

# Сравнение звуковых форматов AAC и MP3

Алексей Лукин

Первые идеи об использовании психоакустической маскировки для компрессии аудиоданных относятся к 1979 году. Однако соответствующие аудиокодеры начали получать широкое распространение лишь с середины 90-х годов, когда вычислительных мощностей персональных компьютеров стало хватать для воспроизведения сжатого аудио в реальном времени и появился стандарт MPEG-1 Audio Layer 3, более известный как MP3. Аудиоформаты с компрессией стали незаменимыми при передаче звука через интернет, обеспечивая «практически прозрачное» качество стереозвука (то есть закодированный сигнал для большинства слушателей неотличим от оригинала) при битрейтах выше 128 кбит/с. С основными принципами формата MP3 можно познакомиться в статьях К. Гласмана (2...8/2005)

Развитие методов сжатия данных и психоакустики постепенно приводило к тому, что стандарт MP3 стал «тесным» для реализации новых идей в кодировании аудио. В результате, к 1997 году институтом Фраунгофера (Fraunhofer IIS), который в начале 90-х создал MP3, а также компаниями Dolby, AT&T, Sony и Nokia – был разработан новый метод компрессии аудио – Advanced Audio Coding (AAC), вошедший в стандарты MPEG-2 и MPEG-4. Основными отличиями от стандарта MP3 стали:

- поддержка более широкого набора форматов (вплоть до 48 каналов) и частот дискретизации звука (от 8 кГц до 96 кГц);
- более эффективный и простой банк фильтров: гибридный банк фильтров MP3 был заменен обычным MDCT (модифицированным дискретным косинусным преобразованием);
- более широкие пределы варьирования частотно-временного разрешения в банке фильтров – в восемь раз (в MP3 – в три раза) – привели к улучшению кодирования транзиентов (переходных процессов) и стационарных участков аудиосигнала;
- более качественное кодирование частот выше 16 кГц;

- более гибкий режим кодирования стереосигналов, позволяющий переключаться в режим M/S (“joint stereo”) независимо в различных частотных полосах;
- дополнительные возможности компрессии: технология формирования шума во временной области (TNS), предсказание MDCT-коэффициентов по времени (long term prediction), режим параметрического кодирования стереосигнала (parametric stereo), синтез шумов (perceptual noise substitution), технология восстановления высоких частот (SBR).

Благодаря этим особенностям, стандарт AAC способен достигать более гибкого и эффективного, а значит – и более качественного кодирования звука. В результате широкого распространения формата MP3, стандарт AAC до сих пор не приобрел сравнимой с MP3 популярности. Тем не менее AAC является основным форматом в популярном интернет-магазине iTunes Store, плеерах iPod, iTunes, телефоне iPhone, игровых приставках PlayStation 3, Nintendo Wii и в цифровом радиовещании DAB+/DRM.

Рассмотрим основные особенности AAC подробнее.

## Банк фильтров

Как и другие психоакустические аудиокодеры, AAC работает по следующей схеме. Входной сигнал пропускается через *банк фильтров* – преобразование, переводящее сигнал из временной области в частотно-временную область (аналогично построению спектрограммы). Параллельно с этим *психоакустическая модель* анализирует сигнал и определяет пороги психоакустической маскировки. Далее спектральные коэффициенты сигнала на выходе банка фильтров *квантуются* так, чтобы спектр шума по возможности (если позволяет битрейт) оказался ниже порогов маскировки и не был слышен. Квантованные коэффициенты *сжимаются без потерь* в выходной файл формата AAC. Таким образом, сам банк фильтров не сжимает сигнал, он лишь переводит его в форму, более пригодную для сжатия.

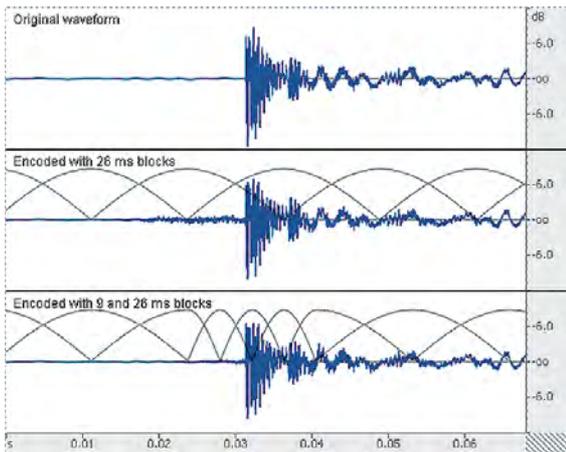


Рис. 1. Эффект пред-эха. Сверху вниз: исходный сигнал; сигнал после сжатия с фиксированным размером окна (видно сильное пред-эхо); сигнал после сжатия с переменным размером окна (пред-эхо уменьшилось)

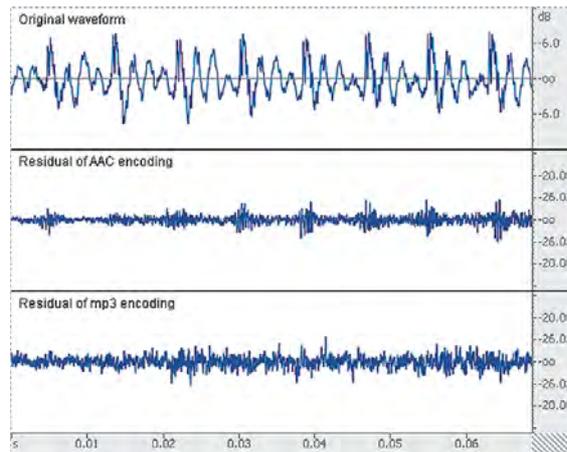


Рис. 2. Технология TNS: формирование амплитудной огибающей шума квантования. Сверху вниз: исходный сигнал; ошибка квантования алгоритмом AAC с технологией TNS; ошибка квантования алгоритмом MP3 с тем же битрейтом

Особенностью каждого банка фильтров является его частотное разрешение, то есть число частотных полос, на которые он делит спектр сигнала. В большинстве банков фильтров, используемых для сжатия звука, число полос составляет несколько сотен. Это означает, что в силу соотношения неопределенностей такие банки фильтров имеют временное разрешение порядка нескольких десятков миллисекунд. Когда спектральные коэффициенты сигнала квантуются, то вносимая ошибка квантования при декодировании сигнала распространяется по времени на всю длину окна банка фильтров. В некоторых случаях это приводит к нежелательному эффекту, называемому пред-эхом (pre-echo). Он проявляется, когда ошибка квантования от транзientа (резкого всплеска энергии в сигнале) распространяется по времени на предшествующий транзientу участок времени и становится слышна (рис. 1). Чтобы уменьшить этот эффект, применяют банки фильтров с переменным частотно-временным разрешением. Например, в MP3 используется переключение временного разрешения банка фильтров между 26 и 9 мс. Для стационарных сигналов используются окна длиной 26 мс, дающие хорошее частотное разрешение, а для транзиентов используются окна длиной 9 мс, уменьшающие эффект пред-эха (см. рис. 1).

В алгоритме AAC также используется переключение размера окон MDCT. При этом разница в размере окон восьмикратная: 6 и 48 мс (256 и 2048 отсчетов). Благодаря этому алгоритм способен адаптироваться к более широкому диапазону сигналов и достигать лучшей степени компрессии.

### Технология TNS – формирование амплитудной огибающей шума

Одной из проблем современных психоакустических кодеров аудиосигнала является работа с транзиентами (переходными процессами в аудиосигнале). Для обеспечения прозрачного кодирования нужно обеспечить попадание шума квантования под порог маскировки, зависящий от времени. Однако на практике этому требованию трудно удовле-

творить вблизи переходных процессов, т.к. шум квантования, возникший при кодировании, распространяется по времени при декодировании на всю длину окна MDCT. Это может приводить к значительным превышениям шумом квантования порогов временной маскировки.

Технология TNS (temporal noise shaping, формирование шума во временной области) в стандарте AAC позволяет управлять распространением шума квантования по времени в пределах каждого окна MDCT. Технология TNS основана на подобию (частотно-временном дуализме) амплитудной огибающей сигнала и огибающей его спектра, а также на использовании линейного предсказания (LPC) по частоте при квантовании спектра.

Хорошо известно, что для сигналов со спектром, сильно отличающимся от белого (например, тональных сигналов), использование линейного предсказания (LPC) во временной области позволяет эффективно «отбеливать» спектр и кодировать такие сигналы путем их разложения на коэффициенты предсказания и сравнительно небольшую по амплитуде ошибку предсказания (residual). При декодировании фильтр линейного предсказания формирует спектр ошибки согласно спектру исходного сигнала.

В кодере AAC линейное предсказание используется противоположным образом: для предсказания отсчетов спектра в частотной области. Разность исходных и предсказанных коэффициентов MDCT квантуется согласно порогам маскировки (в традиционных кодерах квантуются исходные коэффициенты MDCT). Коэффициенты линейного предсказания также записываются в выходной файл. При декодировании сигнала фильтр линейного предсказания, применяемый к разностному сигналу в частотной области (включающему ошибку квантования), формирует амплитудную огибающую исходного сигнала (и ошибки квантования) во временной области. Таким образом, амплитудная огибающая ошибок квантования становится близкой к амплитудной огибающей исходного сигнала (рис. 2).

Технология TNS снижает эффект пред-эха и заметность ошибок квантования на некоторых гармонических сигналах с импульсным характером звукоизвлечения (речь, некоторые духовые и струнно-смычковые инструменты). На рис. 2 сравниваются ошибки квантования, вносимые в вокальный сигнал алгоритмами AAC и MP3 с одинаковыми битрейтами. Вместе с общим понижением ошибки квантования (в силу большей эффективности AAC) наблюдается формирование амплитудной огибающей ошибки квантования по времени согласно огибающей исходного сигнала.

В стандарте AAC технология TNS может применяться к отдельным частотным полосам спектра независимо или отключаться совсем.

### **Технология SBR – восстановление высоких частот**

Достоверная передача широкого частотного диапазона – важное требование для качественного кодирования. Однако передача каждой следующей октавы звукового диапазона в полтора-два раза повышает требования к битрейту для традиционного аудиокодера. Чтобы снизить битрейт и при этом сохранить высокие частоты в кодируемом материале, была создана технология искусственного синтеза высоких частот SBR (spectral band replication).

Технология основывается на том, что наш слух анализирует высокие частоты с меньшей точностью, чем средние и низкие. Для создания эффекта присутствия высоких частот необязательно математически точно реконструировать форму волны, а достаточно лишь восстановить некоторые существенные психоакустические параметры сигнала на высоких частотах. К таким существенным параметрам относятся частотно-временное распределение (огибающая) энергии сигнала и степень его тональности/зашумленности.

Идея алгоритма такова. При кодировании осуществляется анализ высоких частот в исходном аудиосигнале и извлекаются их параметры: в первую очередь – амплитудная огибающая в нескольких (обычно в восьми) частотных полосах. Далее высокие частоты из записи удаляются и кодируются только оставшиеся низкие и средние частоты. При этом в выходной файл также добавляется сравнительно небольшой поток информации о параметрах утраченных высоких частот.

При воспроизведении сначала декодируется сигнал низких и средних частот. Далее (в случае его наличия в плеере) начинает работу декодер SBR. Первым шагом он осуществляет синтез высокочастотного сигнала путем транспонирования (точнее – частотного сдвига) имеющихся средних частот. Поскольку степень тональности/зашумленности спектра на средних и высоких частотах примерно равна, то в результате этого шага получается высокочастотный сигнал с правдоподобной структурой спектра. На втором шаге декодер SBR использует дополнительную сохраненную информацию о высоких частотах для придания им

нужной амплитудной огибающей в каждой частотной полосе. В результате получается сигнал, у которого высокие частоты полностью синтезированы из средних, но при этом сохраняют звучание исходных высоких частот.

Технология SBR может быть пристроена ко многим существующим методам кодирования аудио. Например, SBR в сочетании с MP3 называется MP3 PRO, а SBR в сочетании с AAC называется HE-AAC (high efficiency AAC). В основном, SBR используется при кодировании с относительно низкими битрейтами: 64 кбит/с и ниже. Технология позволяет значительно расширить частотный диапазон аудиосигнала с минимальным увеличением битрейта (несколько кбит/с).

### **Технология Parametric stereo**

Передача стереосигнала обычно требует от кодера почти в 2 раза большего битрейта, чем передача монофонического сигнала. При этом стереоканалы можно кодировать как независимо, так и после M/S преобразования. В последнем случае на S-канал зачастую тратится меньший битрейт, чем на M-канал. Этот режим кодирования также называется joint stereo. В стандарте AAC этот режим может включаться и отключаться кодером независимо для каждой частотной полосы.

Для более эффективного кодирования стереосигналов на совсем низких битрейтах (16...32 кбит/с) была разработана технология параметрического кодирования стереопанорамы (parametric stereo). Она заключается в том, что стереосигнал перед кодированием сводится к моно, но в выходной файл добавляется небольшой поток (2...3 кбит/с), содержащий информацию о стереопанораме исходного стереофайла. Этот поток содержит (в сжатом виде) своеобразную «карту панорамы» для частотно-временной плоскости.

На стадии декодирования к полученному монофоническому сигналу применяется частотно-зависимое панорамирование. Это можно производить одновременно с декодированием, применяя к изначально равным коэффициентам MDCT левого и правого каналов соответствующие амплитудные множители.

Технология Parametric stereo дает хорошее впечатление об исходной стереопанораме звука ценой лишь небольшого увеличения битрейта по сравнению с кодированием моносигнала. Однако она не позволяет добиться полностью прозрачного звучания, так как неспособна учесть все нюансы стереопанорамы, например фазовые сдвиги между стереоканалами.

Технология Parametric stereo была включена в стандарт HE-AAC v2.

### **Технология PNS – генерация шумов**

Для дополнительного увеличения эффективности кодирования шумовых сигналов в стандарте AAC предусмотрена технология PNS (perceptual noise substitution) для синтеза шумов. Известно, что наше ухо более чувствительно к амплитудному спектру сигнала, чем к фазовому. Поэтому вме-

сто кодирования MDCT-коэффициентов исходного сигнала в шумовых областях можно лишь передавать параметры шума: его мощность в зависимости от частоты и времени.

Так и работает технология PNS. При кодировании идентифицируются участки спектра, представляющие собой шум, и соответствующие группы MDCT-коэффициентов исключаются из процесса кодирования. Частотная полоса помечается как шумовая, и для нее запоминается общая энергия шума.

При декодировании в частотные полосы, помеченные как шумовые, подставляются псевдослучайные MDCT-коэффициенты с требуемой общей мощностью. В результате в указанных частотных диапазонах синтезируется шум, близкий по звучанию к исходному шуму.

### **Технология Long term prediction – предсказание по времени**

Психоакустическое кодирование тональных сигналов требует более высокого локального отношения сигнал/шум, чем кодирование шумовых сигналов (например, 20 дБ и 6 дБ соответственно). А это, в свою очередь, требует повышенного битрейта. Однако MDCT-коэффициенты тональных сигналов являются предсказуемыми по времени. Это обстоятельство позволяет эксплуатировать их зависимость по времени для уменьшения битрейта.

В стандарте AAC предусмотрен режим Long term prediction, в котором MDCT-коэффициенты дополнительно кодируются по времени с помощью линейного предсказания. Термин “long term” означает, что предсказание осуществляется не по соседним отсчетам, а по отсчетам, отстоящим на наиболее вероятный период тона на данной частоте.

### **Квантование и сжатие MDCT-коэффициентов**

Аналогично стандарту MP3, в AAC используется нелинейное квантование MDCT-коэффициентов и сжатие их методом Хаффмана. Коэффициенты MDCT квантуются после возведения в степень 0,75, что позволяет увеличивать ошибку квантования для мощных сигналов и уменьшать ее для слабых сигналов в пределах каждой частотной полосы. Таким образом осуществляется дополнительное неявное формирование спектра шума.

После квантования MDCT-коэффициенты сжимаются с помощью набора фиксированных таблиц Хаффмана. В стандарте AAC этих таблиц больше, чем в MP3, и шире возможности по группировке коэффициентов. Это приводит к дополнительному увеличению сжатия.

### **Качество звучания**

При оценке качества звучания аудиокодеров обычно используются субъективные тесты. Слушателям представляются фрагменты сжатых разными кодерами записей, и они оценивают чистоту звучания каждого фрагмента по шкале от 1 до 5. Лучшим кодером считается тот, который способен достичь более высокого качества звучания по сравнению с конкурентами при заданном битрейте.

Достаточно авторитетным интернет-источником, где приведены результаты таких тестов, является сайт <http://www.rjmorim.com/test/>. На нем представлены тестирования различных кодеков на множестве битрейтов. Приведенные результаты в целом хорошо согласуются с другими источниками. Приведем несколько результатов для кодеров MP3 и AAC, помогающих сравнить их качество.

Лучшим кодером MP3 является бесплатный Lame. Однако на большинстве битрейтов он уступает в качестве более новым стандартам сжатия. На высоких битрейтах (выше 128 кбит/с) это отставание невелико, и лидером является кодер Ogg Vorbis.

На битрейте 64 кбит/с преимущество AAC уже становится ощутимым. В варианте HE-AAC алгоритм зарабатывает оценку 3,68. Это примерно соответствует Lame с битрейтом 96 кбит/с и означает преимущество AAC над MP3 примерно в 1,5 раза. Оценка Lame с битрейтом 128 кбит/с – 4,29.

На битрейте 32 кбит/с кодер AAC от компании Nero серьезно выигрывает в качестве по сравнению с MP3: оценки 3,23 и 1,72 соответственно. Однако AAC лишь ненамного опережает формат MP3PRO, получивший оценку 3,08. Это указывает, что технология SBR действительно значительно улучшает качество при низких битрейтах.

### **Выводы**

Благодаря примененным в стандарте AAC новым технологиям, данный формат имеет заметное преимущество перед MPEG-1 Layer 3 (MP3), позволяя достигать лучшего качества звука при таких же битрейтах. Особенно сильный выигрыш наблюдается в области низких битрейтов: 96 кбит/с и ниже. Это подтверждает перспективность формата AAC для цифрового радиовещания.

Популярность AAC для распространения музыки в интернете на сегодняшний день остается низкой по сравнению с форматом MP3. Пользователи продолжают предпочитать лучшую переносимость MP3 более сильному сжатию AAC. Значительная часть музыкальных архивов на сайтах, распространяющих музыку, уже изначально находится в формате MP3, и доступа к несжатым записям у провайдеров не имеется. Это значит, что перекодировать такие записи в формат AAC большого смысла не имеет – качество зачастую уже потеряно. Однако новые карманные плееры и некоторые онлайн-магазины уже поддерживают формат AAC, часто – с верификацией легальности контента (что также отпугивает пользователей, предпочитающих не ограничивать себя в копировании музыки).

Будучи весьма перспективным, формат AAC не является единственным наиболее качественным форматом компрессии звука. На высоких битрейтах (выше 128 кбит/с) AAC часто уступает в качестве кодерам форматов Ogg Vorbis и Musepack. На самых низких битрейтах (менее 32 кбит/с) AAC может уступать параметрическим кодерам звука, в том числе – специализированным кодерам для сжатия речи. Однако в диапазоне средне-низких битрейтов AAC на данный момент сохраняет пальму первенства. ●