ЧАСТЬ I

КОМПЬЮТЕР КАК СТУДИЯ

Прежде чем монтировать звук, нужно понимать инструмент. Эта часть о том, как компьютер хранит звук в виде чисел и что из этого следует на практике, как устроена звуковая станция и откуда в ней берутся задержки, как звук и изображение договариваются о времени и как материал переходит от площадки и монтажа изображения к звукорежиссёру без потерь.

Глава 1. Звук в цифре
Глава 2. Компьютер как студия
Глава 3. Время фильма
Глава 4. Файлы и обмен

ГЛАВА
1 ЗВУК В ЦИФРЕ

[Уайатт, Эмиес]

Всё, с чем работает звукорежиссёр в компьютере, — это числа. Микрофон превращает колебания воздуха в колебания электрического напряжения. Затем *аналого-цифровой преобразователь* (АЦП; англ. *analog-to-digital converter*, ADC) много тысяч раз в секунду измеряет это напряжение и записывает каждое измерение числом. При прослушивании *цифро-аналоговый преобразователь* (ЦАП; *digital-to-analog converter*, DAC) делает обратное: превращает числа снова в напряжение, которое идёт на усилитель и громкоговорители. Такие преобразователи стоят в каждой звуковой карте, в каждом рекордере и в каждой камере. Между ними звук существует только как последовательность чисел, и всё, что с ним происходит, — монтаж, обработка, сведение — это арифметика.

Зачем звукорежиссёру знать, как устроены эти числа? Кажется, достаточно один раз договориться: все записывают и сдают звук в 48 кГц и 24 бита, и проблем не будет. Внутри одной студии так и бывает. Но фонограмма фильма собирается из материала, который приходит отовсюду. Композитор присылает музыку в 44,1 кГц, потому что так привыкла работать музыкальная индустрия. Библиотека эффектов записана в 96 или 192 кГц. Рекордер с площадки отдаёт 32-битные файлы. Режиссёр приносит архивную запись, фрагмент из интернета или голосовое сообщение с телефона. А на выходе нужна не одна версия, а несколько: для кинотеатра, для телевидения, для платформы, иногда для бытового формата. На каждом таком стыке приходится решать, можно ли просто положить файл на дорожку, нужно ли его преобразовать, чем и что при этом потеряется.

Ошибка на стыке не всегда слышна сразу. Музыка может незаметно заиграть быстрее и выше, чем написана. Пик, которого не видно на индикаторе, может стать причиной отказа в приёмке. Звук, несколько раз пересжатый для передачи, теряет чистоту верхних частот. Понять, откуда всё это берётся и что с этим делать, можно, если разобраться в трёх вещах: как часто измеряется звук, насколько точно записывается каждое измерение и что происходит, когда звук сжимают для передачи.

> **Биографическая справка**
> Хилари Уайатт (*Hilary Wyatt*) — британский монтажёр звука: с 1987 года монтирует диалоги и эффекты для кино и телевидения, на британском фильме «Sexy Beast» (2000) руководила всей работой над диалогами (*dialogue supervisor*). Тим Эмиес (*Tim Amyes*) — звукооператор и звукорежиссёр перезаписи, работал на Scottish Television и в сети ITV, преподавал звук для кино и телевидения. Их книга «Audio Post Production for Television and Film» выдержала несколько изданий; русский перевод третьего издания вышел в 2006 году под названием «Монтаж звука в теле- и кинопроизводстве». Многие главы этой книги опираются на неё, а то, что за два десятилетия изменилось, дополнено современными источниками.

Отсчёты и частота дискретизации

Одно измерение называется *отсчётом* (*sample*), а число отсчётов в секунду — *частотой дискретизации* (*sample rate*). Это похоже на киноплёнку. Камера не снимает движение непрерывно, а фиксирует 24 отдельных кадра в секунду, и при показе глаз соединяет их в движение. Звук «снимается» так же, только намного чаще. В кино и на телевидении стандартом служит 48 000 отсчётов в секунду, то есть 48 кГц (килогерц — тысяча раз в секунду). Музыкальная индустрия со времён компакт-диска работает на 44,1 кГц, а при записи эффектов и в обработке нередко применяют 96 кГц и выше.

Естественный вопрос: что происходит между отсчётами? Ведь кадры фильма тоже теряют то, что было между ними. Ответ даёт *теорема отсчётов*. В России её называют теоремой Котельникова, за рубежом — теоремой Найквиста — Шеннона.

Чтобы понять теорему, нужно одно представление о звуке. Любой звук, даже очень сложный,

можно разложить на простые колебания разной частоты. Низкий гул — это колебания, повторяющиеся несколько десятков раз в секунду. Человеческий голос в основном занимает область от сотни до нескольких тысяч колебаний в секунду. Шипящие, звон, шелест листвы поднимаются ещё выше, до десятка с лишним тысяч. Чем выше частота колебания, тем выше звук. Человек в лучшем случае слышит колебания примерно до 20 000 в секунду, то есть до 20 кГц, и с возрастом эта граница опускается.

Теперь сама теорема. Если в звуке нет колебаний с частотой выше половины частоты дискретизации, то по отсчётам его можно восстановить *точно*, вместе со всем, что происходило между ними. Проще говоря, на каждое колебание звука должно приходиться больше двух отсчётов. Тогда отсчёты успевают проследить даже самое быстрое колебание, которое есть в звуке, а более медленные — тем более, и между отсчётами не остаётся ничего, что могло бы потеряться. Половину частоты дискретизации называют *частотой Найквиста*. При 48 кГц это 24 кГц: такая запись сохраняет все колебания до 24 000 в секунду, то есть весь слышимый диапазон с небольшим запасом. Поэтому стандартные частоты дискретизации и выбраны чуть больше удвоенных 20 кГц.

Историческая справка

Мысль о том, что непрерывный сигнал можно передать отдельными значениями, созревала в нескольких странах независимо. В 1915 году британский математик Эдмунд Уиттекер (*Edmund Whittaker*) вывел формулу, по которой функция восстанавливается по своим значениям в равноотстоящих точках. К связи и звуку он её не применял. В 1928 году американский инженер Гарри Найквист (*Harry Nyquist*), исследуя телеграфную связь в компании AT&T, показал, что по линии с ограниченной полосой частот можно передать не больше определённого числа независимых импульсов в секунду — вдвое больше ширины этой полосы. В 1933 году молодой советский инженер Владимир Котельников подготовил к Первому Всесоюзному съезду по вопросам технической реконструкции дела связи работу «О пропускной способности „эфира“ и проволоки в электросвязи». В ней теорема впервые была сформулирована так, как её понимают сегодня: сигнал, в котором нет частот выше некоторой границы, полностью определяется отсчётами, взятыми с удвоенной частотой этой границы. За пределами СССР работу долго не знали. В 1949 году американский математик Клод Шеннон (*Claude Shannon*), основатель теории информации, опубликовал статью «Связь при наличии шума» с ясным доказательством теоремы, и в англоязычной литературе она носит его имя вместе с именем Найквиста. Котельников впоследствии стал академиком и руководил, в частности, радиолокационными исследованиями планет. До звука теорема дошла нескоро. Передачу сигнала цифровыми кодами отсчётов, *импульсно-кодовую модуляцию*, изобрёл в конце 1930-х годов англичанин Алек Ривз (*Alec Reeves*), но цифровые звукозаписывающие аппараты появились лишь на рубеже 1960–1970-х, когда электроника стала достаточно быстрой.

Наложение спектров

Условие «нет колебаний выше половины частоты дискретизации» — не формальность. В кино есть знакомый пример того, что бывает, когда его нарушают. Колесо кареты в старом вестерне вращается вперёд, а на экране кажется, что оно медленно крутится назад или стоит на месте. Камера делает 24 кадра в секунду, а спицы движутся быстрее, чем кадры успевают это проследить: между двумя кадрами спица почти доходит до места соседней, и глаз соединяет кадры в ложное медленное движение.

Со звуком происходит то же самое. Колебание с частотой выше частоты Найквиста не исчезает, а записывается как другое, более низкое колебание, которого в исходном звуке не было. На рис. 1.1 тон 40 кГц, записанный при 48 кГц, дал те же самые отсчёты, что и тон 8 кГц. По отсчётам их уже нельзя различить, и при воспроизведении прозвучит 8 кГц. Это явление называется *наложением спектров* (*aliasing*, на профессиональном жаргоне — «алиасинг»). Частота «отражается» от частоты Найквиста, как от зеркала: 40 кГц превращаются в $48 - 40 = 8$ кГц, 30 кГц — в 18 кГц. В отличие от исходного звука, отражённые призвуки не связаны с ним музыкально: они не совпадают с его гармониками и поэтому звучат чужеродно, металлически.

Поэтому перед аналого-цифровым преобразователем стоит *антиалиасинговый фильтр*: он убирает всё выше частоты Найквиста до того, как сигнал будет измерен. На выходе работает фильтр другого назначения — *восстанавливающий* (*reconstruction filter*). Он сглаживает выход ЦАП и убирает копии спектра выше частоты Найквиста, которые возникают при воспроизве-

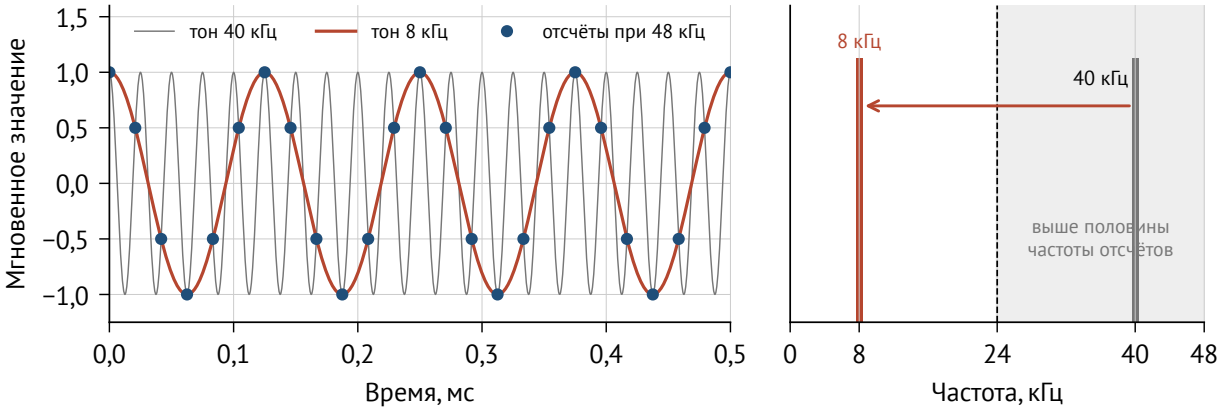

Рис. 1.1. Наложение спектров. Слева: тон 40 кГц (серый) и тон 8 кГц (красный) проходят через одни и те же отсчёты при частоте дискретизации 48 кГц, поэтому после записи они неразличимы. Справа: та же картина на шкале частот — составляющая выше 24 кГц «отражается» вниз. Сигналы расчётные.

дении. Наложение спектров он не устраняет и не создаёт: оно может появиться только при дискретизации. В современных преобразователях оба фильтра встроены, и при записи наложение почти не встречается.

Сегодня наложение спектров возникает в другом месте — внутри компьютера, и звукорежиссёр сталкивается с ним чаще, чем думает. Представим обычную ситуацию: на голос поставили плагин сатурации, чтобы придать ему плотности, и на звуках «с» и «ш» появился звенящий металлический призвук, которого в записи не было. Это не свойство голоса и не «характер» плагина, а брак, и устраняется он легко. Плагины, которые искажают звук намеренно (сатурация, эмуляция ленты или лампы, жёсткие лимитеры и клипперы), добавляют к нему новые высокие гармоники. Если гармоника оказалась выше 24 кГц, она отражается вниз и встаёт между настоящими гармониками голоса. Для борьбы с этим плагины используют *передискретизацию* (*oversampling*): внутри они работают на повышенной частоте, убирают всё лишнее фильтром и только потом возвращаются к 48 кГц. Переключатель «2×», «4×» или «HQ» в плагине — это именно она. Проверка простая: включить передискретизацию и сравнить. Если призвук исчез, это было наложение спектров, и передискретизацию оставляют включённой. Цена — большая нагрузка на процессор и иногда небольшая задержка (о задержках — в главе 2).

Для любопытных

Теорема отсчётов говорит больше, чем «нужно хотя бы два отсчёта на период». Она даёт формулу, по которой непрерывный сигнал вычисляется по отсчётам $x[n]$, взятым с периодом $T = 1/f_s$:

$$x(t) = \sum_n x[n] \cdot \mathrm{sinc}\left(\frac{t - nT}{T}\right), \qquad (1.1)$$

где $\mathrm{sinc}(u) = (\sin \pi u)/(\pi u)$. Каждый отсчёт вносит в результат функцию «sinc», а их сумма проходит и через все промежуточные точки. Именно поэтому цифровой сигнал на выходе ЦАП не «ступенчатый»: ступеньки, которые рисуют некоторые программы при сильном увеличении, — лишь способ показать отсчёты, а не то, что звучит. Частота наложения для тона f при $f_s/2 < f < f_s$ равна $f_s - f$.

Когда частоты не совпадают

Для фильма выбирают 48 кГц. С этой частотой работают монтаж изображения, видеоформаты и требования к сдаче, и она удобно делится на частоты кадров: на один кадр при 24 к/с приходится ровно 2000 отсчётов, при 25 к/с — ровно 1920. Частота 44,1 кГц пришла из музыкальной индустрии вместе с компакт-диском, и музыка до сих пор часто приходит именно в ней. Главное правило проекта: все файлы в нём должны быть в одной частоте, а преобразование частоты делается намеренно и хорошим алгоритмом, а не случайно.

Что бывает при нарушении этого правила, легко представить. Композитор прислал трек в 44,1 кГц, его положили на дорожку проекта 48 кГц, а программа приняла его отсчёты за отсчёты 48 кГц, то есть стала проигрывать 48 000

отсчётов в секунду вместо 44 100. Музыка зазвучит быстрее в $48/\,44{,}1 \approx 1{,}088$ раза, то есть на 8,8 %: трёхминутный трек закончится почти на 15 секунд раньше. Высота поднимется почти на полтора полутона. Музыкант заметит это сразу, а монтажёр, который не слышал трек раньше, может и не заметить, пока не начнёт разъезжаться синхрон. Это не художественный эффект и не «особенность файла», а брак импорта. Современные программы обычно предупреждают о разнице частот или преобразуют файл сами, но при обмене материалом между программами и студиями такая ошибка встречается до сих пор. Проверить просто: посмотреть частоту файла в его свойствах и сравнить с частотой проекта.

Правильный путь — *преобразование частоты дискретизации* (*sample rate conversion*, SRC; на жаргоне — «ресемплинг»). Оно не меняет ни скорость, ни высоту звука: программа вычисляет, какими были бы отсчёты, если бы тот же звук сразу записали в 48 кГц. По сути, это та же теорема в действии: по старым отсчётам восстанавливается непрерывный сигнал, и он заново измеряется в новых точках. Сделать это можно несколькими способами. Большинство станций предлагают преобразование при импорте: в Pro Tools это флажок Apply SRC в окне импорта, в Cubase и Nuendo — параметр преобразования к частоте проекта в настройках импорта. Reaper умеет проигрывать в проекте файлы любой частоты, пересчитывая их на лету; качество такого пересчёта задаётся параметром Resample mode в настройках, и для окончательного рендера там же выбирают самый качественный режим. Наконец, файлы можно заранее преобразовать отдельной программой, например iZotope RX или SoX.

Портит ли преобразование звук? Хороший современный алгоритм делает переход 44,1 → 48 кГц практически прозрачным: его погрешности лежат далеко ниже шума любой реальной записи. Вредит не одно преобразование, а лишние: туда и обратно несколько раз, через разные программы, особенно быстрыми алгоритмами низкого качества. Поэтому файл преобразуют один раз, в начале работы, и дальше используют результат.

Файлы 96 и 192 кГц из библиотек эффектов тоже приводят к частоте проекта. Но если звук собираются сильно замедлять или понижать, например делать рёв чудовища из визга или гул из скрипа, это лучше сделать до преобразования. При понижении на октаву колебания 30–40 кГц опускаются в слышимую область и добавляют звуку деталей. В записи 48 кГц их нет в принципе, а в записи 96 кГц они могут быть — но только если их передали микрофон и аналоговый тракт. Многие микрофоны почти не воспринимают звук выше 20 кГц, и тогда повышенная частота ничего не добавит; кроме того, у самого источника должно быть что-то в этой области. Поэтому перед такой обработкой стоит посмотреть на спектр файла, а не полагаться на число в его свойствах. Для речи и большинства эффектов повышенная частота не нужна.

Преобразование частоты не следует путать с намеренным изменением скорости на 0,1 % или 4 %, которое делают при переходе между частотами кадров изображения. Там звук действительно должен стать чуть быстрее или медленнее, чтобы остаться в синхроне; об этом — в главе 3.

Разрядность: насколько точно записан отсчёт

Каждый отсчёт — это число, а число записывается ограниченным количеством двоичных разрядов, бит. Это похоже на линейку: чем больше на ней делений, тем точнее измерение. *Разрядность* (*bit depth*) определяет, сколько «делений» имеется между самым тихим и самым громким значением. При 16 битах их $2^{16} = 65\,536$, при 24 битах около 16,8 миллиона. Каждый отсчёт округляется до ближайшего деления, и ошибка этого округления звучит как очень тихий шум — *шум квантования* (квантованием называют само округление).

У этого шума есть простое правило: каждый бит добавляет около 6 дБ разницы между самым громким сигналом и шумом. Для синусоиды на полной шкале отношение сигнала к шуму квантования равно примерно $6{,}02N + 1{,}76$ дБ, где N — число бит: 98 дБ при 16 битах и 146 дБ при 24 (рис. 1.2, слева). Для сравнения: разница между порогом слышимости и болевым порогом составляет около 120–130 дБ. Формула верна, когда ошибка округления ведёт себя как случайный шум, то есть для достаточно громкого и сложного сигнала или при дизеринге (о нём ниже), и относится ко всей полосе до частоты Найквиста. Реальные преобразователи до 146 дБ не дотягивают: их собственный шум ограничивает диапазон примерно 110–125 дБ. Значит, при 24 битах

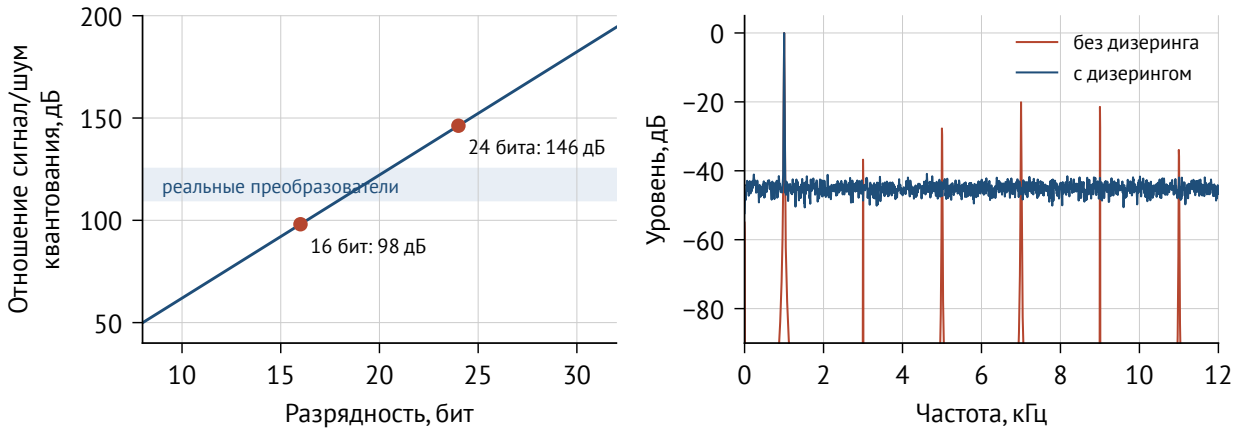

Рис. 1.2. Слева: теоретическое отношение сигнала к шуму квантования в зависимости от разрядности. Затенённая полоса — примерный диапазон реальных преобразователей. Справа: спектр тихого тона 1 кГц (−34 дБ от полной шкалы), уменьшенного до 8 бит без дизеринга (красный) и с дизерингом (синий). Разрядность занижена нарочно, чтобы эффект был виден.

последние разряды записывают уже шум электроники, и шум квантования в обычной работе практически не слышен.

Что из этого следует на практике? Запись на площадке и работа над фильмом ведутся в 24 битах. Запас позволяет записывать с осторожным уровнем, оставляя место для неожиданного крика или удара, и не бояться, что тихий звук утонет в шуме. 16-битный файл, например из старой библиотеки или архива, можно спокойно положить в 24-битный проект: станция дополнит его нулями, звук не изменится, но и лучше не станет. Отдельно преобразовывать его не нужно.

Плавающая точка внутри компьютера

Внутри современных программ звук хранится не в целых числах, а в числах *с плавающей точкой* (*floating point*), обычно 32-битных. Такое число состоит из 24 значащих разрядов и множителя-порядка, как запись $1{,}234 \times 10^5$. Точность остаётся примерно той же, что у 24 бит, но диапазон огромен: значение может превысить «полную шкалу» в тысячи раз, и ничего не обрежется.

Это объясняет ситуацию, которая многих пугает зря. Индикатор шины внутри программы показывает +4 dBFS, горит красным, а на выходе звук чистый. Брак ли это? Нет. Внутри станции звук считается с плавающей точкой и не срезается, поэтому достаточно убавить уровень дальше по цепи, например на мастер-шине. Испортить звук перегрузкой можно только в трёх местах: на входе преобразователя при записи, на выходе к ЦАП и при сохранении в файл с целыми числами (16 или 24 бита). В этих трёх точках шкала снова становится жёсткой, и всё, что выше 0 dBFS, срезается.

Появились и полевые рекордеры, записывающие 32-битные файлы с плавающей точкой. Многие из них используют два преобразователя с разной чувствительностью и объединяют их сигналы, так что внезапный крик или выстрел меньше рискует перегрузить запись. Это сильно облегчает работу звукооператора, но не отменяет перегрузку микрофона и его предусилителя: если аналоговый сигнал искажён до преобразователя, плавающая точка этого не исправит.

Что делать с такими файлами в монтаже? Обычно ничего особенного: современные станции открывают их как есть, и проект можно вести в 32 битах с плавающей точкой. Нужно помнить одну особенность. В таком файле уровень может оказаться выше 0 dBFS: крик, записанный «за шкалой», сохранился целиком, но если файл перевести в 24 бита или проиграть через тракт с целыми числами, его вершины срежутся. Поэтому громкие места сначала опускают уровнем фрагмента или нормализацией и только потом, если этого требует программа или заказчик, переводят файл в 24 бита.

Дизеринг

Когда разрядность уменьшают, например при сохранении сведения из 32 бит в 16 для бытовой версии, отсчёты снова округляются, на этот раз грубее. На громком звуке ошибка округления ведёт себя как случайный шум. Но на очень тихом и простом звуке, например в конце затухания чистого тона, она может перестать быть случайной:

ошибка начинает повторяться вместе с волной и превращается в искажения — гармоники, связанные с сигналом. На рис. 1.2 справа видно, что тихий тон без дизеринга оброс гармониками на 3, 5, 7 кГц и далее. Это расчёт для чистого тона; у шумного или сложного звука ошибка ближе к случайной, и такие искажения проявляются слабее или не проявляются вовсе.

В работе это выглядит так. Сведение экспортировали в 16 бит, и там, где звук уходит в тишину, — в хвосте отзвучавшего колокола, в конце музыкальной фразы, в затухании реверберации — появился шершавый хрип, а сам звук обрывается раньше, чем должен. Это брак, и устраняется он одной настройкой при экспорте.

Дизеринг (*dither*) — это добавление перед округлением очень слабого, специально подобранного шума. Он разрушает связь ошибки с сигналом: вместо гармоник остаётся ровный тихий шум, а тихий звук продолжает слышаться внутри этого шума, а не превращается в хрип. Звучит как парадокс: шум, добавленный намеренно, делает запись чище. Но слух намного терпимее к ровному шуму, чем к искажениям, которые меняются вместе со звуком. Существуют и формы дизеринга с перераспределением шума по частотам (*noise shaping*), при которых шум сдвигается туда, где слух менее чувствителен.

Когда нужен дизеринг? Обязательно — при уменьшении разрядности до 16 бит, например при экспорте бытовой версии. При экспорте в 24 бита из 32-битного проекта многие его тоже включают, и это правильная привычка: он ничего не портит. Но на практике разница здесь обычно не слышна, потому что ошибка округления лежит на уровне около −144 dBFS, далеко ниже шума любой реальной записи. Если заказчик указывает, нужен ли дизеринг, следует выполнять его требование. Многократное добавление дизеринга не вредит катастрофически, но накапливает шум, поэтому его не включают на каждом промежуточном файле, а только на окончательном.

Уровни: полная шкала и истинный пик

Цифровая шкала имеет потолок — самое большое число, которое можно записать. Его называют 0 dBFS (децибелы относительно полной шкалы, *decibels full scale*). Все уровни отсчитываются вниз от этого потолка: −6 dBFS означает половину наибольшей амплитуды, −20 dBFS — одну десятую. В файле с целыми числами уровни отсчётов не могут быть выше 0 dBFS; внутри программы, где счёт идёт с плавающей точкой, уровень может быть и положительным, как описано выше. Для настройки трактов используют опорный тон: в американской практике его уровень обычно −20 dBFS, в европейской −18 dBFS. Это не «норма громкости фильма», а лишь отметка, по которой сверяют оборудование. О громкости как таковой речь пойдёт в главе 12.

Здесь есть тонкость, из-за которой случаются неприятные сюрпризы при сдаче. Индикатор пика в течение всего фильма ни разу не дошёл до 0 dBFS, а заказчик отклонил файл из-за превышения пика. Ошибки заказчика здесь нет. Обычный индикатор пика показывает наибольший *отсчёт*. Но настоящий сигнал, который восстановит ЦАП, проходит и между отсчётами, и его вершина может оказаться выше самого большого отсчёта. На рис. 1.3 все отсчёты находятся в пределах шкалы, а восстановленная волна поднимается на 3 дБ выше. Такая вершина называется *межотсчётным пиком* (*inter-sample peak*), а уровень с её учётом — *истинным пиком* (*true peak*). Сам файл от этого не испорчен, но ЦАП, кодек для передачи или преобразователь частоты могут на такой вершине перегрузиться. Поэтому требования к сдаче обычно задают предел именно истинного пика, а увидеть его может только специальный измеритель.

Конкретный предел зависит от заказчика и формата: для телевидения и стриминговых платформ часто указывают −1 или −2 dBTP, где TP означает истинный пик. Чем сильнее материал будет потом сжат кодеком с потерями, тем больший запас разумно оставить, поскольку при декодировании пики могут вырасти. Если требований нет, запас около 1 дБ истинного пика — распространённая осторожная практика. Измеритель истинного пика есть в большинстве современных индикаторов громкости, а лимитер с режимом контроля истинного пика (*true peak*, *ISP*) удерживает сигнал в нужных пределах.

Сжатие звука

Несжатый звук занимает много места: один канал 48 кГц, 24 бита — это 1152 кбит/с, а стерео около 17 МБ в минуту. Для передачи через интернет и в бытовых форматах звук сжимают.

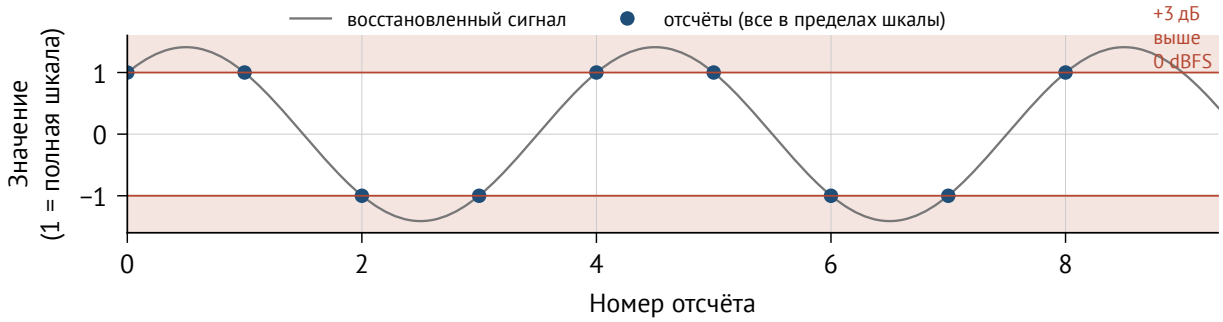

Рис. 1.3. Межотсчётный пик. Тон с частотой в четверть частоты дискретизации и неудачной фазой: отсчёты не превышают полной шкалы, а сигнал между ними выше на 3 дБ. Это худший случай для одиночного тона; у реальных сигналов превышение обычно меньше, но после лимитера и при сжатии с потерями встречается часто.

Сжатие бывает двух видов. *Сжатие без потерь* (*lossless*; FLAC, ALAC) похоже на архиватор: файл становится меньше примерно вдвое, а после распаковки получаются в точности те же отсчёты. *Сжатие с потерями* (*lossy*; MP3, AAC, Opus, а также кодеки кинотеатров и телевидения) работает иначе. Оно опирается на свойства слуха: тихий звук рядом по частоте с громким маскируется и не слышен (это явление подробно разобрано в первой книге, «Монтаже диалогов»), поэтому его можно записать грубо или не записывать вовсе. Так файл уменьшается в 10–20 раз. При достаточной скорости потока разница на слух может быть почти незаметной. На низкой скорости потока сжатие выдаёт себя «булькающими» призвуками на фоне, смазанными шипящими и потерей верхних частот.

Беда начинается при повторном сжатии. Типичная история: в монтаж положили временную музыку или эффект, скачанные в MP3. Несколько недель всё звучало прилично, а после кодирования для сдачи этот фрагмент стал «мыльным»: верха смазались, на тихих местах появились призвуки. Кодек сдачи здесь ни при чём: на артефакты первого сжатия наложились артефакты второго. Кодек «прячет» свои ошибки под маскировкой, рассчитанной на исходный звук, а после обработки, эквализации и сведения маскировка уже другая, и спрятанное становится слышным. На высокой скорости потока несколько поколений бывают почти неотличимы, на низкой ухудшение заметно быстро.

Отсюда правила для монтажа. Исходники хранятся и монтируются без потерь, в WAV или BWF (о метаданных BWF — в главе 4). Временные файлы со сжатием с потерями заменяют исходниками до окончательного сведения, и делать это лучше по списку, а не по памяти: после недель работы недостатки временного файла просто перестают замечать. Если же другого источника нет, например интервью есть только в виде ролика из интернета, файл переводят в WAV с параметрами проекта. Качество это не вернёт, но остановит дальнейшие потери: при монтаже и экспорте он больше не будет пересжиматься.

Историческая справка

Цифровой звук вошёл в обиход в начале 1980-х вместе с компакт-диском, записанным в формате 16 бит / 44,1 кГц. Необычная частота 44,1 кГц — наследство первых цифровых студийных записей: до появления специальных цифровых магнитофонов отсчёты записывали на видеомагнитофоны, размещая их в строках телевизионного кадра, и частота получилась из параметров телевизионной развёртки (например, 3 отсчёта в строке × 245 строк × 60 полей в секунду = 44 100). Во второй половине 80-х на съёмочной площадке появился цифровой кассетный магнитофон DAT, который быстро вытеснил четвертьдюймовую ленту и надолго стал стандартом полевой записи; запись сразу в файлы вытеснила его уже в 2000-е. Первые цифровые звуковые станции, такие как Synclavier и Audiofile, работали с 8 или 16 дорожками и маленьким встроенным диском, а профессиональные видеоформаты конца 80-х и 90-х несли четыре канала звука 16 или 20 бит при 48 кГц. Из этого времени в кино и на телевидении осталась частота 48 кГц.

От отсчётов к компьютеру

Отсчёты и их точность — это то, что хранится в файле. Как компьютер проигрывает, обрабатывает и сводит их в реальном времени и откуда берётся задержка, — тема следующей главы.