ГЛАВА 3

ВРЕМЯ ФИЛЬМА

[Уайатт, Эмиес]

Звук в компьютере измеряется отсчётами, а фильм — кадрами. Пока оба счёта согласованы, монтажёр об этом не думает: звуковой фрагмент стоит в нужном кадре и остаётся там до конца фильма. Трудности начинаются, когда программы и устройства по-разному понимают, сколько кадров в секунде, или когда их часы идут с чуть разной скоростью. Тогда звук либо сразу стоит не на месте, либо медленно уходит от изображения, и к концу длинной сцены рассинхрон становится заметен.

Зачем звукорежиссёру разбираться в частотах кадров, если частоту проекта можно просто взять из видеофайла? Затем, что ошибки здесь коварны. Рассинхрон из-за перепутанной частоты в начале сцены не виден вовсе и проявляется только через несколько минут. Тайм-код, который назвал монтажёр изображения, может не совпасть с тайм-кодом звуковой станции, хотя оба смотрят на один и тот же кадр. Фильм в телевизионной версии вдруг звучит выше, чем в кинотеатре. Причина каждый раз не в звуке, а в том, как устроен счёт времени, и, не понимая этого, легко «исправить» синхрон так, что станет только хуже. Восприятие рассинхрона зрителем подробно разобрано в главе 4 «Монтажа диалогов»; здесь речь пойдёт о том, как его не создать технически.

Частоты кадров

Кино исторически снимается с частотой 24 кадра в секунду. Европейское телевидение работает с 25 кадрами, американское и японское — с 30, а точнее с 29,97. Эти две телевизионные традиции называют по их системам цветного телевидения: PAL (*Phase Alternating Line*, Европа, в том числе Россия) и NTSC (*National Television System Committee*, США и Япония). К основным частотам добавились удвоенные (48, 50, 59,94, 60) для съёмки быстрого движения и замедления. В табл. 3.1 приведены те, что встречаются чаще всего.

За округлёнными числами 23,976 и 29,97 стоят точные дроби: $24 \cdot 1000/1001$ и $30 \cdot 1000/1001$. Они на 0,1 % медленнее «круглых» частот. Происхождение этой разницы историческое, но последствия вполне современные: многие камеры по умолчанию снимают именно 23,976, и монтаж изображения идёт на этой частоте, а кинотеатр показывает 24.

Разница в 0,1 % кажется ничтожной, но она накапливается. Представим: монтажёр изображения прислал сцену на 23,976 к/с, а проект в звуковой станции по ошибке настроен на 24. Звук, поставленный по тайм-коду начала, в первых кадрах совпадает идеально. Через минуту он уходит на 60 мс — это уже больше кадра, и на крупном плане губы перестают совпадать. Через десять минут расхождение больше полусекунды,

Таблица 3.1. Распространённые частоты кадров. Значения рассчитаны; «23,976» и «29,97» — округлённые записи точных дробей 24000/1001 и 30000/1001

Частота, к/с	Длительность кадра, мс	Отсчётов на кадр при 48 кГц	Где встречается
24	41,667	2000	Кинопроекция, цифровой кинопакет (DCP)
23,976	41,708	2002	Съёмка «под кино» для видео и стриминга в странах NTSC; монтаж многих фильмов
25	40,000	1920	Телевидение и видео в странах PAL, в том числе в России
29,97	33,367	1601,6	Телевидение стран NTSC
30	33,333	1600	Видео без привязки к телевизионному стандарту
50 и 59,94	20,000 и 16,683	960 и 800,8	Спорт, репортаж, съёмка движения

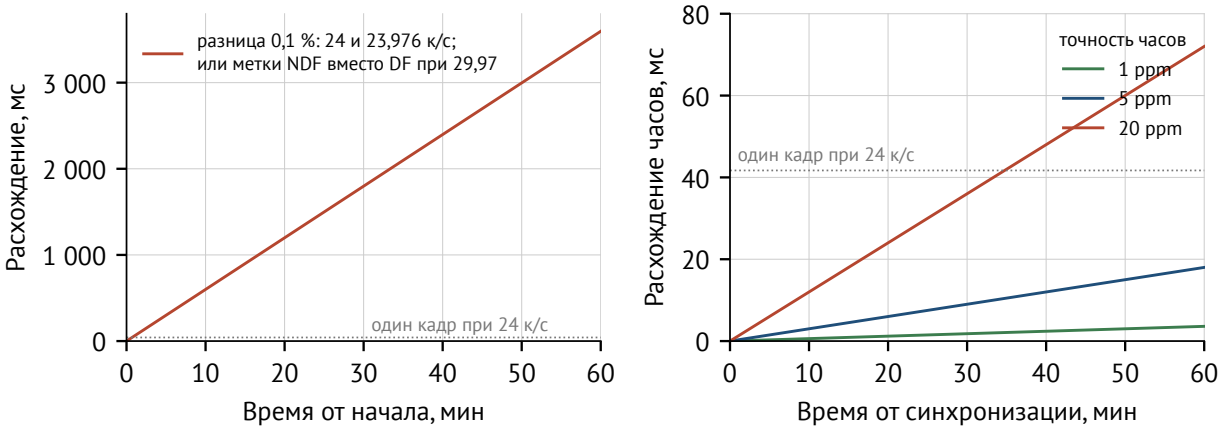

Рис. 3.1. Слева: накопление рассинхрона при разнице частот 0,1 %, например если материал 23,976 к/с играет как 24 к/с. Справа: расхождение двух независимых часов с разной точностью после одной синхронизации. Пунктир — длительность одного кадра при 24 к/с. Значения рассчитаны; точность реальных устройств указывает их производитель.

через час — 3,6 с (рис. 3.1, слева). Главный признак такой ошибки — рассинхрон, который *растёт* равномерно, примерно на 60 мс в минуту. Двигать звук здесь бесполезно: подвинув его в одном месте, вы испортите другое. Нужно сравнить частоту видеофайла, проекта и звуковых файлов и привести всё к одной.

Третий столбец таблицы показывает ещё одну деталь. При 24, 25 или 23,976 к/с в одном кадре целое число отсчётов, а при 29,97 — нет: 1601,6. Граница кадра попадает между отсчётами, и картина повторяется только через пять кадров (8008 отсчётов). Поэтому при монтаже на 29,97 некоторые программы округляют положение границы кадра до отсчёта, и склейка «точно по кадру» может оказаться на долю отсчёта в стороне. На слух это не влияет, но это ещё одна причина задавать частоту проекта один раз и не менять её.

Историческая справка

Частота 29,97 появилась в начале 1950-х, когда в США к чёрно-белому телевидению с 30 кадрами в секунду добавили цветной сигнал. Чтобы цветовая составляющая не мешала звуковой, частоту кадров немного понизили, на 0,1 %. Кино, перенесённое на такое видео, тоже пришлось замедлить на 0,1 %, отсюда 23,976. В Европе телевидение было основано на частоте электросети 50 Гц и работало с 25 кадрами в секунду. Фильм, снятый на 24 кадра, при показе по европейскому телевидению просто ускоряли до 25 кадров: он шёл на 4 % быстрее, а звук становился выше. При переносе плёнки на видео в системе NTSC кадры повторяли в сложном порядке (так называемая схема 3:2), чтобы из 24 кадров получить 60 полей.

Тайм-код: адрес, а не время

Тайм-код (*timecode*) — это номер кадра, записанный в виде «часы:минуты:секунды:кадры», например 01:02:10:15. Его стандарты описаны в документах американского Общества инженеров кино и телевидения (SMPTE, *Society of Motion Picture and Television Engineers*) и Европейского вещательного союза (EBU, *European Broadcasting Union*). Важно понимать, что тайм-код — это *адрес* кадра, а не измерение времени. Он говорит, какой это кадр по счёту, но ничего не говорит о том, сколько длится кадр. Чтобы перевести тайм-код в секунды, нужно знать частоту. Одна и та же метка 00:10:00:00 при 24 и при 23,976 к/с соответствует разному реальному времени: во втором случае на 0,6 с позже.

В фильме тайм-код выполняет две задачи. Первая — найти материал: звукооператор и камера пишут одинаковый тайм-код, и по нему звук каждого дубля находится автоматически. Вторая — поставить звук на место: зная тайм-код начала файла, станция ставит его точно в нужный кадр.

Пропуск кадров

При 29,97 к/с возникает неудобство. Если считать кадры по 30 в секунду, метки отстают от настоящих часов: через час на часах 01:00:00:00, а тайм-код показывает 00:59:56:12, на 3,6 с меньше. Для телевидения, где программа должна длиться ровно столько, сколько указано в сетке вещания, это неприемлемо. Поэтому есть тайм-код *с пропуском кадров* (*drop-frame*, DF): в начале каждой минуты пропускаются две метки (номера кадров 00 и 01), кроме каждой десятой

минуты. За час так пропускается 108 меток, и тайм-код почти совпадает с часами: остаётся расхождение около 0,1 кадра в час, примерно 86 мс за сутки.

Пропускаются именно *метки*, а не кадры. Изображение и звук не теряют ни одного кадра, меняется только то, как их пронумеровали. Поэтому смешение DF и обычного тайм-кода (*non-drop-frame*, NDF) не портит материал, но сбивает адреса. В работе это выглядит так: монтажёр изображения называет тайм-код реплики, которую надо поправить, а в звуковой станции по этому адресу стоит другой кадр, и чем ближе к концу ролика, тем больше разница — до тех же 3,6 с за час. Это не сбой синхрона: один и тот же материал прочитан в одной программе с пропуском кадров, а в другой без него. Подсказку даёт разделитель перед кадрами: тайм-код DF обычно пишут с точкой с запятой или точкой (01:00:00;00), а NDF — с двоеточием. Режим согласовывают в настройках проекта (в Pro Tools — в окне *Session Setup*, в Reaper — в настройках проекта, раздел *Video/Frame rate*) и указывают в документах обмена.

Для 23,976 к/с стандартного пропуска кадров нет: такой тайм-код считает 24 кадра в секунду и отстаёт от часов на те же 0,1 %. Для кино это не страшно: важна не реальная длительность, а совпадение звука и изображения, которые считаются одинаково.

> **Для любопытных**
>
> Почему пропуск 108 меток в час почти точно компенсирует 0,1 %? За час при точной частоте 30000/1001 проходит $3600 \cdot 30000/1001 \approx 107\,892,1$ кадра, а без пропуска тайм-код насчитал бы $3600 \cdot 30 = 108\,000$ меток. Разница 107,9 кадра. Пропуск по две метки в 54 минутах из 60 даёт 108. Остаток $108 - 107,9 \approx 0,1$ кадра в час и есть то небольшое расхождение, которое накапливается за сутки.

Где живёт тайм-код сегодня

Раньше тайм-код записывался на ленту отдельной дорожкой в виде звукового сигнала (продольный тайм-код, LTC, *linear timecode*) или в служебные строки изображения (VITC, *vertical interval timecode*). Сегодня он чаще всего хранится в метаданных файлов. Звуковой файл формата BWF (*Broadcast Wave Format*, WAV с дополнительным описанием) содержит отметку времени начала — число отсчётов от полуночи — и сведения о частоте кадров (подробнее о метаданных — в главе 4). Видеофайл камеры хранит тайм-код первого кадра. Станция, прочитав эти отметки, ставит файл на место сама.

Продольный тайм-код как звуковой сигнал при этом не исчез. Его до сих пор используют, чтобы передать время между устройствами, например записать на свободную звуковую дорожку камеры, у которой нет входа тайм-кода. Его характерный скрежет узнаёт любой, кто хоть раз случайно вывел такую дорожку в мониторы.

Синхронизация на площадке

Чтобы звук и изображение совпадали, камера и рекордер должны иметь одинаковое время. Для этого их *синхронизируют* (*jam sync*): в начале дня от одного генератора тайм-кода передают время во все устройства. После этого каждое устройство считает время само, по своим внутренним часам. И здесь появляется проблема: никакие часы не идут абсолютно точно. Их точность выражают в миллионных долях, ppm (*parts per million*). Расхождение в 1 ppm — это 3,6 мс за час, в 20 ppm — 72 мс за час, то есть почти два кадра (рис. 3.1, справа).

Обычные внутренние часы камер и рекордеров бывают не точнее нескольких или десятков ppm. Специальные генераторы тайм-кода, которые крепят к камере и рекордеру, обычно гораздо точнее; конкретное значение указывает их производитель. Поэтому на долгих съёмках синхронизацию повторяют несколько раз в день, а на длинных дублях используют внешние генераторы на всех устройствах.

Если синхронизация всё-таки ушла, это видно по характерному признаку: расхождение растёт внутри дубля, но гораздо медленнее, чем при перепутанной частоте. В начале пятиминутного дубля хлопушка совпадает идеально, а к концу губы и звук разошлись на кадр. Камера и рекордер честно записали каждый своё время, просто их часы шли с чуть разной скоростью. Исправляют это двумя способами. Можно измерить сдвиг в начале и в конце дубля и чуть изменить скорость звука, чтобы обе точки совпали: в Reaper это скорость воспроизведения фрагмента (*playback rate*), в Pro Tools — обработка с изменением длительности. Можно разрезать дубль в паузах и совместить куски по отдельности. Второй способ проще и надёжнее, если в дубле есть паузы.

Хлопушка (*clapper*) остаётся надёжным запасным вариантом. Удар половинок даёт резкий звук и хорошо видимый кадр, и по ним звук совмещают с изображением вручную. Кроме того, современные программы монтажа изображения и звука умеют совмещать материал по форме волны: они сравнивают звук камеры (которая почти всегда пишет черновой звук) со звуком рекордера и находят сдвиг. Этот способ хорошо работает при чистом черновом звуке и хуже — в шуме или при большом расстоянии камеры от источника.

Изменение скорости

Иногда материал приходится переводить с одной частоты кадров на другую. Два случая встречаются чаще всего.

Ускорение 24 → 25. Чтобы показать кинофильм на телевидении с частотой 25 к/с, его часто просто ускоряют: каждый кадр показывают за 1/25 секунды. Фильм становится на 4 % короче, а звук при простом ускорении — на 71 цент выше, то есть почти на три четверти полутона (цент — сотая доля полутона). Отсюда знакомая многим история: в телевизионной версии голоса звучат чуть выше и бодрее, а музыка — не в той тональности, что в кинотеатре. Это не ошибка конкретного канала, а прямое следствие ускорения без сохранения высоты. Сегодня звук при таком переводе обычно обрабатывают с сохранением высоты, но это решение нужно принимать явно: на голосах и музыке разница в три четверти полутона хорошо слышна, а обработка с сохранением высоты может давать свои призвуки на сложных звуках, поэтому такие места проверяют на слух.

Изменение на 0,1 %. Переход между 24 и 23,976 к/с меняет скорость всего на 0,1 %, а высоту — на 1,7 цента. Высоту так никто не услышит, но длительность меняется, и без пересчёта звук уйдёт на 60 мс за минуту. Поэтому при таком переходе звук пересчитывают: либо ускоряя или замедляя его вместе с изображением, либо, если изображение пересчитано без изменения скорости, оставляя как есть. Чтобы не ошибиться, нужно точно знать, что сделано с изображением, и спрашивать об этом монтажёра изображения, а не угадывать.

Есть и ловушка со стороны площадки. Некоторые звукооператоры для съёмки на 24 к/с, которую будут монтировать на 23,976, записывают звук с частотой 48,048 кГц. При воспроизведении на 48 кГц он замедляется ровно на 0,1 % вместе с изображением. Приём удобный, но такой файл легко принять за обычный 48 кГц, и тогда синхрон начинает уходить на те же 60 мс в минуту. Поэтому, если звук с площадки «уплывает» равномерно, первым делом смотрят частоту в свойствах файлов и в отчёте звукооператора.

> **Для любопытных**
>
> Изменение скорости в r раз без сохранения высоты сдвигает частоты в r раз, что в центах составляет $c = 1200 \log_2 r$. Для $r = 25/24$ это 71 цент, для $r = 1001/1000$ — 1,7 цента. Длительность меняется в $1/r$ раз: на 4 % и 0,1 % соответственно.

Изображение в звуковой станции

Звукорежиссёр работает с изображением: станция проигрывает видеофайл синхронно со звуком. Здесь есть две ловушки.

Первая — частота. Видеофайл, частота проекта и частота тайм-кода должны совпадать. Если видео 23,976 к/с, а проект настроен на 24, станция может проигрывать видео правильно, но показывать тайм-код и ставить звук по другой частоте. Проверить это просто. При экспорте для звука монтажёр изображения обычно выводит тайм-код прямо в кадр. Этот вшитый тайм-код должен совпадать с тайм-кодом станции на протяжении всего файла, а не только в начале. Если разница растёт — неверна частота проекта; если она постоянна — неверен тайм-код начала видеофайла. В обоих случаях исправляют настройки, а не двигают звук.

Вторая — задержка монитора. Телевизоры, проекторы и многие мониторы обрабатывают изображение и показывают его с опозданием, иногда на несколько десятков миллисекунд и больше. Звук из громкоговорителей при этом не задерживается. В результате звук кажется опережающим изображение, хотя в файле всё синхронно. Признак такой ситуации — звук опережает губы на всём материале одинаково, и только на одном экране: на другом мониторе рассинхрон исчезает. Если не знать об этом, легко «исправить» синхрон, сдвинув звук, и испортить его для всех остальных зрителей. Правильный путь — измерить задержку специальным тестовым роликом со вспышкой и щелчком и задать поправку в станции (в Pro Tools это параметр *Video Sync Offset*, похожие настройки есть и

в других станциях) или включить на телевизоре режим с минимальной обработкой, часто он называется игровым. В профессиональных студиях задержку тракта изображения проверяют при настройке зала.

Сколько рассинхрона допустимо? Ответ зависит от того, что видно в кадре: губы на крупном плане требуют большей точности, чем шаги на общем. Отраслевые рекомендации для телевидения, например EBU R37, указывают допуск порядка нескольких десятков миллисекунд, причём опережение звука зритель переносит хуже, чем отставание: в жизни звук никогда не приходит раньше изображения. Для монтажёра практическое правило проще: ошибка в один кадр обычно уже видна на крупном плане речи, поэтому всё, что сдвигает синхрон на кадр и больше, нужно находить и устранять.

От времени к файлам

Тайм-код и частота кадров позволяют поставить звук на место. Но чтобы он туда попал, материал должен дойти от площадки и монтажа изображения до звуковой станции вместе со всеми сведениями: какой это дубль, с какого микрофона, к какой сцене он относится. Как это устроено, рассматривается в следующей главе.