

Тим Кинтцель

Руководство программиста по работе со звуком



WAVE

MIDI

MOD

FFT

MPEG

ADPCM

μ -Law

MP3


Addison
Wesley


DMK
ИЗДАТЕЛЬСТВО

ДЛЯ ПРОГРАММИСТОВ

ББК 32.973.26-018.2

К41

Кинтцель Т.

К41 Руководство программиста по работе со звуком = A Programmer's Guide to Sound: Пер. с англ. – М.: ДМК Пресс. – 432 с., ил. (Серия «Для программистов»).

ISBN 5-94074-014-6

Издание содержит подробную информацию о хранении, обработке и компрессии звуковых файлов. Разработчики, которые хотят использовать в своих приложениях возможность обработки звука, найдут в этой книге ответы на все интересующие их вопросы: применение низкоуровневых средств работы со звуком для Win32, Mac OS и UNIX; декомпрессия данных в форматах MPEG, IMA ADPCM; чтение файлов WAVE, VOC, AIFF и AU; воспроизведение файлов MIDI и MOD.

В этой книге вы также найдете обзор смежных тем, в частности синтеза звучания музыкальных инструментов, восприятия звуков человеком, цифровой фильтрации и преобразований Фурье.

Для каждого раздела приведены полные исходные тексты программ, иллюстрирующие все обсуждаемые принципы.

Книга предназначена для звукоинженеров, желающих пополнить свои знания о программировании звука, а также для программистов, которым необходима информация о работе с аудиоматериалами. Некоторые разделы книги могут быть полезны специалистам, разрабатывающим программные средства обработки сигналов для геофизики, телеметрии, связи, систем автоматического управления и сбора данных.

ББК 32.973.26-018.2

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the publisher. Printed in the United States of America. Published simultaneously in Canada.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но, поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 0-201-41972-6 (англ.)

ISBN 5-94074-014-6 (рус.)

Copyright © by Tim Kientzle

© Перевод на русский язык, оформление ДМК Пресс



Содержание

Введение	13
Часть I. Основы	15
Глава 1. От выдолбленной колоды к киберпространству	16
Что такое звук?	17
Естественные звуки	18
Электронные звуки	18
Звуки, которые мы слышим	20
Цифровой звук	20
Дополнительные источники информации	21
Глава 2. Восприятие звука человеком	23
Частота и высота тона	23
Высота тона в музыке	25
Высота тона и частота звука	29
Шум	30
Громкость, амплитуда и мощность	31
Общее качество звука	33
Глава 3. Хранение звука в цифровом виде	35
Дискретные форматы звука	35
Амплитудно-импульсная модуляция (АИМ)	35
Широтно-импульсная модуляция (ШИМ)	36
Импульсно-кодовая модуляция (ИКМ)	37
Побочные эффекты дискретизации	38
Наложение спектров	39
Наложение спектров в выходном сигнале	43
Передискретизация	43
Погрешность квантования	45
Сглаживание сигнала	46
Ограничение сигнала	47
Выборки с плавающей точкой	48

Глава 4. Структура для звука в C++	49
Класс AudioAbstract	51
Стереозвук	54
Согласование	54
Чтение и запись целых чисел	59
Класс SineWave	60

Часть II. Особенности работы со звуком

в различных системах

Глава 5. Объекты-проигрыватели	64
Проигрыватели: основы	64
Реализация очереди	67
Открытие звукового файла	70
Глава 6. Воспроизведение звука в Windows	73
Воспроизведение	74
Вызов подсистемы	76
Выбор устройства	78
Пример приложения Windows	81
Глава 7. Воспроизведение звука в Mac OS	84
Вызов подсистемы	85
Двойная буферизация	86
Пример программы для Mac OS	87
Воспроизведение файла в Mac OS	90
Глава 8. Система UNIX и звук в сети	92
Серверы и потоки	93
Вызов подсистем и сообщения	94
Конфигурирование потока и подключение к серверу	97
Пример приложения для UNIX	100

Часть III. Компрессия

Глава 9. Компрессия звука	102
Сжатие информации без потерь	102
Нелинейная ИКМ	103
Дифференциальная ИКМ	104

Адаптивная ДИКМ	105
Методы компрессии с прогнозированием	105
Кодирование в частотных поддиапазонах	106
Компрессия речи человека	106
Использование опорных точек	107
Прогрессивная компрессия	108
Глава 10. Классы декомпрессоров	110
8-битная ИКМ	111
Знаковая 8-битная ИКМ	112
Беззнаковая 8-битная ИКМ	112
16-битная ИКМ	113
16-битная ИКМ формата MSB	113
16-битная ИКМ формата LSB	114
Глава 11. Нелинейные звуковые форматы	116
Логарифмическая компрессия	116
Компрессия m-Law	117
Компрессия A-Law	121
Свойства логарифмического кодирования	122
Глава 12. Дифференциальная ИКМ	125
Два вида ДИКМ-кодирования	125
Глава 13. Компрессия IMA ADPCM	129
Общее описание работы	130
Декомпрессия	131
Компрессия	132
Практическая реализация IMA ADPCM	132
Вариант Microsoft IMA ADPCM	133
Вариант Apple IMA ADPCM	138
Сравнение модулей Microsoft и Apple	141
Несколько слов об IMA ADPCM	141
Как я разбирался с форматами ADPCM	142
Соавторы	143
Глава 14. Аудио MPEG	144
Обзор стандартов MPEG	144
Стандарт MPEG-1	144
Стандарт MPEG-2	145

Стандарт MPEG-3	146
Стандарт MPEG-4	146
Аудио MPEG	147
Общая структура	147
Обработка потока байтов	151
Заголовок фрейма MPEG	153
Слоты и фреймы	157
Синтез в поддиапазонах	159
Матрицирование	161
Коэффициенты взвешивания	167
MPEG-кодирование стереозвука	170
Уровень Layer 1	170
Хранение распределения в Layer 1	171
Масштабные коэффициенты в Layer 1	172
Хранение отсчетов в Layer 1	172
Передискретизация и масштабирование в Layer 1	174
Уровень Layer 2	175
Хранение распределения в Layer 2	177
Масштабные коэффициенты в Layer 2	181
Чтение выборок в Layer 2	182
Передискретизация в Layer 2	184
Уровень Layer 3	185
Программа чтения файлов MPEG	186
Источники дополнительной информации	187
Часть IV. Основные форматы файлов	189
Глава 15. Формат файла AU	190
Идентификация файлов AU	192
Чтение файлов AU	192
Запись файлов AU	195
Простой фильтр AU	196
Глава 16. Формат файла VOC	197
Идентификация файлов VOC	197
Заголовок файла формата VOC	197
Блоки данных	198
Блок признака окончания (тип 0)	199
Блок звуковых данных (тип 1)	199

Блок продолжения данных (тип 2)	201
Блок паузы (тип 3)	202
Маркерный блок (тип 4)	202
Текстовый блок (тип 5)	203
Циклы повторения (типы 6 и 7)	203
Блок расширения (тип 8)	205
Блок расширения (тип 9)	206
Другие блоки	207
Чтение файлов VOC	208
Глава 17. Формат файла WAVE	212
Идентификация WAVE-файлов	212
О файлах формата RIFF и IFF	212
Формат WAVE. Обзор	213
Класс WaveRead	214
Чтение файлов формата WAVE	215
Контейнеры	216
Контейнер RIFF WAVE	219
Блок fmt	220
Создание объекта декомпрессора	222
Блок data	224
Текстовые блоки	225
Глава 18. Форматы файлов AIFF и AIFF-C	226
Идентификация файлов AIFF	227
Класс AiffRead	228
Чтение файлов AIFF	229
Контейнеры	229
Контейнер FORM AIFF	231
Блок FVER	232
Блок COMM	233
Блок SSND	236
Глава 19. Формат файла IFF/8SVX	238
Идентификация файлов IFF/8SVX	238
Обзор формата IFF/8SVX	239
Чтение файлов формата IFF/8SVX	239
Общие замечания	240
Контейнеры	241
Контейнер FORM 8SVX	242

Блок VHDR	243
Блок BODY	246
Ошибки IFF/8SVX	247
Текстовые блоки	248

Часть V. Форматы музыкальных файлов 249

Глава 20. Программирование музыки 250

Музыкальные инструменты 251

Ноты 252

Глава 21. Синтез музыкальных инструментов 254

Семплы 254

Генератор синусоидального сигнала 260

Управление амплитудой 261

Управление огибающей в семплах 262

Другие методы управления 263

Частотная модуляция и синтез 264

Колесящая струна 264

Реализация алгоритма струны 266

Тестирование нот 273

Глава 22. Стандарт MIDI 277

Файлы формата Standard MIDI 278

Идентификация файлов MIDI 278

Блок заголовка MIDI 279

Дорожки MIDI 280

Целые значения переменной длины 281

Относительное время 281

Чтение дорожек MIDI 281

События MIDI 283

Соглашения по нумерации 284

Статус выполнения 284

Управление событиями MIDI 285

Чтение событий MIDI 286

Эксклюзивные системные сообщения 287

Мета-события 288

Синхронизация MIDI 291

Стандарт General MIDI 294

Загружаемые семплы 297

Проигрыватель MIDI	298
Постобработка потока событий MIDI	301
Базовый и расширенный MIDI	304
Воспроизведение потока событий MIDI	306
Каналы MIDI	308
Механизмы aftertouch и Pitch Wheel	311
Контроллеры и режимы	312
Реализация	317
Настройка нот в MIDI	318
Музыкальные инструменты MIDI	319
Предостережения	323
Канальный протокол MIDI	323
Команды реального времени	323
Другие форматы файлов MIDI	324
Стандарты MIDI	325
Глава 23. Формат файла MOD	326
Идентификация файлов MOD	327
Аппаратные средства Amiga	327
Формат MOD: обзор	329
Инструменты	329
Синхронизация	329
Ноты	329
Такты	330
Трафареты	331
Плей-лист	331
Детали формата	331
Инструменты	332
Ноты	337
Хранение музыкального произведения	340
Общая структура файла	343
Воспроизведение файлов MOD	346
Метод PlayBeat	349
Воспроизведение с эффектами	352
Часть VI. Обработка звука	371
Глава 24. Преобразования Фурье	372
Основы преобразования Фурье	372
Синусоидальные сигналы	373

Как измерить одну частоту	375
Учет фазы	378
Реализация ДПФ	379
Масштабирование	379
Как оценить весь спектр	380
«Короткие» БПФ	380
Разложение «длинных» БПФ	382
Двухточечное БПФ	382
Четырехточечное БПФ	383
Формальный вывод БПФ	383
Программирование БПФ	384
Скорость	388
Эксперименты с БПФ	389
Использование БПФ	390
Выбор окна	393
Замечания	393
Глава 25. Фильтрация	395
Наложение синусоид	395
КИХ-фильтр	396
Проектирование простых КИХ-фильтров	396
Реализация КИХ-фильтров	399
Фазовый сдвиг	401
БИХ-фильтры	401
Синтез с использованием фильтров	402
Часть VII. Приложения	403
Приложение А. Где можно взять исходные коды	404
Приложение В. Краткое введение в С++	408
Приложение С. Стиль программирования	420
Приложение D. Используемые системы	422
Приложение E. Повторное использование моего исходного программного кода	424
Алфавитный указатель	426



Глава 1. От выдолбленной колоды к киберпространству

Звук – неотъемлемая часть того, что определяет наше восприятие окружающего мира. На протяжении всей нашей жизни тысячи звуков сплетаются в причудливый ковер, от которого мы зависим в большей степени, чем можем себе представить. Я слышал, что младенцы, еще находясь в лоне матери, успокаиваются, слыша негромкий стук ее сердца. Когда наш слух ослабевает из-за простуды или гриппа, оттого что звуки воспринимаются не как обычно, мир удаляется от нас и становится не таким «объемным». Тихая, поросшая лесом долина заполнена не только листвой деревьев, но также и сонмом слабых звуков, издаваемых насекомыми, птицами и шепчущими порывами ветра.

У компьютера, как и у всех окружающих нас предметов, есть такой звуковой портрет. Пощелкивание клавиш клавиатуры и кнопок мыши, тихий шелест устройств чтения компакт-дисков, жужжание вентиляторов источников питания, характерное посвистывание устанавливающих соединение модемов – это всего лишь часть звуков, из которых складывается общая звуковая картина работающего компьютера.

В наши дни, когда высококачественный синтезированный звук окончательно утвердился в качестве стандарта для настольных компьютеров, повседневный звуковой пейзаж постепенно все больше и больше проникает в виртуальную реальность. Сейчас он используется в основном в играх и разнообразных системах вещания по сети, однако для него будут найдены и другие области применения. Например, система «говорящего» справочника позволит пользователю ПК сконцентрировать внимание на той или иной неполадке, вместо того чтобы отыскивать причину сбоя в справочнике. Некоторые программисты тоже пришли к выводу, что звуковое сопровождение отдельных действий программ упрощает процесс поиска ошибок.

Тем не менее хранение и вывод звуковой информации может оказаться более сложной задачей, чем кажется на первый взгляд, так как при ее решении необходимо учитывать ряд взаимосвязанных характеристик:

- *объем памяти.* Хранение высококачественной звукозаписи требует большого объема памяти либо применения сложных методов компрессии;
- *быстродействие (полоса частот).* При использовании несжатого звука необходимо обеспечить непрерывную и устойчивую его передачу, что особенно трудно в сетевых приложениях. При использовании сжатого звука вы должны быть уверены, что процессор сможет произвести декомпрессию достаточно быстро. На практике не исключено, что компрессия окажется слишком сильной (или наоборот);

- *время ожидания (задержка)*. При воспроизведении звука неизбежны временные задержки. Они возникают в связи с тем, что определенное время затрачивается на считывание данных с диска, настройку аппаратуры, обеспечивающей работу со звуком, установку связи в сети и запуск процесса обработки звука. Во многих приложениях недопустима задержка даже на сотую долю секунды.

Перечисленные характеристики могут встречаться в самых разных комбинациях. Предположим, вы разрабатываете программу для проведения двусторонней конференции в сети. Модемы обладают ограниченной пропускной способностью, поэтому для обеспечения высокого качества передаваемого звука придется использовать эффективные средства компрессии. Однако многие алгоритмы сжатия звука устроены так, что за один раз обрабатывается блок данных значительного объема. При применении подобного алгоритма перед сжатием и передачей данных вам приходится ждать, пока наберется блок звуковой информации достаточного объема. Это может привести к появлению длительной задержки. Требуется особая тщательность для соблюдения баланса между ограничениями на ширину полосы пропускания и длительностью задержки.

Что такое звук?

Формально звук представляет собой волны изменения давления, которые двигаются в сжимающейся среде. Строго говоря, характеристики звука зависят от свойств материала, где он распространяется, но общие принципы одни и те же¹. Молекулы (воздуха, воды, металла) стремятся занять положение, равноудаленное от всех своих соседей. Поэтому как только молекулы в какой-нибудь области окажутся ближе друг к другу, чем к своим соседям, они начнут отталкиваться друг от друга, стремясь разойтись на прежнее расстояние, сжимая соседние частицы.

На рис. 1.1 показано, как это происходит. Более темные участки рисунка соответствуют областям, в которых молекулы прижаты друг к другу теснее (это области повышенного давления). На более светлых участках показаны области, где молекулы относительно разрежены. После того как в области сжатия они разойдутся на прежнее расстояние, произойдет сжатие соседней области, молекулы которой также будут стремиться разойтись, что приведет к сжатию следующей, и т.д.

Происходящее можно рассматривать с нескольких позиций. Если проследить за отдельной молекулой, то обнаружится, что она колеблется туда-сюда, не перемещаясь на большое расстояние. Однако если вы отодвинетесь и рассмотрите область повышенного давления как единый объект, то увидите, что рис. 1.1 действительно схематично изображает движение *волны давления*.

¹ Курс физики в обычном колледже состоит из нескольких разделов, каждый из которых начинается словами: «Все, о чем мы до сих пор говорили при изучении этого курса, не соответствует действительности. Давайте разберемся с тем, что происходит на самом деле». К акустике это относится в той же степени, что и к теории относительности. В данном разделе мы сознательно используем упрощенный подход.

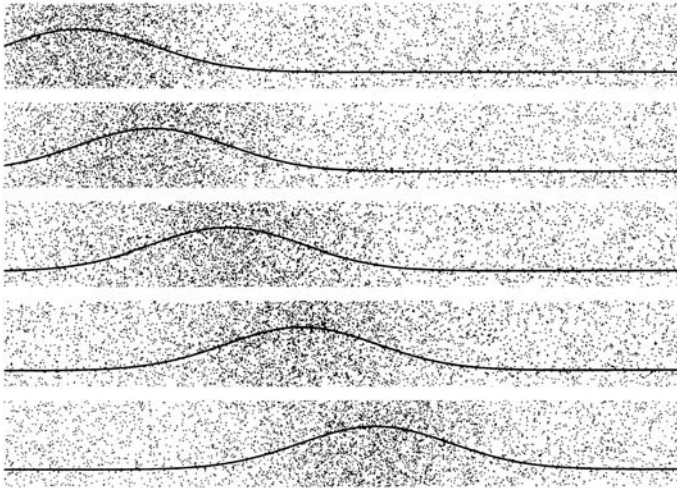


Рис. 1.1. Движущаяся волна давления

Естественные звуки

Базовые элементы, составляющие звук, – это вибрирующие объекты и волны изменения давления. Если ударить в барабан, его мембрана будет вибрировать. Поднимаясь, она сжимает воздух над собой, и образуемая в результате волна давления движется по воздуху до тех пор, пока не надавит на барабанную перепонку слушателя.

Подобным образом ведут себя и другие источники звука. Музыкант дует во флейту и заставляет вибрировать находящийся внутри нее столб воздуха. Открывая и закрывая отверстия, расположенные на инструменте, флейтист может управлять этой вибрацией. Если провести мокрым пальцем по ободку хрустального стакана, он начнет вибрировать.

Иногда вибрация имеет сложный характер. Голос человека формируется в результате взаимодействия вибраций голосовых связок, воздуха в легких, горле, ротовой полости и синусовых пазух, а также в жидкой среде его тела. Источником звука в акустической гитаре или скрипке являются колебания струны, которые создают вибрацию воздуха внутри инструмента. Губы тромбониста заставляют вибрировать воздух внутри тромбона. Практически все звуки, издаваемые этими инструментами, зависят от формы и характеристик воздушной камеры. Звуки, которые вы слышите ежедневно, почти всегда являются результатом сложного взаимодействия различных вибрирующих элементов.

Электронные звуки

Все слышимые нами звуки – это, в некотором смысле, последовательности волн давления. Если мы возьмем очень чувствительный измеритель давления и подсоединим его к перьевому графопостроителю, то получим линию,

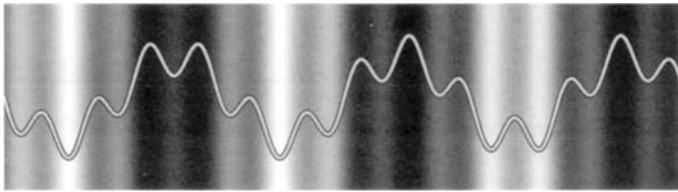


Рис. 1.2. Преобразование звука из колебаний воздуха в колебания электрического сигнала

поднимающуюся вверх при повышении давления и опускающуюся вниз при его уменьшении. Эта связь наглядно представлена на рис. 1.2: волнистая линия поднимается вверх при высоком давлении (зоны высокого давления показаны темным фоном) и опускается, когда давление падает.

Построить такой измеритель давления очень просто. Для начала возьмем кусок бумаги или тонкой пленки пластика. Когда давление воздуха начнет повышаться, воздух станет давить на бумагу и она сместится назад. Когда давление упадет, бумагу потянет вперед. Теперь приклеим к этому листочку небольшой виток проволоки и поместим поблизости магнит. По мере возрастания или падения давления воздуха лист бумаги будет перемещаться вперед или назад, а вместе с ним станет двигаться и виток проволоки. При перемещении провода в магнитном поле генерируется слабый переменный электрический ток. Изменения этого сигнала в точности соответствуют колебаниям давления воздуха, как показано на рис. 1.3.

Характерно, что подобный микрофонный эффект может быть обратимым: колебания электрического сигнала, подаваемого на проволочную петлю, заставят колебаться бумагу. А эти колебания, в свою очередь, создадут волны давления в прилегающем слое воздуха. Таким образом, принципы работы микрофона и динамика одни и те же.

Возможность преобразования звука из колебаний воздуха в переменные электрические сигналы лежит в основе таких привычных нам вещей, как радио, телефон и магнитофон. Это явление, кроме того, предоставляет широкое поле деятельности тем, кому нравится экспериментировать со звуком. Есть множество методов обработки звука, которые предпочтительнее реализовать с помощью электронных устройств, чем акустических приспособлений. Есть и такие, которые, наоборот, проще реализуются с помощью акустических, а не электронных средств. Например,

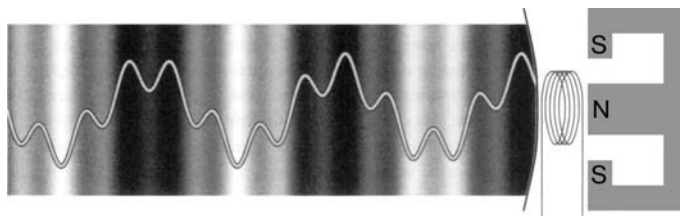


Рис. 1.3. Принцип действия микрофона

джазовые трубачи для изменения звучания своих инструментов использовали различные материалы, вплоть до мокрых тряпок, и казу (духовой музыкальный инструмент). Их влияние на звук трудно имитировать при помощи электроники. Напротив, с помощью акустических приспособлений практически невозможно добиться многих звуковых эффектов, которые легко достигаются при применении электронных цепей (в качестве примера упомянем столь любимые электрогитаристами эффекты «хорус» и «дисторшн»).

Звуки, которые мы слышим

Прежде чем перейти к следующему разделу и приступить к обсуждению цифрового представления звуков, мне бы хотелось познакомить вас с очень важным аспектом. В конечном счете, звуки, которые мы производим, нужны для того, чтобы их слышали. В связи с этим необходимо понимать, что и как слышит человек. Например, хотя наши уши представляют собой чрезвычайно чувствительные к изменению давления органы, очень немногие люди ощущают изменения давления, связанные с переменной погоды. Также человек не воспринимает звуки, используемые летучими мышами для ориентации.

То обстоятельство, что люди могут слышать звуки только в определенном диапазоне частот, приводит к интересным последствиям. Часто вы можете упростить звуковые данные или механизм компрессии, не оказав тем самым значительного воздействия на конечный результат. Чаще всего этим пользуются при наличии ограничений на полосу пропускания или объем памяти.

Посмотрим, например, что получится, если вы попытаете применить очень высококачественный алгоритм компрессии на компьютере средней мощности. Если используемое программное обеспечение недостаточно быстро обрабатывает аудиоданные, воспроизведение звука может прерваться и продолжиться только после прохождения данных через систему. В такой ситуации вы добьетесь большего впечатления, применяя менее точные методы. Хотя получаемые в результате преобразований аудиоданные в техническом смысле менее качественны, можно поддерживать устойчивую передачу, избавляясь тем самым от более заметных периодических пропусков звучания.

Этот простой и эффектный пример иллюстрирует один из компромиссов, которые часто используются при обработке звука. В следующей главе мы займемся *психоакустикой* – наукой, изучающей, восприятие людьми звуков.

Цифровой звук

Подобно тому как существует возможность преобразования звука из волн давления в воздухе в аналоговый электрический сигнал, допустимо преобразование меняющегося электрического сигнала в последовательность цифровых значений и наоборот. Однако, поскольку аналоговое и цифровое представления звука коренным образом различаются, при таком преобразовании неизбежна потеря информации. Хитрость заключается в том, чтобы понять, какая информация пропадает, из чего станет ясно, какую необходимо хранить.

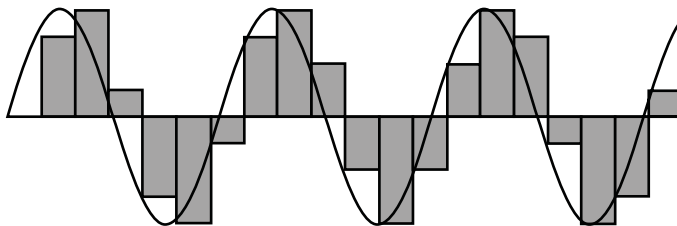


Рис. 1.4. Преобразование электронного представления

В цифровом виде звуковая волна представляется в виде ряда чисел (называемых *отсчетами* или *выборками*), которые соответствуют величине давления или электрическому напряжению в последовательные моменты времени. При квантовании аналогового сигнала вы столкнетесь с двумя основными проблемами. На рис. 1.4 показаны два варианта представления звуковой волны. Гладкая черная линия – это исходная электронная версия; числовая запись представлена последовательностью серых прямоугольников.

Во-первых, любое значение в последовательности цифрового представления соответствует определенному промежутку времени; каждый серый блок имеет некоторую толщину. Во-вторых, цифровое представление дискретно; набор допустимых значений для записи возможной высоты каждого блока ограничен. В результате высоты блоков не совсем точно совпадают с волной. Более подробно эти вопросы обсуждаются в третьей главе.

Две названные проблемы служат основными источниками погрешности цифрового представления звукового сигнала. Величину этой погрешности можно контролировать, изменяя конкретные параметры формата представления звука, но полностью от нее избавиться нельзя. Бессмысленно ставить вопрос: «Как устранить погрешность?» – скорее надо сформулировать его так: «Какая ошибка в меньшей степени повлияет на работу моего приложения?». Под воздействием других факторов, таких как объем данных и скорость обработки, вам в определенных ситуациях придется смириться с увеличением погрешности, обусловленной другими источниками, для того чтобы избежать более неприятных ошибок.

Дополнительные источники информации

Обработка звука – это огромная область знания, и в одной книге невозможно рассмотреть все ее аспекты. Прочитав данное руководство, вы приобретете знания о том, как с помощью самостоятельно написанных программ работать с файлами распространенных форматов, и у вас накопится коллекция программ на C++, которые помогут вам на первых порах. Однако существует множество других моментов, необходимых для изучения. Ниже приведен список источников информации, которые, на мой взгляд, необходимо упомянуть:

- Glenn White *“The Audio Dictionary”* (University of Washington Press, 1985). Эта книга представляет собой исчерпывающий справочник по терминам,

- используемым во всех областях, связанных с аудио. В книге изложена как история, так и подробная техническая информация;
- F. Richard Moore “*The Elements of Computer Music*” (Prentice-Hall, 1990). Это наиболее полный справочник по компьютерной музыке. Мур использовал разработанную им систему `music` в качестве базиса для описания всего, что связано с компьютерной музыкой, начиная от основополагающих аспектов синтеза звучания инструментов и заканчивая увлекательным разделом, посвященным случайным композициям;
 - Ken Pohlmann “*Principles of Digital Audio*”, Third Edition (McGraw-Hill, 1995). Эта книга представляет собой тщательный анализ цифрового звука с точки зрения аппаратного обеспечения. В частности, Полман тщательно объясняет, как при разработке аппаратного обеспечения учитываются вопросы наложения спектров и ошибки квантования;
 - Richard G. Lyon “*Understanding Digital Signal Processing*” (Addison Wesley Longman, 1997). Эта книга – отличное введение в сложный вопрос цифровой обработки звука. Особо отмечу описание проектирования цифровых фильтров, которое включает в себя много практических советов по программированию;
 - в журнале “*Computer Music Journal*” рассматриваются вопросы из всех областей, связанных с производством компьютерной музыки, в том числе и форматы аудиофайлов. Web-страница этого журнала содержит примеры статей и полезный архив звуков: <http://www.mitpress.mit.edu/Computer-Music-Journal/>.



Глава 2. Восприятие звука человеком

При тщательном выборе формата записи звуковой информации и метода обработки звука приходится учитывать множество критериев, в том числе объем данных и скорость процессора. Один из самых важных вопросов – насколько хорошо будет это звучать.

Если бы достаточно было купить соответствующий измеритель, направить его на громкоговоритель, воспроизводящий звук и получить значение, характеризующее качество звучания, все было бы очень просто. Производители дорогой стереоаппаратуры могли бы гордиться тем, что в их системах «качество звучания составляет 99,2%», а алгоритмы обработки звука точно измеряли бы ухудшение качества звучания. Создание столь полезного устройства, безусловно нереально, что отчасти обусловлено существенным, хотя и едва уловимым различием, существующим между характеристиками звука, которые можно *измерить* и теми, которые можно *услышать*. В конечном счете, единственные «приборы», которым можно доверять, – это расположенные у нас по бокам головы.

Наука, занимающаяся изучением того, как слышит человек, называется *психоакустикой* (это раздел *психофизики*). В течение последнего столетия ученым удалось создать довольно четкое описание механизма, благодаря которому человек слышит. Он исключительно сложен и запутан.

В дальнейшей дискуссии мы рассмотрим различия между единицами измерения восприятия, то есть понятиями, относящимися к ощущаемым человеком характеристикам звука, и понятиями, относящимися к математическим и физическим характеристикам звука.

Частота и высота тона

Музыкальные инструменты труба и туба похожи друг на друга, однако первая издает более «высокие» звуки. Описывая различие между их звучаниями, люди говорят о *высоте тона* – у трубы «тон выше», чем у тубы. Высота тона тесно связана с физической характеристикой звука, называемой *частотой*.

Период синусоидальной волны – это время, затрачиваемое на один полный цикл. Частота измеряется как количество полных циклов за одну секунду, единица измерения – 1 герц (краткое обозначение – Гц). Мы также будем измерять частоту в килогерцах [кГц], один килогерц равен тысяче герц. Для справки: люди слышат синусоидальные волны в диапазоне приблизительно от 30 до 20000 Гц.

Это определение частоты основано на том, что синусоидальные волны являются *периодическими*: форма колебаний волн в точности сохраняется. Несмотря

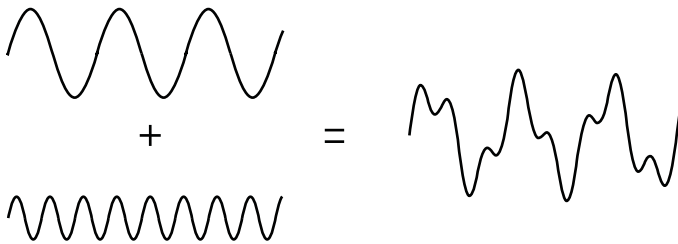


Рис. 2.1. При сложении двух синусоидальных волн образуется новый звук

на то, что большая часть математического аппарата, используемого для работы со звуком, основывается на базовом постулате повторяемости сигнала (в частности, на преобразовании Фурье, о котором мы поговорим в главе 24), в реальной жизни очень немногие звуки обладают свойством периодичности.

На практике непосредственно частота имеет значение только для синусоидальных волн.

К счастью, любой звук можно представить набором синусоидальных колебаний. И наоборот, любой звук можно синтезировать, сложив подходящий набор синусоидальных волн. Уравнение, проиллюстрированное рис. 2.1, работает в обе стороны: если читать его слева направо, оно показывает, как получить звук, сложив две синусоидальные волны, а справа налево – как разложить сложный сигнал на две отдельные синусоидальные составляющие.

Поскольку любой звук раскладывается на синусоидальные волны, мы можем построить *частотный спектр* звука. Спектр частот звуковой волны представляет собой график зависимости амплитуды от частоты. На рис. 2.2 показана амплитуда составляющих звук волн. В данном случае спектр очень прост: мы видим, что звук составлен волнами двух частот и амплитуда одной из них вдвое больше, чем у другой. В главе 24 рассказывается о том, как можно вычислить спектр частот звука.

В действительности люди говорят о частоте сложного звука. Обычно эта характеристика измеряется путем обработки частотного спектра выделением частоты синусоидальной волны, имеющей максимальную амплитуду. Однако данный подход не всегда можно использовать.

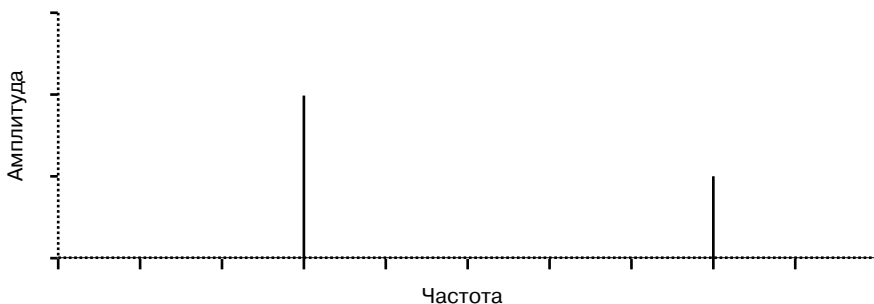


Рис. 2.2. Спектр частот звуковой волны, показанной на рис. 2.1

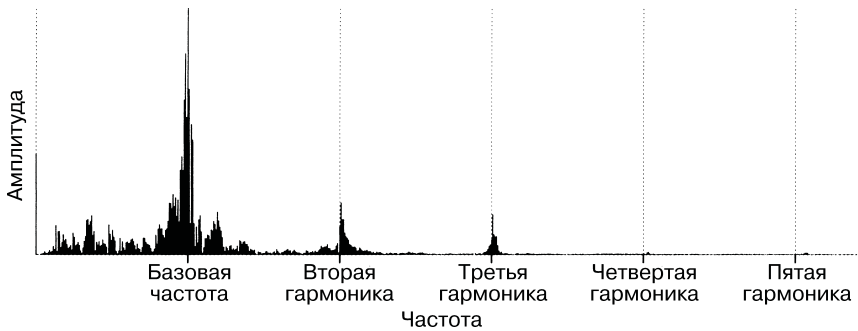


Рис. 2.3. Спектр частот реального музыкального инструмента

Рассмотрим рис. 2.3. Подобно спектрам частот большинства музыкальных инструментов этот спектр содержит последовательность равномерно расположенных пиков, называемых *гармониками*. Они соответствуют частотам, кратным некоторой *базовой частоте*. Характеристический звук (или *тембр*) музыкального инструмента нередко основывается на относительной громкости различных гармоник. Обычно базовая частота является также частотой самой сильной из составляющих звук синусоидальных волн.

Судя по всему, воспринимаемую человеком высоту тона всего звука в целом наиболее точно передает базовая частота. Данная зависимость достаточно устойчива, что позволяет использовать в некоторых компьютеризированных инструментах метод идентификации сыгранных нот по фундаментальной частоте звука. Этот метод называется *слежением за базовой частотой* или *слежением за высотой тона*. В результате, компьютеризированные инструменты могут автоматически следовать за мелодией, сыгранной на другом инструменте.

При рассмотрении частоты сложного звука, на наш взгляд, было бы лучше оперировать понятиями, относящимися к ощущениям человека: частотой сложного звука является частота синусоидальной волны такой же воспринимаемой высоты тона.

Высота тона в музыке

Большинство музыкантов западной школы пользуются набором высот тона, называемых *нотами*. Такой набор называется *равномерно темперированным звукорядом*, который представляет собой результат постепенной эволюции теории музыки, разработанной математиками Древней Греции.

Древние греки заслужили признание благодаря тому, что заметили зависимость между частотой и высотой тона. Они обнаружили, что в зависимости от длины колеблющейся струны, получаются звуки различной высоты тона (длина струны напрямую связана с частотой), и создали изящную математическую теорию, которая объяснила, почему комбинации определенных частот созвучны. Они открыли, например, что, если частота одного звука ровно вдвое превосходит частоту другого, такие звуки особенно подходят друг к другу. Сейчас это отношение

известно под названием *октава*. Оно получило настолько широкое распространение, что мы теперь часто относим отличающиеся на октаву тональности к одной и той же ноте. С течением времени музыканты Запада пришли к соглашению, что каждая октава должна быть поделена на 12 нот. Но, как я объясню чуть позже, возникло некоторое разногласие относительно того, какая именно частота должна соответствовать каждой из этих нот.

Греки сделали попытку объяснить, почему одни пары звуков образовывали *консонанс* – вместе они звучали согласованно, в то время как другие пары образовывали *диссонанс*. Разработанная ими теория основывалась на относительных частотах. Согласно этой теории, два звука образовывали консонанс, если отношение их частот можно было свести к небольшим числам. Пусть, например, имеются две идеальные синусоидальные волны, частота одной – 660 Гц, другой – 440 Гц. Тогда отношение их частот составит 660:440, или 3:2. Так как 3 и 2 – небольшие числа, можно было бы ожидать, что совместное звучание этих двух нот будет приятным. Согласно данной теории, высота нот определяется относительно: начинают с одной ноты, называемой *тоникой*, а затем устанавливают остальные ноты в соответствии с отношениями их частот к частоте тоники.

Самыми важными из этих соотношений являются *пятое* (отношение 3:2) и *третье* (отношение 5:4). Если мы назовем начальную ноту До, то сможем, используя предложенные греческими учеными отношения, определить систему из 12 нот, называемую *натуральным или чистым звукорядом*. Отношения для этих нот приведены в табл. 2.1. В первой колонке даны общепринятые названия нот западной школы. (Значок # читается как «диез», а b – как «бемоль»). Во второй и третьей колонке содержатся соответственно названия и отношения для различных музыкальных *интервалов*. Например, нота Ми – это «третья после До».

Чистый звукоряд позволяет получать очень благозвучные аккорды. Например, отношение 4:5:6 называется главной триадой (трихордом) и получается в результате объединения тоники с ее четвертой и пятой нотами (скажем, До-Ми-Соль).

Несмотря на это, чистый звукоряд – не очень гибкая структура. Если вы, например, настроите инструмент так, чтобы отношения его тональностей точно соответствовали чистому звукоряду, то не у всех нот будут третьи. При настройке До-мажор, приведенной в табл. 2.1, вы не сможете сыграть третью ноту после Ми (которая должна была бы иметь отношение 5:4 к ноте Ми или 25:16 к До). Наиболее подходящей нотой окажется Соль#, чья частота будет приблизительно на 3% выше, чем требуется.

Эта несогласованность имеет четко выраженные последствия в музыке. Например, если вы настроите свой любимый музыкальный инструмент в соответствии с чистым До-мажорным звукорядом, который приведен в табл. 2.1, и попытаетесь сыграть главную триаду, начинающуюся с ноты Ми, то получившийся аккорд будет звучать слегка фальшиво.

В зависимости от вашего вкуса это либо серьезная проблема, либо возможность придать вашим музыкальным композициям характерный оттенок. История знает примеры как одного подхода к этой проблеме, так и другого.

Таблица 2.1. Отношения частот для чистого звукоряда

Нота	Интервал	Отношение к тонике
C (До)	Тоника	1:1
C#/Db (До-диез/Ре-бемоль)		16:15
D (Ре)	Вторая	9:8
D#/Eb (Ре-диез/Ми-бемоль)	Минорная третья	6:5
E/Fb (Ми/Фа-бемоль)	Третья	5:4
F/E# (Фа/Ми-диез)	Четвертая	4:3
F#/Gb (Фа-диез/Соль-бемоль)		7:5
G (Соль)	Пятая	3:2
G#/Ab (Соль-диез/Ля-бемоль)		8:5
A (Ля)	Шестая	5:3
A#/Bb (Ля-диез/Си-бемоль)		9:5
B/Cb (Си/До-бемоль)	Седьмая	15:8
C/B# (До/Си-бемоль)	Октава	2:1

С появлением стандартного нотного письма (*общепринятая практическая нотация*) люди все в большей и большей степени стали рассматривать 12-тоновую шкалу как последовательность 12 отношений, называемых *полутонами*. Каждый полутон – это отношение частот двух последовательно идущих нот. При использовании чистого звукоряда, безусловно, не все полутона одинаковы. В табл. 2.2 к чистому звукоряду добавлена колонка, в которой приведены отношения между последовательно стоящими нотами (например, До-диез и До).

Было сделано немало попыток «поправить» расположение нот путем небольшого смещения частот некоторых из них. По правде говоря, тот вид чистого звукоряда, который представлен в данной книге, сформировался только в XVI столетии. В соответствии с настройкой, первоначально предложенной греческим математиком Пифагором, вся шкала конструировалась на основе последовательно идущих пятых, или чистых квинт. Данные *Пифагорова строя* также включены в табл. 2.2. Последовавшие попытки согласования шкалы настройки и чистого звукоряда Пифагора называли *темперациями*.

Настойчивые попытки принять единую темперацию в конце концов привели к появлению равномерной темперации, в которой все полутона были одинаковы. Для этого требовалось сместить частоты нот таким образом, чтобы отношения частот у стоящих друг за другом нот сравнялись. Так как у нас 12 нот, а одна октава по-прежнему соответствует соотношению частот 2:1, в итоге коэффициент отношения между последовательными тональностями получается равным $\sqrt[12]{2}$ – иррациональному числу, приблизительно равному 1,0595.

В табл. 2.3 сравниваются отношения частот для чистой интонации, шкалы настройки Пифагора и равномерно темперированной системы.

Равномерная темперация представляет собой изящный компромисс. Этот звукоряд позволяет использовать в качестве базиса любую из нот и по ходу дела стирает

Таблица 2.2. Отношения частот нот для различных звукорядов

Нота	Чистый звукоряд	Чистые полутона	Интервалы Пифагора	Полутона Пифагора
C (До)	1,0	1,0667	1,0	1,0679
C#/Db (До-диез/Ре-бемоль)	1,0667	1,0547	1,0679	1,0535
D (Ре)	1,125	1,0667	1,125	1,0679
D#/Eb (Ре-диез/Ми-бемоль)	1,2	1,0417	1,2013	1,0535
E/Fb (Ми/Фа-бемоль)	1,25	1,0667	1,2656	1,0679
F/E# (Фа/Е-диез)	1,3333	1,05	1,3515	1,0535
F#/Gb (Фа-диез/Соль-бемоль)	1,4	1,0714	1,4238	1,0535
G (Соль)	1,5	1,0667	1,5	1,0679
G#/Ab (Соль-диез/Ля-бемоль)	1,6	1,0417	1,6018	1,0535
A (Ля)	1,6667	1,08	1,6875	1,0679
A#/Bb (Ля-диез/Си-бемоль)	1,8	1,0417	1,8020	1,0535
B/Cb (Си/До-бемоль)	1,875	1,0667	1,8984	1,0535
C/B# (До/Си-бемоль)	2,0		2,0	

Таблица 2.3. Отношения частот нот в различных звукорядах

Нота	Чистый звукоряд	Пифагоров строй	Равномерная температура
C (До)	1,0	1,0	1,0
C#/Db (До-диез/Ре-бемоль)	1,0667	1,0679	1,0595
D (Ре)	1,125	1,125	1,1225
D#/Eb (Ре-диез/Ми-бемоль)	1,2	1,2013	1,1892
E/Fb (Ми/Фа-бемоль)	1,25	1,2656	1,2599
F/E# (Фа/Е-диез)	1,3333	1,3515	1,3348
F#/Gb (Фа-диез/Соль-бемоль)	1,4	1,4238	1,4142
G (Соль)	1,5	1,5	1,4983
G#/Ab (Соль-диез/Ля-бемоль)	1,6	1,6018	1,5874
A (Ля)	1,6667	1,6875	1,6818
A#/Bb (Ля-диез/Си-бемоль)	1,8	1,8020	1,7818
B/Cb (Си/До-бемоль)	1,875	1,8984	1,8877
C/B# (До/Си-бемоль)	2,0	2,0	2,0

различия между тонами. Исторически сложилось так, что ряд тонов стал ассоциироваться с определенной тональностью. В значительной степени на это повлияла неизбежная несогласованность, возникающая при попытке построить 12-тоновую шкалу, основываясь на идеальных третьих и четвертых.

Определив частоту базовой ноты, вы можете использовать указанные выше отношения для того, чтобы задать весь звукоряд. Чаще всего музыканты применяют в качестве базовой ноту Ля440, то есть ноту Ля с частотой звука 440 Гц. Воспользовавшись равномерной температурой, вы получите частоты, указанные в табл. 2.4.

Таблица 2.4. Частоты нот в соответствии с равномерной температурой (Гц)

C (До)	32,70	65,41	130,81	261,63	523,25	1046,50	2093,00
C#(До-диез)	34,65	69,30	138,59	277,18	554,36	1108,73	2217,46
D (Ре)	36,71	73,42	146,83	293,66	587,33	1174,66	2349,32
D# (Ре-диез)	38,89	77,78	155,56	311,13	622,25	1244,51	2489,02
E(Ми)	41,20	82,41	164,81	329,63	659,25	1318,51	2637,02
F(Фа)	43,65	87,31	174,61	349,23	698,46	1396,91	2793,83
F#(Фа-диез)	46,25	92,50	185,00	369,99	739,99	1479,98	2959,96
G (Соль)	49,00	98,00	196,00	391,99	783,99	1567,98	3135,96
G#(Соль-диез)	51,91	103,83	207,65	415,31	830,61	1661,22	3322,44
A (Ля)	55,00	110,00	220,00	440,00	880,00	1760,00	3520,00
A#(Ля-диез)	58,27	116,54	233,08	466,16	932,33	1864,65	3729,31
B(Си)	61,74	123,47	246,94	493,88	987,77	1975,53	3951,07
C(До)	65,41	130,81	261,63	523,25	1046,50	2093,00	4186,01

Одним из аргументов, свидетельствующих в пользу принятия равномерной температуры, является то, что ее применение позволяет сыграть все ноты на одном физическом инструменте, который не придется постоянно перестраивать. Эта проблема, однако, была полностью устранена после появления электронных инструментов. В то время как взятый в отдельности физический инструмент непросто перенастроить с До-мажор на Ми-мажор в пределах чистого звукоряда, электронный инструмент может мгновенно автоматически перенастраиваться, что позволит использовать подобные перемены даже в пределах одной песни. В некоторых современных композициях вообще не применяется 12-тоновый ряд. *Микротональные* композиции, ряд в которых поделен особым образом, или *атональные*, где фиксированные частоты для нот вообще не задаются, исполняют электронные инструменты.

Высота тона и частота звука

Теория древних греков, согласно которой высота тона и частота – одно и то же, привлекает своей простотой. Тем не менее проведенные в последнее время эксперименты, в ходе которых были сделаны попытки измерить ощущение высоты тона человеком, выявили ряд ситуаций, где эта простая идея срабатывала не совсем точно. Например, у громких звуков высота тона ниже, чем у тихих звуков той же частоты.

Теория звука, разработанная греками, также перестает работать на очень высоких или на очень низких частотах. Если тщательно протестировать правильно настроенное пианино, обнаружится, что высокие ноты настроены на несколько бóльшие, а низкие – на несколько меньшие частоты, чем можно было бы ожидать. Не так давно исследователи подтвердили то, что настройщики пианино уже знали давным-давно: человек обычно слышит высокие ноты чуть ниже, чем они есть на самом деле. Аналогичным образом, низкие ноты слышатся чуть выше. Установленное греками родство между высотой тона и частотой справедливо

только для среднего диапазона нот и теряет точность для более высоких или низких звуков.

Картина становится еще более запутанной, когда речь заходит о сложных звуках. Выше я упомянул, что у большинства музыкальных инструментов звучание состоит из сильных компонент на частотах (гармониках), кратных некоторой базовой частоте. Чаще всего эта базовая частота является еще и частотой самой сильной синусоидальной волны в звуке. В результате, мы зачастую, построив спектр частот звука и выделив в нем самую громкую составляющую, можем заранее определить, какую высоту тона будет иметь звук. Однако звук на базовой частоте в ряде случаев бывает совсем слабый или вообще отсутствует, а мы по-прежнему слышим звук той же самой высоты тона. Известно, что у некоторых инструментов, в том числе у фортепиано, существуют *растянутые частичные тона*, где частоты появления гармоник не совсем соответствуют значениям, кратным базовой частоте.

Шум

Если вы получили спектр частот, похожий на тот, что показан на рис. 2.2, то имеет смысл определить частоты с максимальной амплитудой и воспользоваться ими для описания свойств звука. Но что делать, если спектр звука имеет вид графика, показанного на рис. 2.4? В такой ситуации невозможно выделить частоту или хотя бы несколько частот, с помощью которых можно было бы попробовать охарактеризовать звук.

Звук, образованный колебаниями всего диапазона частот, подобный тому, спектр которого показан на рис. 2.4, называется *шумом*. Толкование этого слова, принятое в технике, отличается от общепризнанного. Свист высокого тона (издаваемый, например, старым монитором) может считаться шумом в бытовом смысле. Но у этого звука есть четко определенный спектр частот, и, следовательно, он *не может* считаться шумом в техническом смысле этого слова. Я буду использовать слово «шум» только в данном техническом смысле.

Шум, как ни странно, встречается повсюду. Большинство ударных инструментов, включая барабаны, деревянные бруски и тарелки, производит преимущественно шум. Много шумовых составляющих содержит голос человека. Шум

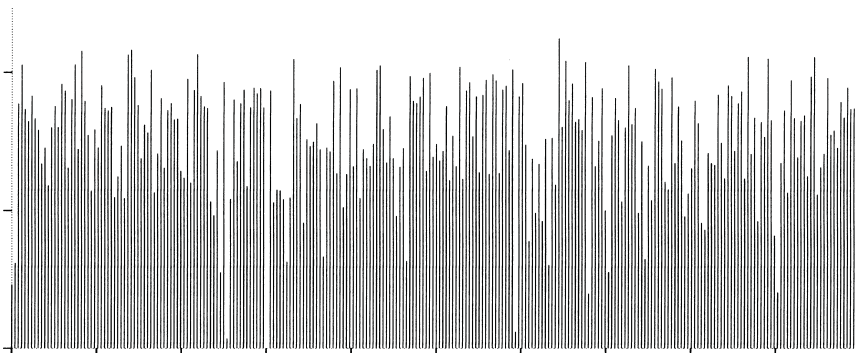


Рис. 2.4. Спектр частот шума