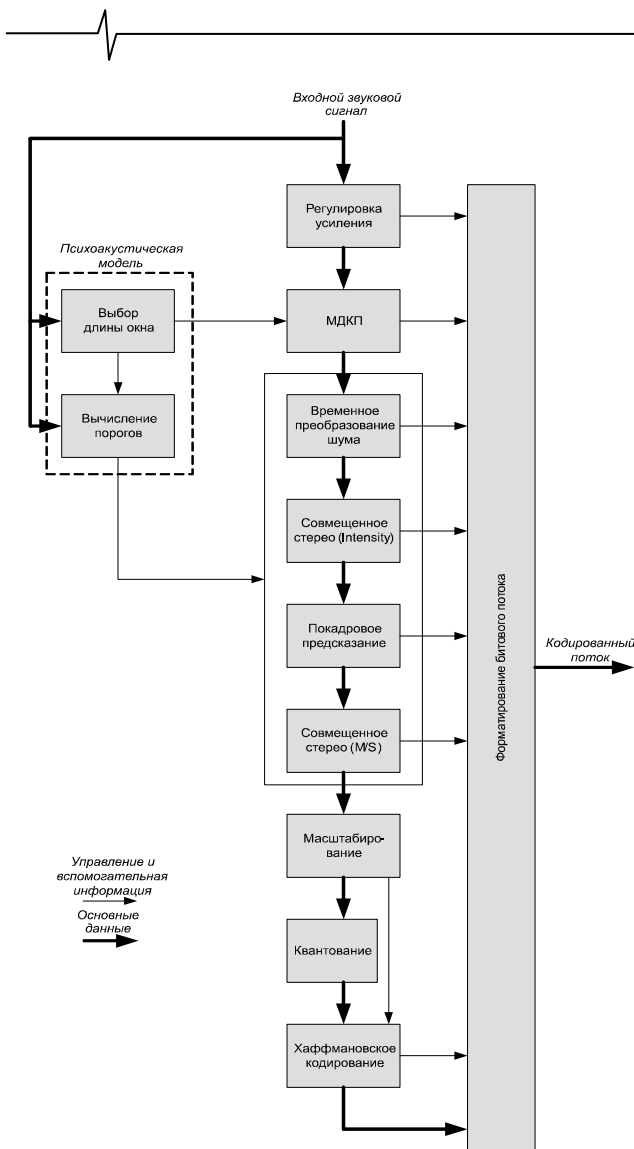


Рис. 1. Схема кодера AAC

Отметим, что выбор осуществляется между двумя типами окон – Кайзера-Бесселя и синусным. Выбор длины взвешивающего окна и, соответственно, длины МДКП для одного кадра осуществляется исходя из следующих альтернатив: 2048 либо 256 при длине кадра 1024 отсчета или 1920 либо 240 при длине кадра 960 отсчетов. Короткие преобразования (256 и 240) осуществляются на том же интервале кадра 8 раз подряд, при этом используются соответствующие короткие взвешивающие окна. При обнаружении участка резкого изменения уровня сигнала с целью локализации этой динамической особенности данный блок выносит решение о применении коротких взвешивающих окон. Здесь же происходит оценка количества бит, необходимых для кодирования очередного анализируемого кадра.

МДКП – метод модифицированного косинусного преобразования (MDCT – Modified Discrete Cosine Transform). МДКП осуществляет преобразование временного звукового сигнала в частотную область. Формула преобразования выглядит следующим образом:

$$X_k = 2 \sum_{n=0}^{2N-1} w_n x_n \cos\left(\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2}\right) \left(k + \frac{1}{2}\right)\right), \text{ где}$$

X_k -- значения спектральных компонент,
 x_n -- отсчеты входного сигнала,
 w_n -- коэффициенты взвешивающего окна,
 n – временной индекс, $0 \leq n \leq 2N-1$,
 k -- частотный индекс, $0 \leq k \leq N-1$,
 $2N$ -- длина окна преобразования.

Здесь размерность входного сигнала в два раза больше размерности выходного. Это связано с тем, что МДКП в отличие, например, от обычного косинусного преобразования работает с перекрывающимися (на 50%) областями сигнала. Этим решается или, по крайней мере, ослабляется проблема блочности, то есть, появления немаскируемых артефактов на границах преобразования при последовательной обработке сигнала, длительность которого больше, чем длина окна используемого преобразования [3-5]. Кроме этого, для предотвращения разрывов фазы на границах окна преобразования и, как следствие, появления мешающих артефактов в этих областях, в алгоритм AAC вводятся взвешивающие временные окна. Решение о длине окна и типе окна, как уже было сказано, принимается в блоке психоакустической модели кодека.

Метод временного преобразования формы шума квантования (TNS - Temporal Noise Shaping). Данный модуль служит для борьбы с последствиями процедуры квантования сигнала, что проявляется, например, в частом появлении немаскируемых, то есть различных человеческим ухом, пре-эхо сигналов при наличии в сигнале резких фронтов, особенно характерных для речевых сигналов. Суть метода состоит в том, чтобы на основе спектрального анализа звукового сигнала предсказать ошибку квантования (forward prediction или опережающее предсказание) и ввести ее в сигнал (“вычесть” ее). Таким образом, при восстановлении сигнала немаскируемые артефакты, вызванные квантованием, будут частично подавлены [6].

Совмещенное стерео (Joint stereo): Intensity-стерео или M/S-стерео. Данный модуль служит для сокращения информационной избыточности, обусловленной высокой корреляцией между левым и правым каналами стереосигнала. При этом метод Intensity-стерео является алгоритмом сжатия с потерей информации, а M/S – алгоритмом без её потери. Соответствующему блоку декодера передается информация о выбранном типе стереокодирования. Алгоритм Intensity использует корреляционную зависимость между верхними частями спектров стереопары звукового сигнала. При этом в выходной поток передаются нижние части спектров левого и правого каналов в полосе до 6 кГц, а оставшаяся верхняя часть полосы передается в виде масштабированной невозстанавливаемой на стороне декодера суммы спектров левого и правого каналов. Алгоритм M/S (middle/side = среднее/разность) анализирует возможность передачи вместо сигналов левого (L) и правого (R) каналов пары $M=(L+R)/2$ и $S=(L-R)/2$ с точки зрения минимизации выходного битового потока. При этом оцениваются распределения мощностей для обоих вариантов и выносится решение о передаче пары L/R либо M/S.

Блок покадрового предсказания (Prediction). Данный модуль используется для сокращения информационной избыточности, связанной с высокой степенью

корреляции между спектрами последовательных кадров. При этом на стороне кодера используется так называемый алгоритм обратного предсказания (backward prediction), осуществляющий итеративное квантование и восстановление данных с целью предотвращения распространения ошибки предсказания. Добавим, что блок кадрового предсказания включается только тогда, когда при первоначальном анализе звукового сигнала на уровне психоакустической модели принимается решение об отсутствии в данном кадре резких переходов и участков «нестационарности спектра» (так называемые transients).

Блок масштабирования спектральных компонент (Scaling). В этом модуле, на стороне кодера, происходит умножение спектральных компонент на масштабные множители (scalefactors). Данная процедура используется для управления в частотной области формой шумов квантования (во временной области этим занимается блок TNS, см. выше). Сами масштабрующие множители передаются вместе с потоком. Количество коэффициентов определяется разбиением спектра сигнала на так называемые «масштабные полосы» (scalefactor bands), каждой из которых соответствует один коэффициент.

Блок квантования (Quantization) - осуществляет неравномерное квантование спектральных компонент сигнала и разделен на 3 уровня. Верхний уровень занимается вычислением количества бит, в которые нужно «уложить» поток, проверяет не равны ли все спектральные компоненты нулю и, если нет, вызывает процедуру следующего уровня – так называемый внешний итеративный цикл, который в свою очередь вызывает процедуру внутреннего итеративного цикла. Внутренний цикл квантует входной вектор, увеличивая шаг квантования до тех пор, пока выходной вектор все еще укладывается в отведенные ему количество бит. После выполнения внутреннего цикла внешний цикл оценивает искажения, вызванные последней итерацией квантования внутри каждой масштабной полосы (см. описание блока Scaling) и, если ошибка больше пороговой, уменьшает значение масштабующего коэффициента для тех масштабных полос, в которых было допущено превышение порога ошибки.

Блок хаффмановского кодирования (Huffman coding) представляет собой статистический кодер, работающий с парами или квартетами передаваемых квантованных спектральных коэффициентов и с единичными значениями масштабных коэффициентов. Для этого предусмотрено 10 основных кодовых книг кодирования для спектральных компонент и одна для масштабных коэффициентов, различающихся между собой тем, над чем они определены (пары или квартеты значений), максимальным значением входных величин, а также знаковостью самой кодовой книги (знаковая или беззнаковая). Порядок кодирования спектральных коэффициентов – от нижней частоты к верхней.

Блок формирования битового потока (Bitstream formatter). Данный блок компонует выходной битовый поток, который включает в себя следующие основные типы управляющей информации:

- квантованные спектральные коэффициенты звукового сигнала,
- масштабрующие коэффициенты модуля "Scaling",
- информация для восстановления входных уровней звукового сигнала в случае использования на стороне кодера блока "Gain control",
- информация о типе и длине использовавшихся взвешивающих окон,
- информация, необходимая для декодирования временно-го преобразования формы шума (TNS),
- информация об использовавшемся типе объединенного стерео,
- информация об использовавшихся на стороне кодера параметрах хаффмановского кодирования.

Структура декодирующего устройства. Декодирование осуществляется в обратном (относительно кодирования) порядке. Одноименные блоки на рис. 2 представляют собой соответствующие декодеры. На вход декодера поступает упакованный битовый поток, основная часть которого состоит из квантованных и закодированных с помощью хаффмановского кодера спектральных компонент.

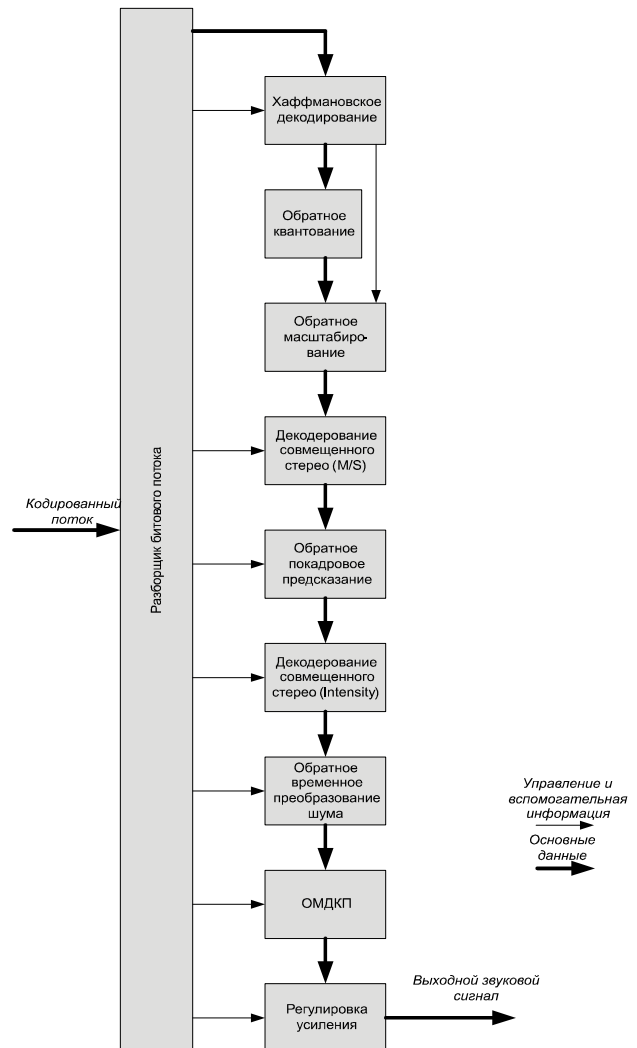


Рис. 2. Схема декодера AAC

Вектор развития стандарта звукового кодера AAC можно обозначить с помощью таких символических формул:

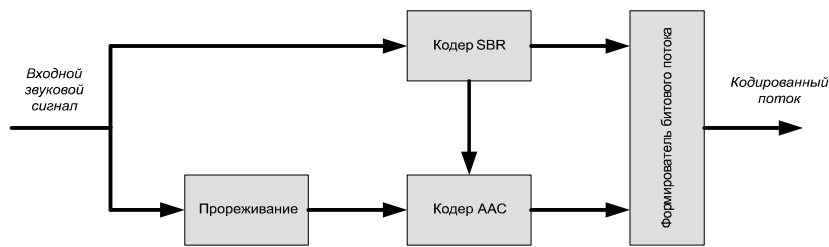


Рис. 3. Общая схема кодера HE-AAC v.1 (SBR)

«HE-AAC v.1» = «AAC» + «SBR»

«HE-AAC v.2» = «AAC» + «SBR» + «PS»

Метод спектрального дублирования или отбраживания (SBR – Spectral Band Replication) основывается на двух вводимых в рамках данной концепции качественных постулатах:

1. В верхней части спектра звукового сигнала содержится в значительной степени избыточная с точки зрения слухового восприятия информация. Для слухового восприятия важны только мощностные характеристики высокочастотных составляющих, а фазовые соотношения могут быть искажены без потери воспринимаемого качества восстановленного звукового сигнала.

2. На некотором временном интервале существует высокая корреляционная зависимость между мощностными характеристиками верхней и нижней половинами спектра звукового сигнала.

На рис. 3 приведена общая схема HE-AAC кодера первой версии.

Суть SBR-алгоритма состоит в том, чтобы не передавать полностью верхнюю половину спектра (коэффициенты МДКП) звукового сигнала. Вместо этого передается информация о мощности спектральных компонент отдельных подполос верхней половины спектра. Кроме этого на стороне передачи анализируется мощность шумовых составляющих этих подполос и их уровень также передается декодеру для возможности восстановления исходного отношения сигнал/шум. Мощность высокочастотных спектральных компонент оценивается линейным предсказателем 2-го порядка с помощью матрицы ковариаций огибающих спектра. Как видно из рис. 3, SBR-кодер работает с непрореженным по времени сигналом, в то время как на AAC-кодер подается уже прореженный в два раза сигнал (блок "Прореживание"). Важно отметить, что блок прореживания должен обладать некоторой "интеллектуальностью", а именно, содержать в себе НЧ-фильтр, предотвращающий зеркальное наложение верхней половины спектра на нижнюю при осуществлении собственно самого прореживания. Помимо перечисленных алгоритмов SBR-кодер отслеживает и оценивает возможность возникно-

вения в высокочастотной части спектра мощных индивидуальных узкополосных сигналов (в стандарте они именуются "синусоидальными компонентами"), образующихся, например, вследствие возникновения в звуковом сигнале участков нестационарности спектра (так называемые "transients"). Информация о таких компонентах спектра также передается декодеру. Таким образом, SBR-декодер получает следующую информацию для восстановления звукового сигнала:

- мощностные коэффициенты подполос верхней половины спектра декодируемого сигнала (так называемые *огигающие спектра*);
- мощностные коэффициенты шума для восстановления соотношения сигнал/шум;
- положение и мощность индивидуальных узкополосных сигналов, не поддающихся линейному предсказанию.

Заметим, что данная информация об огибающих спектра передается не напрямую, а в виде разностей оригинальных и предсказанных значений. Поэтому на стороне декодера вновь осуществляется линейное предсказание по низкочастотной части спектра, получаемой от AAC-декодера, после чего производится восстановление (по мощности) огибающих верхней половины спектра сигнала.

Таким образом, полностью передается только нижняя часть спектра сигнала, а высокочастотные составляющие восстанавливаются по сопроводительной составляющей, занимающей лишь небольшой процент передаваемой информации. Технология SBR получила широкое признание и существует множество источников, кроме самого стандарта [2], в которых можно почерпнуть более подробную информацию о деталях этого алгоритма. Это прежде всего сайт компании Coding Technologies (www.codingtechnologies.com/products/sbr.htm), а также следующий ряд статей по этой теме: [7-10].

Параметрическое стерео кодирование (PS - Parametric Stereo) осуществляет преобразование стереосигнала в моносигнал (downmix), сопровождая его дополнительным набором параметров, позволяющих на стороне декодера восстановить стереобраз. Данное преобразование относится к классу кодирования с информационными потерями. На рис. 4 изображена схема PS-кодера.

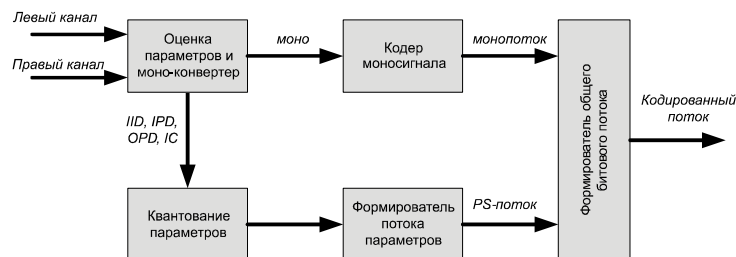


Рис. 4. Схема PS-кодера

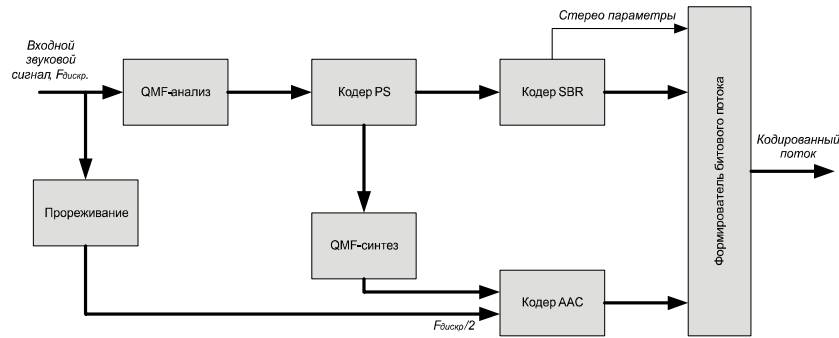


Рис. 5. Схема кодера HE-AAC v.2

В блоке "Parameter estimation" осуществляется оценка следующих четырех типов параметров:

- межканальные разности уровней (IID - Inter-channel Intensity Differences);
- межканальные разности фаз (IPD - Inter-channel Phase Differences);
- общий фазовый сдвиг (OPD - Overall Phase Difference);
- межканальная когерентность (IC - межканальная когерентность).

Последний параметр (IC) измеряется как максимум кросс-корреляции функции фаз или времени. Перечисленных параметров достаточно для того, чтобы восстановить стереообраз на стороне приема. Для восстановления этого стереообраз без заметных на слух искажений под поток этой дополнительной информации достаточно отвести 9-10 кбит/с. При этом в канале будет передаваться только моносигнал (после процедуры преобразования моно-стерео). Перечисленные 4 параметра квантуются (блок "Parameter quantization") и кодируются как дополнительные части монопотока (в блоке "Bit stream multiplexer") таким образом, чтобы обеспечить обратную совместимость при воспроизведении данного потока моно-кодеками, игнорирующими эту дополнительную информацию. Более подробно с технологией параметрического стереокодирования можно ознакомиться, например, в [11-14].

Общая схема HE-AAC-кодека второй версии, включающего в себя кодеры Spectral Band Replication (SBR) и Parametric Stereo (PS) приведена на рис. 5:

Заметим здесь еще раз, что кодеры PS и SBR работают с двойной частотой дискретизации сигнала относительно кодера AAC.

В заключение отметим еще раз, что в ЗАО «НИИР-КОМ» разработан звуковой кодек, совместимый со стандартом HE-AAC v.1. Испытания, проведенные специалистами этой фирмы, подтвердили превосходство стандарта AAC над самым популярным на сегодняшний день форматом сжатия звука MP3 для всех скоростей кодирования и всех типов звуковых сигналов. При этом установлено, что при скорости цифрового потока 24 кбит/с кодек HE-AAC в состоянии обеспечить качество FM-стереовещания, а при скорости 32 кбит/с - качество Audio-CD. Направление дальнейших работ по совершенствованию этого продукта связано с созданием кодека, поддерживающего технологию параметрического стереокодирования (PS) и, следовательно, совместимого с версией HE-AAC v.2, что безусловно позволит добиться повышения качества звука при за-

данной скорости цифрового потока, либо осуществлять передачу звука в каналах с меньшей пропускной способностью, сохраняя при этом качественные характеристики передаваемого звукового сигнала.

Литература

1. ISO/IEC MPEG-2 13818-7 AAC
2. ISO/IEC MPEG-4 14496-3 Audio
3. Henrique S. Malvar, Signal Processing With Lapped Transforms (Artech House: Norwood MA, 1992).
4. John P. Princen and Alan B. Bradley, "Analysis/synthesis filter bank design based on time domain aliasing cancellation", IEEE Trans. Acoust. Speech Sig. Proc. ASSP-34 (5), 1153-1161 (1986).
5. A. W. Johnson and A. B. Bradley, "Adaptive transform coding incorporating time domain aliasing cancellation", Speech Comm. 6, 299-308 (1987).
6. Jurgen Herre, "Temporal Noise Shaping, Quantization and Coding Methods in Perceptual Audio Coding: a Tutorial Introduction", AES 17th International Conference on High Quality Audio Coding
7. Per Ekstrand, "Bandwidth extension of audio signals by spectral band replication", Proc.1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), Leuven, Belgium, November 15, 2002
8. Gilbert A. Soulodre and Michel Lavoie, "Subjective Evaluation of MPEG Layer II with Spectral Band Replication", Audio Engineering Society, Convention Paper 6185, Presented at the 117th Convention, 2004 October 28-31 San Francisco, CA, USA
9. O. Kunz, "SBR explained: White paper", Coding Technologies
10. A. Gröschel, M. Schug, M. Beer and F. Henn, "Enhancing audio coding efficiency of MPEG Layer-2 with Spectral Band Replication for DigitalRadio (DAB) in a backwards compatible way", 114th AES Convention, Amsterdam, 2003, March
11. E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in Proc. 114th AES Convention, Amsterdam, The Netherlands, Mar. 2003, Preprint 5852.
12. C. Faller and F. Baumgarte, "Estimation of auditory spatial cues for binaural cue coding," in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Orlando, Florida, USA, May 2002.
13. J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "High-quality parametric spatial audio coding at low bitrates," in Proc. 116th AES Convention, Berlin, Germany, May 2004.
14. W. Oomen, E. Schuijers, H. Purnhagen, and J. Engdegard, "MPEG4-Ext2: CE on low complexity parametric stereo," ISO/IEC JTC1/SC29/WG11 MPEG2003/M10366, Dec. 2003.