

А. П. НЕМИРКО
Л. А. МАНИЛО
А. Н. КАЛИНИЧЕНКО

**МАТЕМАТИЧЕСКИЙ
АНАЛИЗ
БИОМЕДИЦИНСКИХ
СИГНАЛОВ
И ДАННЫХ**



МОСКВА
ФИЗМАТЛИТ®
2017

УДК 51.761
ББК 32.81, 53.4
Н 50



*Издание осуществлено при поддержке
Российского фонда фундаментальных
исследований по проекту 16-17-00060,
не подлежит продаже*

Немирко А.П., Манило Л.А., Калининченко А.Н. **Математический анализ биомедицинских сигналов и данных.** — М.: ФИЗМАТЛИТ, 2017. — 248 с. — ISBN 978-5-9221-1720-3.

В книге рассмотрены математические методы анализа многомерных биомедицинских данных, задачи автоматизации медицинской диагностики методами теории статистических решений, теории таблиц решений и блок-схем алгоритмов. Подробно рассмотрены методы обработки и анализа биомедицинских сигналов и примеры применения анализа таких сигналов в медицинских диагностических и мониторинговых системах.

Книга может быть полезной специалистам, а также студентам и аспирантам соответствующего профиля.

ISBN 978-5-9221-1720-3

© ФИЗМАТЛИТ, 2017

© А. П. Немирко, Л. А. Манило,
А. Н. Калининченко, 2017

ОГЛАВЛЕНИЕ

Введение	5
Глава 1. Методы классификации образов и принятия диагностических решений	7
1.1. Методы распознавания образов	7
1.2. Предварительный анализ медицинских данных в пространстве меньшей размерности. Метод главных компонент	15
1.3. Линейный дискриминант Фишера	21
1.4. Дискриминантный анализ многоклассовой задачи с использованием критерия Фишера	29
1.5. Непараметрические методы обучения (алгоритм персептрона)	35
1.6. Построение линейных разделяющих функций с применением методов линейного программирования	41
1.7. Выравнивание символьных последовательностей	46
1.8. Медицинская диагностика на основе теории статистических решений	52
1.9. Таблицы решений и блок-схемы алгоритмов в медицинской диагностике	58
1.10. Характеристики точности и цены диагностики	68
1.11. Базы данных физиологических сигналов physionet.org	73
Глава 2. Методы анализа биомедицинских сигналов	79
2.1. Дискретизация и квантование сигналов	79
2.2. Цифровая фильтрация сигналов	86
2.3. Цифровые методы спектрального оценивания сигналов	118
2.4. Получение цифровых оценок корреляционных функций	129
2.5. Идентификация сложных биосигналов с использованием оценок условной энтропии	133
2.6. Распознавание биосигналов с нелинейными свойствами по параметрам аппроксимированной энтропии	138
2.7. Анализ фазовых портретов биомедицинских сигналов	150

Глава 3. Анализ электрокардиографических сигналов в мониторинговых и диагностических системах	161
3.1. Основные задачи и этапы непрерывного анализа ЭКГ	161
3.2. Характеристики электрокардиосигнала и помех.	163
3.3. Предварительная цифровая фильтрация ЭКГ	166
3.4. Обнаружение опасных аритмий по спектральному описанию электрокардиосигнала	173
3.5. Алгоритм обнаружения QRS-комплекса ЭКГ	179
3.6. Анализ форм ЭКГ и распознавание нарушений ритма	184
3.7. Анализ ишемических изменений ЭКГ	192
Глава 4. Анализ сигналов вариабельности сердечного ритма, электроэнцефалограмм и спирограмм в медицинских системах	196
4.1. Анализ вариабельности сердечного ритма и артериального давления	196
4.2. Анализ ВСР на основе модели управления водителем сердечного ритма.	203
4.3. Оценка глубины анестезии по ЭЭГ на основе спектральной энтропии	211
4.4. Оценка параметров дыхания при искусственной вентиляции легких	224
4.5. Автоматический анализ формы спирографических петель по их сигнатурам	231
Список литературы	239

Введение

Развитие систем диагностики медицинской назначения в значительной степени связано с появлением новых направлений и технологий исследований, требующих применения эффективных математических методов анализа биомедицинских данных. Материал настоящей книги посвящен рассмотрению теоретических основ различных методов анализа многомерных данных и их применению в задачах непрерывного контроля состояния больных. Основное внимание уделено методам распознавания данных на основе теории статистических решений, линейному дискриминантному анализу, применению таблиц решений в задачах медицинской диагностики, а также методам анализа биомедицинских сигналов, нашедшим практическое применение в системах мониторинга контроля различного назначения.

В четырех основных разделах монографии рассматриваются методы классификации образов и принятия диагностических решений; математические методы анализа биомедицинских сигналов; анализ электрокардиографических сигналов в мониторинговых и диагностических системах; анализ сигналов вариабельности сердечного ритма, электроэнцефалограмм и спирограмм в медицинских системах. Каждый из разделов включает теоретический материал: формулировку задачи исследования и способы ее решения. В задачах распознавания большое внимание уделено вопросам выбора пространства признаков и оптимизации процедур классификации. Кроме того, рассмотрен комплекс интересных прикладных задач. Они позволяют оценить возможности используемых теоретических подходов в конкретных медицинских приложениях. Предложены и исследованы оригинальные методы обнаружения QRS-комплекса и распознавания нарушений ритма по ЭКГ, анализа вариабельности сердечного ритма, распознавания биосигналов с нелинейными свойствами по параметрам аппроксимированной энтропии, оценки глубины анестезии по ЭЭГ на основе спектральной энтропии, автоматического анализа формы спирографических петель

по их сигнатурам. Решение каждой из задач сопровождается теоретическим обоснованием выбранного метода анализа данных, исследованием свойств предложенных преобразований на модельных сигналах, оценкой эффективности разработанных алгоритмов на множестве реальных биомедицинских сигналов (ЭКГ, ритмограмма, ЭЭГ, спирограмма).

Материал монографии рассчитан на исследователей и разработчиков медицинских интеллектуальных приборов и систем. Кроме того, представленный теоретический материал с примерами использования рассмотренных методов в медицинской практике представляется полезным для студентов и аспирантов соответствующего профиля, а также специалистов по биомедицинской инженерии и информатике.

Глава 1

МЕТОДЫ КЛАССИФИКАЦИИ ОБРАЗОВ И ПРИНЯТИЯ ДИАГНОСТИЧЕСКИХ РЕШЕНИЙ

1.1. Методы распознавания образов

Методы распознавания образов применяются для задач анализа данных, которые характеризуются малым числом объектов в классах и для которых важно определить степень сходства или различия классов. Часто необходимо создать алгоритм автоматического определения (расознавания) принадлежности объектов к определенным классам. Расознавание образов относится к автоматизации интеллектуальных задач принятия решений и опирается на теорию вероятностей, теорию матриц, математическую логику и математическую лингвистику. Методам распознавания образов посвящена обширная литература ([1.5–1.9] и др.).

Можно привести следующие примеры применения методов распознавания образов в различных областях человеческой деятельности:

- диагностика заболеваний [1.1, 1.10];
- распознавание и прогнозирование угрожающих жизни состояний больного, психофизиологических состояний человека по комплексу электрофизиологических показателей [1.11];
- автоматический анализ и интерпретация биомедицинских сигналов и изображений [1.1, 1.12];
- другие задачи: машинное зрение; распознавание речи; идентификация человека по голосу, отпечаткам пальцев, изображению глаза; чтение печатного и рукописного текста, топографических карт и т. п.

Представление образов в многомерном пространстве признаков

Допустим, мы имеем m классов объектов $\omega_1, \omega_2, \dots, \omega_m$. Каждый поступающий объект должен быть отнесен к одному из этих m классов и каждый объект описывается набором из n измерений (действительных чисел) $x_1, x_2, \dots, x_n$, которые называются признаками. Такой набор может быть представлен точкой в n -мерном евклидовом пространстве, или, иначе говоря, вектором признаков $\mathbf{X} = (x_1, x_2, \dots, x_n)$ в n -мерном признаковом пространстве, где $x_1, x_2, \dots, x_n$ — декартовы координаты конца вектора $\mathbf{X}$. Например, при диагностике состояний больного мы имеем два класса: ω_1 — А-состояние, ω_2 — В-состояние.

Каждый поступающий большой описывается 3 признаками (показателями), полученными по результатам его объективного обследования:

- x_1 — частота пульса,
 - x_2 — частота дыхания,
 - x_3 — среднее артериальное давление.
- В этом случае состояние одного больного, описываемое точкой (вектором) $\mathbf{X} = (x_1, x_2, x_3)$ в трехмерном пространстве, должно быть отнесено к одному из двух классов: ω_1 или ω_2 (рис. 1.1).

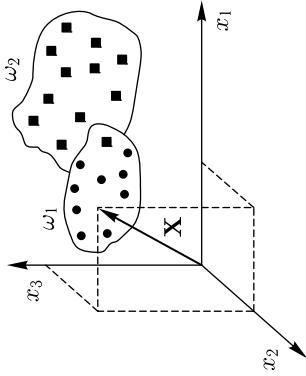


Рис. 1.1. Описание классов в признаковом пространстве

Разделяющие поверхности в признаковом пространстве. Дискриминантные функции

Таким образом, классификатор объектов является устройством, которое отображает точки из E^n в множество классов $\{\omega_1, \omega_2, \dots, \omega_m\}$ (рис. 1.2). Обозначим через R_i множество точек из E_n , которое отображается в класс ω_i .

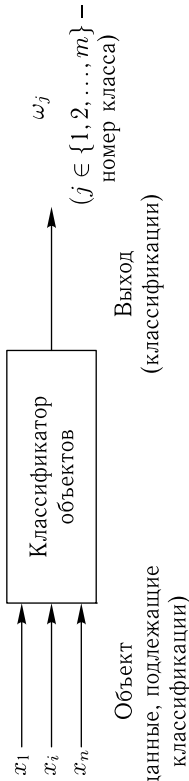


Рис. 1.2. Функция классификатора объектов

Тогда множеству классов $\{\omega_1, \omega_2, \dots, \omega_m\}$ будет соответствовать множество областей из E^n : $\{R_1, R_2, \dots, R_m\}$. Например, на рис. 1.3, где $n = 2$, $m = 3$, любая точка на плоскости в зависимости от ее принадлежности к областям R_1, R_2, R_3 отображается соответственно в один из классов $\omega_1, \omega_2, \omega_3$. Все три области отделены друг от друга поверхностями (кривыми в E^2), которые мы будем называть разделяющими.

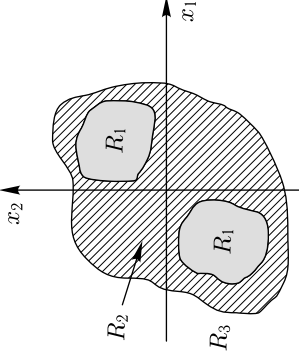


Рис. 1.3. Отображение объектов в множество классов

Разделяющие поверхности любого классификатора объектов можно полностью определить m скалярными функциями: $D_1(\mathbf{X}), \dots, D_m(\mathbf{X})$, называемыми дискриминантными функциями (ДФ). Они выбираются так, чтобы для всех $\mathbf{X} \in R_i$ выполнялось $D_i(\mathbf{X}) > D_j(\mathbf{X})$ при $i, j = 1, \dots, m; i \neq j$. Мы также предполагаем, что ДФ непрерывны на разделяющих поверхностях. Тогда поверхность, разделяющая смежные области R_i и R_j , определяется уравнением

$$D_i(\mathbf{X}) - D_j(\mathbf{X}) = 0. \tag{1.1}$$

На рис. 1.4 показаны три ДФ и области решений, порождаемые ими на плоскости x_1, x_2 .

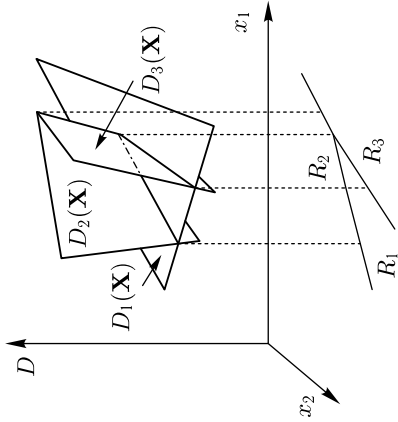


Рис. 1.4. Дискриминантные функции

Введение понятия дискриминантных функций преобразует модель классификатора объектов к виду рис. 1.5, где дискриминаторами названы блоки, вычисляющие дискриминантные функции.

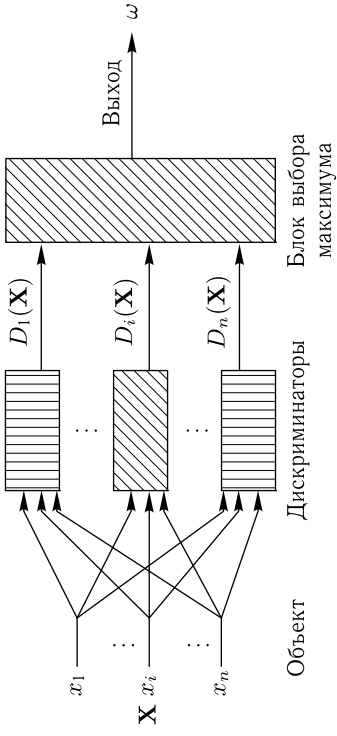


Рис. 1.5. Распознавание с помощью дискриминантных функций

Данные функции могут принадлежать к разным видам (линейные, кусочно-линейные, нелинейные, квадратичные). При решении задач распознавания образов существует проблема выбора вида ДФ.

Линейные дискриминантные функции

К простейшим дискриминантным функциям относятся линейные (ЛДФ), определяемые выражением

$$D_i(\mathbf{X}) = \sum_{k=1}^m w_{ik}x_k + w_{i(n+1)}, \quad i = 1, \dots, m.$$

Для $i = 1$ ЛДФ имеет вид

$$D_1(\mathbf{X}) = w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + \dots + w_{1(n+1)}.$$

В этом случае поверхность, разделяющая смежные области R_i и R_j , в E^n определяется из (1.1):

$$D_i(\mathbf{X}) - D_j(\mathbf{X}) = \sum_{k=1}^n w_k x_k + w_{n+1} = 0, \quad (1.2)$$

где $w_k = w_{ik} - w_{jk}$, $w_{n+1} = w_{i(n+1)} - w_{j(n+1)}$.

Полученное уравнение — это уравнение гиперплоскости в пространстве E^n .

Если $m = 2$, то уравнение (1.2) превращается в следующее:

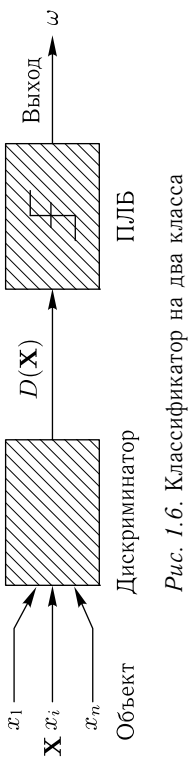
$$D(\mathbf{X}) = D_1(\mathbf{X}) - D_2(\mathbf{X}) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1} = 0,$$

и по знаку новой дискриминантной функции $D(\mathbf{X})$ мы можем отнести входной объект к одному из двух классов — ω_1 или ω_2 .

Если $D(\mathbf{X}) > 0$, то R_1 и входной объект относится к классу ω_1 .

Если $D(\mathbf{X}) < 0$, то R_2 и входной объект относится к классу ω_2 .

Введя пороговый логический блок (ПЛБ), классификатор для этого случая можно изобразить, как показано на рис. 1.6.



Вышеприведенное выражение для $D(\mathbf{X})$ можно также записать в виде

$$D(\mathbf{X}) = \mathbf{W}^T \mathbf{X} - a = 0,$$

где $\mathbf{W}$ — весовой вектор, a — пороговая величина. Удобно сделать вектор $\mathbf{W}$ единичным, разделив все компоненты на норму этого вектора. В этом случае, так как $\mathbf{W}^T \mathbf{X} = |\mathbf{X}| \cos \alpha$ (рис. 1.7), процедура распознавания заключается в вычислении проекции вектора $\mathbf{X}$ на направление весового вектора $\mathbf{W}$ и сравнении полученной величины с порогом a . Также ясно, что разделяющая гиперплоскость проходит перпендикулярно к $\mathbf{W}$ на расстоянии a от начала координат.

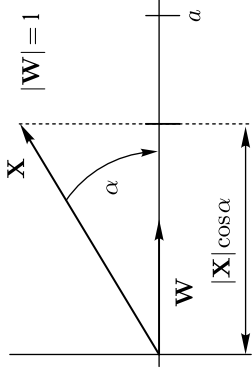


Рис. 1.7. К определению ЛДФ

Существует ряд методов определения $\mathbf{W}$ и a , основанных как на критериях теории статистических решений, так и на максимизации определенного некоторым образом расстояния между классами.

Классификатор ЭЭГ

Пример линейного классификатора — классификатор электроэнцефалограмм (ЭЭГ) (рис. 1.8). На вход устройства подается сигнал ЭЭГ, обозначенный как $z(t)$. В устройстве вычисляется 6 признаков (параметров):

- x_1 — число пересечений нулевой линии,
- x_2 — средняя амплитуда исходной ЭЭГ,
- x_3 — число максимумов и минимумов,
- x_4 — средняя амплитуда 1-й производной,
- x_5 — число точек перегиба,
- x_6 — средняя амплитуда 2-й производной.

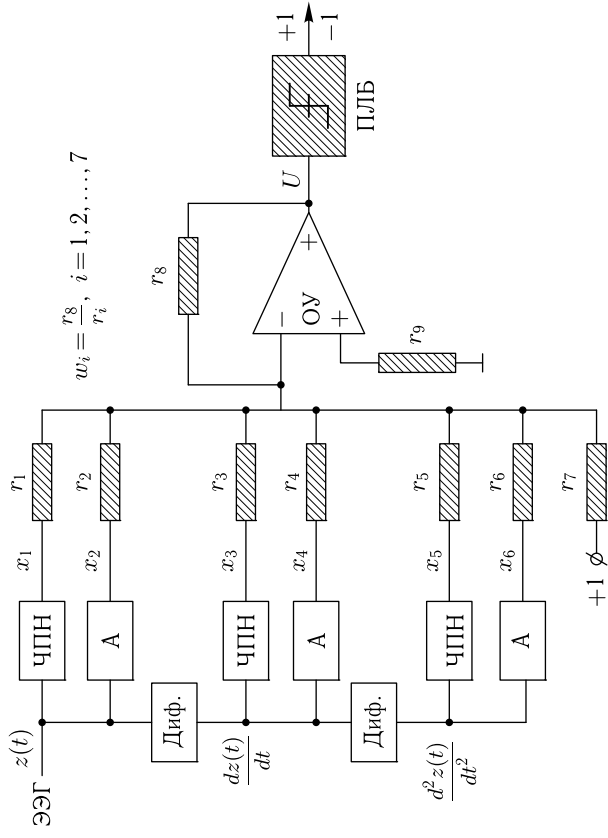


Рис. 1.8. Классификатор ЭЭГ. Обозначения: ЧПН — среднее число пересечений нулевой линии; А — средняя амплитуда сигнала; Диф. — дифференциатор; ОУ — операционный усилитель; ПЛБ — пороговый логический блок

Как видно из рисунка, сигнал U на выходе ОУ равен

$$U = \sum_{i=1}^6 w_i x_i + w_7,$$

где $w_i = \frac{r_8}{r_i}, i = 1, \dots, 7$.

В случае, когда некоторые w_i отрицательны, в схеме перед соответствующими r_i необходимо поставить инверторы. Сигнал на выходе ПЛБ равен $+1$, если $U > 0$, и -1 , если $U < 0$.

Реальная схема классификатора немного сложнее. Она разбивает ЭЭГ не на два, а на четыре класса за счет использования не одного, а трех порогов. В этом случае на рисунке соединения с r_7 удаляется, а правая часть рисунка приобретает вид рис. 1.9. Сравнение с порогом проводится по следующему правилу:

если $\sum_{i=1}^6 w_i x_i < w_7$, то ЭЭГ относится к классу 1;

иначе если $w_7 \leq \sum_{i=1}^6 w_i x_i < w_8$, то ЭЭГ относится к классу 2;

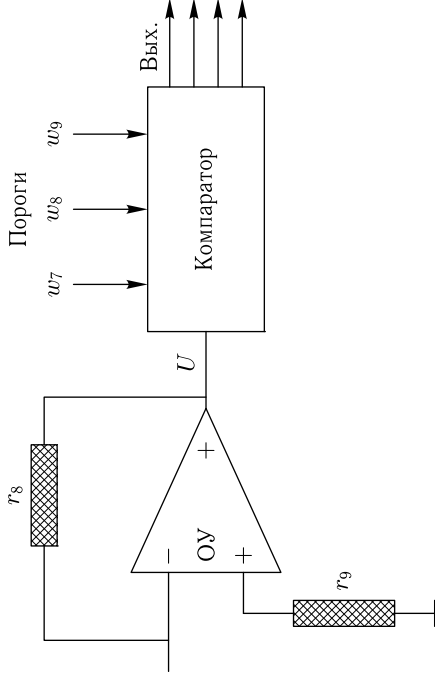


Рис. 1.9. Классификатор ЭЭГ на 4 класса

иначе если $w_8 \leq \sum_{i=1}^6 w_i x_i < w_9$, то ЭЭГ относится к классу 3;

иначе если $\sum_{i=1}^6 w_i x_i > w_9$, то ЭЭГ относится к классу 4.

Данный классификатор используется для определения глубины наркоза, стадий сна, для изучения биологических обратных связей, влияния лекарственных препаратов на ЭЭГ и для других целей.

Классификатор по минимуму расстояния

Пусть в n -мерном признаковом пространстве мы имеем m эталонных векторов $P_1, P_2, \dots, P_m$, каждый из которых является эталоном (шаблоном) своего класса, как изображено на рис. 1.10. Вектор $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$ соответствует классу ω_i . Классификатор по минимуму расстояния относит входной вектор X к тому классу ω_i , для которого расстояние $|X - P_i|$ между X и эталонным

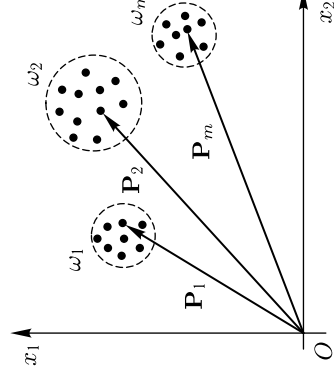


Рис. 1.10. Расположение эталонов

вектором $\mathbf{P}_i$ минимально. Вместо расстояния $|\mathbf{X} - \mathbf{P}_i|$ можно взять его квадрат $|\mathbf{X} - \mathbf{P}_i|^2$. Так как $|\mathbf{X} - \mathbf{P}_i| = \sqrt{(\mathbf{X} - \mathbf{P}_i)^T(\mathbf{X} - \mathbf{P}_i)}$, то $|\mathbf{X} - \mathbf{P}_i|^2 = (\mathbf{X} - \mathbf{P}_i)^T(\mathbf{X} - \mathbf{P}_i)$ и на основании свойства дистрибутивности скалярных произведений векторов имеем

$$|\mathbf{X} - \mathbf{P}_i|^2 = \mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{P}_i - \mathbf{P}_i^T\mathbf{X} + \mathbf{P}_i^T\mathbf{P}_i. \quad (1.3)$$

Произведение $\mathbf{X}^T\mathbf{X} = |\mathbf{X}|$ на расстояние не влияет, а на основании свойства коммутативности скалярных произведений векторов справедливо равенство $\mathbf{X}^T\mathbf{P}_i = \mathbf{P}_i^T\mathbf{X}$, так как

$$\sum_{j=1}^n x_j p_{ij} = \sum_{j=1}^n p_{ij} x_j.$$

Тогда, исключив $\mathbf{X}^T\mathbf{X}$ из выражения (1.3), поменяв знаки и разделив все члены на 2, получим дискриминантную функцию $D_i(\mathbf{X})$ для класса ω_i . Она имеет вид

$$D_i(\mathbf{X}) = \mathbf{X}^T\mathbf{P}_i - \frac{1}{2}\mathbf{P}_i^T\mathbf{P}_i, \quad i = 1, \dots, m. \quad (1.4)$$

Как и ранее, входной вектор $\mathbf{X}$ относится к классу ω_i , если

$$D_i(\mathbf{X}) > D_j(\mathbf{X}), \quad i, j = 1, \dots, m, \quad i \neq j.$$

Выражение (1.4) можно записать также в виде

$$D_i(\mathbf{X}) = x_1 w_{i1} + x_2 w_{i2} + \dots + x_n w_{in} + w_{i(n+1)}, \quad i = 1, \dots, m,$$

где

$$w_{ij} = p_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n;$$

$$w_{i(n+1)} = -\frac{1}{2}(p_{i1}^2 + p_{i2}^2 + \dots + p_{in}^2), \quad i = 1, \dots, m.$$

Таким образом, видно, что классификатор по минимуму расстояния является линейным классификатором. Такой классификатор целесообразно применять в случаях, когда каждый класс имеет свой эталонный объект $\mathbf{P}_i$, вокруг которого группируются остальные объекты этого класса.

При $m = 2$ (два класса)

$$D_1(\mathbf{X}) = \mathbf{X}^T\mathbf{P}_1 - \frac{1}{2}\mathbf{P}_1^T\mathbf{P}_1, \quad D_2(\mathbf{X}) = \mathbf{X}^T\mathbf{P}_2 - \frac{1}{2}\mathbf{P}_2^T\mathbf{P}_2,$$

$$D^O(\mathbf{X}) = D_1(\mathbf{X}) - D_2(\mathbf{X}) = \mathbf{X}^T\mathbf{P}_1 - \mathbf{X}^T\mathbf{P}_2 - \frac{1}{2}(\mathbf{P}_1^T\mathbf{P}_1 - \mathbf{P}_2^T\mathbf{P}_2).$$

Добавим к выражению в скобках член $\mathbf{P}_1^T\mathbf{P}_2$ и вычтем равный ему $\mathbf{P}_2^T\mathbf{P}_1$. Тогда

$$D^O(\mathbf{X}) = \mathbf{X}^T(\mathbf{P}_1 - \mathbf{P}_2) - \frac{1}{2}(\mathbf{P}_1^T\mathbf{P}_1 - \mathbf{P}_1^T\mathbf{P}_2 + \mathbf{P}_2^T\mathbf{P}_1 - \mathbf{P}_2^T\mathbf{P}_2),$$

$$D^O(\mathbf{X}) = \mathbf{X}^T(\mathbf{P}_1 - \mathbf{P}_2) - \frac{1}{2}(\mathbf{P}_1 + \mathbf{P}_2)^T(\mathbf{P}_1 - \mathbf{P}_2).$$

Разделим все члены последнего выражения на $|\mathbf{P}_1 - \mathbf{P}_2|$. Тогда

$$D(\mathbf{X}) = \mathbf{X}^T \frac{\mathbf{P}_1 - \mathbf{P}_2}{|\mathbf{P}_1 - \mathbf{P}_2|} - \frac{1}{2} \frac{(\mathbf{P}_1 + \mathbf{P}_2)^T(\mathbf{P}_1 - \mathbf{P}_2)}{|\mathbf{P}_1 - \mathbf{P}_2|},$$

$$D(\mathbf{X}) = \mathbf{W}^T\mathbf{X} - a, \quad \mathbf{W} = \frac{\mathbf{P}_1 - \mathbf{P}_2}{|\mathbf{P}_1 - \mathbf{P}_2|}, \quad a = -\frac{1}{2} \frac{(\mathbf{P}_1 + \mathbf{P}_2)^T\mathbf{W}}{|\mathbf{P}_1 - \mathbf{P}_2|}.$$

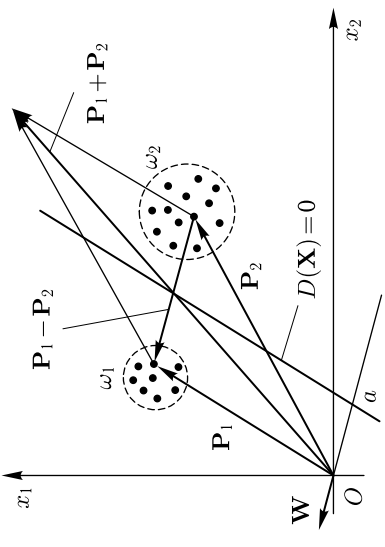


Рис. 1.11. Классификация по минимуму расстояния до эталонов двух классов
На рис. 1.11 показана разделяющая (линейная) граница $D(\mathbf{X}) = 0$ между двумя классами, проведенная для классификации по минимуму расстояния до эталонов классов.

1.2. Предварительный анализ медицинских данных в пространстве меньшей размерности. Метод главных компонент

Во многих задачах обработки и анализа биомедицинских данных исходное число p рассматриваемых, т.е. измеряемых на исследуемых объектах, признаков довольно велико. Для визуализации картины, простоты интерпретации и упрощения расчетов очень часто необходимо представить каждое из наблюдений в виде набора чисел, состоящего из *существенно меньшего*, чем p , числа признаков. При этом оставшиеся признаки могут либо выбираться из числа исходных, либо определяться по какому-либо правилу по совокупности исходных признаков, например как линейные комбинации последних.

Главные компоненты представляют собой новое множество исследуемых признаков

$$y^{(1)}, y^{(2)}, \dots, y^{(p)},$$

каждый из которых получен в результате некоторой линейной комбинации исходных признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, непосредственно

измеренных на объектах. Полученные в результате такого преобразования новые признаки $y^{(1)}, y^{(2)}, \dots, y^{(p)}$ могут обладать рядом удобных статистических свойств, например быть упорядоченными по степени рассеяния в изучаемой совокупности объектов: первый признак обладает наибольшей степенью рассеяния, т.е. наибольшей дисперсией, второй — меньшей и т.д.

Рассмотрим метод вычисления главных компонент [1.13]. Будем предполагать, что исследуемые наблюдения $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ извлечены из некоторой p -мерной генеральной совокупности, определяемой соответствующей вероятностной мерой. Из всех характеристик исследуемой совокупности данных существенное значение имеет ковариационная матрица $\Sigma = [\sigma_{ij}]$, где

$$\sigma_{ij} = M((x^{(i)} - a^{(i)})(x^{(j)} - a^{(j)})), \quad i, j = 1, 2, \dots, p. \quad (1.5)$$

Здесь $a^{(i)}$ — компоненты вектора $\mathbf{a}$ средних значений признаков $x^{(i)}$. Будем считать, что в (1.5) $\mathbf{a} = 0$. Этого всегда можно добиться, рассматривая в качестве исходных признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ не сами измерения, а их отклонения от выборочных средних значений.

Первой главной компонентой исследуемой совокупности наблюдений называется такая нормированная линейная комбинация p исходных признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$:

$$y^{(1)} = l_{11}x^{(1)} + l_{12}x^{(2)} + \dots + l_{1p}x^{(p)} = \mathbf{L}_1^T \mathbf{X}$$

$(\mathbf{L}_1^T = [l_{11} \ l_{12} \ \dots \ l_{1p}], \ l_{11}^2 + l_{12}^2 + \dots + l_{1p}^2 = 1)$, которая среди всех прочих нормированных линейных комбинаций обладает *наибольшей* дисперсией.

Вообще, i -й *главной компонентой* исследуемой генеральной совокупности ($i = 1, 3, \dots, p$) называют такую нормированную линейную комбинацию p исходных признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$:

$$y^{(i)} = l_{i1}x^{(1)} + l_{i2}x^{(2)} + \dots + l_{ip}x^{(p)} = \mathbf{L}_i^T \mathbf{X},$$

которая среди всех прочих линейных нормированных ($l_{i1}^2 + l_{i2}^2 + \dots + l_{ip}^2 = 1$) комбинаций, не коррелированных со всеми предшествующими главными компонентами $y^{(1)}, \dots, y^{(i-1)}$ (т.е. $\text{cov}(y^{(i)}, y^{(j)}) = M(y^{(i)}y^{(j)}) = 0$ для $j < i$), обладает наибольшей дисперсией.

Из определения следует, что, во-первых, главные компоненты $y^{(1)}, y^{(2)}, \dots, y^{(p)}$ пронумерованы в порядке убывания их дисперсий, т.е. $Dy^{(1)} \geq Dy^{(2)} \geq \dots \geq Dy^{(p)}$, где

$$Dy^{(i)} = M(\mathbf{L}_i^T \mathbf{X})^2 = M(\mathbf{L}_i^T \mathbf{X} \mathbf{X}^T \mathbf{L}_i) = \mathbf{L}_i^T \Sigma \mathbf{L}_i, \quad (1.6)$$

и, во-вторых, вектор $\mathbf{L}_i$, определяющий переход от $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ к $y^{(i)}$, является i -м *собственным вектором* ковариационной матрицы Σ , т.е. его компоненты $l_{i1}, l_{i2}, \dots, l_{ip}$ определяются как нормированное ($\sum_{j=1}^p l_{ij}^2 = 1$) решение системы уравнений

$$(\Sigma - \lambda_i \mathbf{I}) \mathbf{L}_i = 0. \quad (1.7)$$

Эта система линейных однородных уравнений имеет ненулевое решение только в том случае, когда ее определитель равен нулю, т.е.

$$|\Sigma - \lambda \mathbf{I}| = 0, \quad (1.8)$$

где $|\cdot|$ означает определитель матрицы, λ — неизвестное число, $\mathbf{I}$ — единичная матрица.

Выражение (1.8), в котором определитель матрицы $(\Sigma - \lambda \mathbf{I})$ является многочленом степени p относительно λ , называют *характеристическим уравнением*, а его корни λ_i , $i = 1, 2, \dots, p$, — *собственными значениями* матрицы Σ .

Каждому собственному значению λ_i , определяемому из (1.8), соответствует свой собственный вектор $\mathbf{L}_i$ матрицы Σ . Составляющие этого вектора $l_{i1}, l_{i2}, \dots, l_{ip}$ для конкретного значения λ_i могут быть найдены путем решения системы уравнений (1.7).

Из сопоставления (1.6)–(1.8) следует, что

$$Dy^{(i)} = \mathbf{L}_i^T \Sigma \mathbf{L}_i = \mathbf{L}_i^T \lambda_i \mathbf{I} \mathbf{L}_i = \lambda_i.$$

Таким образом, ковариационная матрица Σ_Y главных компонент $y^{(1)}, y^{(2)}, \dots, y^{(p)}$ будет иметь вид диагональной матрицы

$$\Sigma_Y = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_p \end{bmatrix}.$$

Учитывая, что преобразование, с помощью которого осуществляется переход от исходных компонент $\mathbf{X}$ к главным компонентам $\mathbf{Y}$ ($\mathbf{Y} = \mathbf{L}^T \mathbf{X}$), *ортogonalно*, исходные переменные $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ можно выразить через главные компоненты:

$$x^{(i)} = l_{i1}y^{(1)} + l_{i2}y^{(2)} + \dots + l_{ip}y^{(p)}$$

(в матричной записи $\mathbf{X} = \mathbf{L}\mathbf{Y}$).

Несмотря на то что для p исходных признаков получено p главных компонент, вклад большей их части в суммарную дисперсию оказывается небольшим. Поэтому, исключив из рассмотрения те компоненты, вклад которых мал, можно существенно сократить размерность исследуемого признакового пространства.

Ортогональное преобразование $\mathbf{Y} = \mathbf{L}^T \mathbf{X}$ оставляет инвариантной обобщенную дисперсию, а также сумму дисперсий компонент исходного вектора $\mathbf{X}$, и, как следствие, суммарная дисперсия $(Dy^{(1)} + Dy^{(2)} + \dots + Dy^{(p)})$ главных компонент, равная сумме собственных чисел $(\lambda_1 + \lambda_2 + \dots + \lambda_p)$, также равна сумме дисперсий $(Dx^{(1)} + Dx^{(2)} + \dots + Dx^{(p)})$ исходных признаков [1.13]. На этом основании принимается решение о том, сколько последних главных компонент можно исключить из дальнейшего рассмотрения. Для этого проводится анализ изменения относительной доли дисперсии

$$q(p') = \frac{Dy^{(1)} + Dy^{(2)} + \dots + Dy^{(p')}}{Dy^{(1)} + Dy^{(2)} + \dots + Dy^{(p)}} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_{p'}}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (1.9)$$

($1 \leq p' \leq p$), вносимой первыми p' главными компонентами, и определяется число компонент, которое целесообразно сохранить. Так, для примера кривой $q(p')$, изображенной на рис. 1.12, очевидно возможно сокращение размерности пространства с $p = 10$ до $p' = 3$, так как вклад семи последних главных компонент в суммарное рассеяние составляет не более 10%.

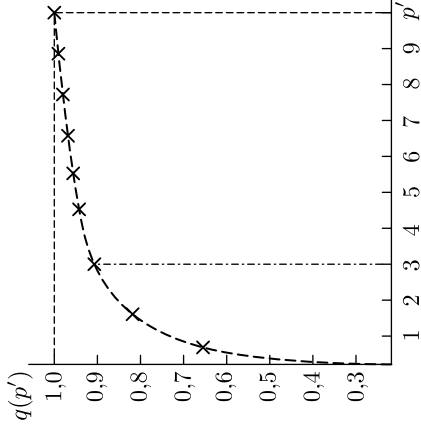


Рис. 1.12. Зависимость относительной доли суммарной дисперсии исследуемых признаков $q(p')$ от числа p' первых главных компонент (случай $p = 10$)

Анализируя лишь $p' = 3$ первых, т.е. наиболее весомых главных компонент, можно объяснить основную долю, а именно 90 % суммарной дисперсии, и выявить особенности рассеяния или группировки наблюдений.

Такой анализ главных компонент целесообразен, если величины $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ — составляющие вектора $\mathbf{X}$ — имеют одинаковую размерность. В случае если признаки $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ измеряются

в различных физических единицах, на этапе предобработки необходимо перейти к безразмерным признакам $x^{*(i)}$, например, с помощью нормирующего преобразования вида

$$x_{\nu}^{*(i)} = \frac{x_{\nu}^{(i)}}{\sqrt{\sigma_{ii}}} \quad (i = 1, 2, \dots, p, \quad \nu = 1, 2, \dots, n),$$

где σ_{ii} получено из (1.5), а затем вычислять главные компоненты относительно этих признаков $x^{*(1)}, \dots, x^{*(p)}$ и их ковариационной матрицы.

Пример. Исходные данные представляют собой результат описания множества реализаций электрокардиограмм — ЭКГ (ритмограмм) набором признаков $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}$. Эти признаки получены в процессе вычисления и анализа параметров энтропии Колмогорова, которая отражает степень сложности (хаотичности) сигналов. Выборка данных включает несколько классов ЭКГ: мерцательная аритмия (МА), нормальный ритм (НР), частая экстрасистолия (ЧЭ). Каждый класс представлен 25 объектами.

Требуется перейти от исходных признаков к главным компонентам, оценить долю дисперсии, которая обеспечивается первыми двумя главными компонентами $(y^{(1)}, y^{(2)})$, и отобразить объекты в пространстве новых признаков.

По данным измерений в соответствии с формулой (1.5) определена выборочная ковариационная матрица $\Sigma = [\sigma_{ij}]$

$$\begin{bmatrix} 0,327 & 0,177 & -4,359 \cdot 10^{-4} & -0,061 & -0,085 \\ 0,177 & 0,112 & 0,018 & -0,023 & -0,052 \\ -4,359 \cdot 10^{-4} & 0,018 & 0,026 & 0,015 & -1,636 \cdot 10^{-3} \\ -0,061 & -0,023 & 0,015 & 0,022 & 0,016 \\ -0,085 & -0,052 & -1,636 \cdot 10^{-3} & 0,016 & 0,064 \end{bmatrix}.$$

Решая в соответствии с (1.8) уравнение пятой степени относительно λ

$$\begin{vmatrix} 0,327 - \lambda & 0,177 & -4,359 \cdot 10^{-4} & -0,061 & -0,085 \\ 0,177 & 0,112 - \lambda & 0,018 & -0,023 & -0,052 \\ -4,359 \cdot 10^{-4} & 0,018 & 0,026 - \lambda & 0,015 & -1,636 \cdot 10^{-3} \\ -0,061 & -0,023 & 0,015 & 0,022 - \lambda & 0,016 \\ -0,085 & -0,052 & -1,636 \cdot 10^{-3} & 0,016 & 0,064 - \lambda \end{vmatrix} = 0,$$

находим собственные значения $\lambda_1 = 0,463$; $\lambda_2 = 0,047$; $\lambda_3 = 0,038$; $\lambda_4 = 2,7 \cdot 10^{-3}$; $\lambda_5 = 8,1 \cdot 10^{-4}$. Они упорядочены в порядке убывания их значений.

Подставляя последовательно численные значения $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ в систему уравнений (1.7) и решая ее относительно неизвестных $\mathbf{L}_i = [l_{i1} \ l_{i2} \ l_{i3} \ l_{i4} \ l_{i5}]^T, i = 1, 2, 3, 4, 5$, находим

$$\mathbf{L}_1 = \begin{bmatrix} -0,835 \\ -0,469 \\ -0,015 \\ 0,148 \\ 0,246 \end{bmatrix}, \quad \mathbf{L}_2 = \begin{bmatrix} -0,228 \\ 0,431 \\ 0,724 \\ 0,453 \\ -0,182 \end{bmatrix}, \quad \mathbf{L}_3 = \begin{bmatrix} 0,217 \\ 0,132 \\ 0,172 \\ 0,089 \\ 0,947 \end{bmatrix},$$

$$\mathbf{L}_4 = \begin{bmatrix} 0,421 \\ -0,575 \\ 0,018 \\ 0,696 \\ -0,085 \end{bmatrix}, \quad \mathbf{L}_5 = \begin{bmatrix} 0,162 \\ -0,469 \\ 0,667 \\ -0,53 \\ -0,039 \end{bmatrix}.$$

В качестве главных компонент получаем

$$\begin{aligned} y^{(1)} &= -0,84x^{(1)} - 0,47x^{(2)} - 0,02x^{(3)} + 0,15x^{(4)} + 0,25x^{(5)}, \\ y^{(2)} &= -0,23x^{(1)} + 0,43x^{(2)} + 0,72x^{(3)} + 0,45x^{(4)} - 0,18x^{(5)}, \\ y^{(3)} &= 0,22x^{(1)} + 0,13x^{(2)} + 0,17x^{(3)} + 0,09x^{(4)} + 0,95x^{(5)}, \\ y^{(4)} &= 0,42x^{(1)} - 0,58x^{(2)} + 0,02x^{(3)} + 0,70x^{(4)} - 0,09x^{(5)}, \\ y^{(5)} &= 0,16x^{(1)} - 0,47x^{(2)} + 0,67x^{(3)} - 0,53x^{(4)} - 0,04x^{(5)}. \end{aligned}$$

Здесь под $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}$ подразумеваются отклонения параметров энтропии от их средних значений.

Вычисление относительной доли суммарной дисперсии, обусловленной пятью главными компонентами, в соответствии с формулой (1.9) дает следующие результаты:

$$\begin{aligned} q(1) &= \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5} = 0,84, \\ q(2) &= \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5} = 0,925, \\ q(3) &= \frac{\lambda_1 + \lambda_2 + \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5} = 0,994, \\ q(4) &= \frac{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5} = 0,999, \\ q(5) &= 1. \end{aligned}$$

По результатам вычислений можно сделать вывод о том, что почти вся информация (92,5%) об изменчивости энтропийных характеристик наблюдаемой выборки ЭКГ-данных (75 объектов) содержится в первых двух главных компонентах.

На рис. 1.13 приведено изображение множества ритмограмм в пространстве двух первых главных компонент для трех классов ЭКГ (МА, НР, ЧЭ).

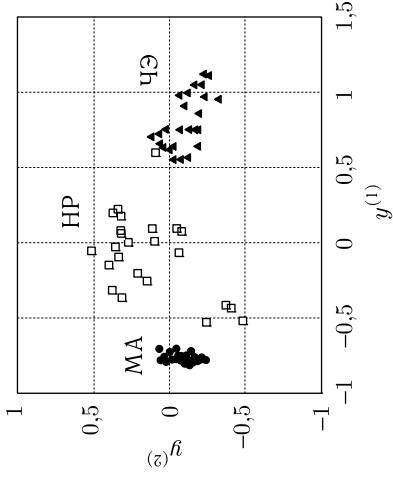


Рис. 1.13. Представление ритмограмм классов: мерцательная аритмия (МА), нормальный ритм (НР) и частая экстрасистолия (ЧЭ) в пространстве двух первых главных компонент

Как видно из рисунка, описание объектов параметрами энтропии позволяет получить группировку объектов, соответствующую трем классам рассматриваемых ЭКГ-сигналов. Наиболее плотная группировка объектов наблюдается в классе МА. Возможно линейное разделение классов, причем, используя энтропийные характеристики ритмограмм, можно эффективно решить задачу распознавания опасного нарушения (МА) на фоне альтернативных групп аритмий (ЧЭ и НР с выраженной дыхательной аритмией).

1.3. Линейный дискриминант Фишера

Построение решающих функций для распознавания двух классов объектов ω_1 и ω_2 по некоторому описанию $\mathbf{X} = (x_1, x_2, \dots, x_L)$ представляет собой непростую задачу. Известный байесовский подход, основанный на вычислении отношения правдоподобия, оптимален, поскольку связан с наилучшим критерием, минимизирующим вероятность ошибочного решения [1.5]. Однако на практике его трудно реализовать, так как для проведения вычислений необходимо знать условные плотности вероятностей, процедура оценивания которых довольно сложна. Более простым оказывается линейный классификатор, который можно построить, используя разные критерии оптимизации параметров решающей функции, в том числе критерий Фишера [1.6].

Для двухклассовой задачи линейная дискриминантная функция (ЛДФ), являющаяся разделяющей гиперплоскостью, имеет вид

$$D(\mathbf{X}) = \mathbf{W}^T \mathbf{X} - a = 0, \quad \text{или} \quad \mathbf{W}^T \mathbf{X} = a, \quad (1.10)$$

где $\mathbf{W} = (w_1, w_2, \dots, w_L)$ — весовой вектор единичной длины, a — скалярная пороговая величина.

Алгоритм распознавания, позволяющий отнести неизвестный объект $\mathbf{X}$ к одной из групп, в этом случае имеет следующий вид:

если $\mathbf{W}^T \mathbf{X} < a$, то ω_1 (1-й класс);

если $\mathbf{W}^T \mathbf{X} \geq a$, то ω_2 (2-й класс).

На рис. 1.14 изображены данная ЛДФ и соответствующий ей алгоритм распознавания.

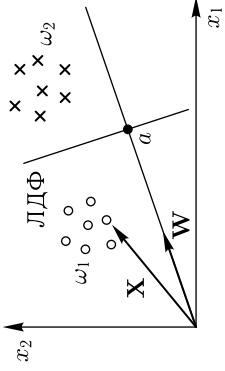


Рис. 1.14. Пример построения линейной дискриминантной функции

Для того чтобы определить составляющие вектора $\mathbf{W}$ и порог a , можно воспользоваться критерием минимизации ошибки классификации. Однако для этого потребуется довольно сложная итеративная процедура поиска параметров решающего правила. Эту задачу можно решить, применяя другой критерий оптимизации — критерий Фишера. Он позволяет найти такой вектор $\mathbf{W}$, при котором проекции точек классов на выбранное направление разделяются наилучшим образом. Порог a определяется исходя из критерия оптимальности разбиения объектов на два класса. В этом случае алгоритм поиска неизвестных параметров $\mathbf{W}$ и a можно представить следующим образом:

- 1) найти $\mathbf{W}$, используя критерий Фишера;
 - 2) спроецировать точки обоих классов на прямую, определяемую положением $\mathbf{W}$;
 - 3) решить одномерную задачу по поиску наилучшей величины a , например по критерию минимума ошибок классификации.
- Если допустить, что оба класса распределены по нормальному закону с одинаковыми ковариационными матрицами $\Sigma_1 = \Sigma_2 = \Sigma$, то из теории известно, что решение 1-го шага вышеприведенного алгоритма определяется выражением

$$\mathbf{W} = \Sigma^{-1}(\mathbf{M}_1 - \mathbf{M}_2), \quad (1.11)$$

где $\mathbf{M}_1$ и $\mathbf{M}_2$ — векторы средних значений двух классов. В случае $\Sigma_1 \neq \Sigma_2$ в качестве Σ можно использовать усредненную ковариационную матрицу в виде

$$\Sigma = \frac{1}{2} (\Sigma_1 + \Sigma_2). \quad (1.12)$$

Так как после вычисления $\mathbf{W}$ по формуле (1.11) его надо проецировать для получения вектора единичной длины, т. е. разделить на $|\mathbf{W}|$, то коэффициент $\frac{1}{2}$ в (1.12) можно отбросить. Тогда $\Sigma = \Sigma_1 + \Sigma_2$.

Оценки $\mathbf{M}_1$, $\mathbf{M}_2$ и Σ_1 , Σ_2 определяются выражениями

$$\mathbf{M}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{X}_i^{(1)},$$

$$\mathbf{M}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{X}_i^{(2)},$$

$$\Sigma_j = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (\mathbf{X}_i^{(j)} - \mathbf{M}_j)(\mathbf{X}_i^{(j)} - \mathbf{M}_j)^T, \quad j = 1, 2,$$

где $\mathbf{X}_i^{(1)}$ — i -й объект из 1-го класса; $\mathbf{X}_i^{(2)}$ — i -й объект из 2-го класса; n_1 — число объектов 1-го класса; n_2 — число объектов 2-го класса.

Если спроецировать точки объектов каждого класса на направление $\mathbf{W}$, то каждый вектор $\mathbf{X}$ превратится в соответствующий скаляр y : $y_1^{(1)} = \mathbf{W}^T \mathbf{X}_1^{(1)}$, $y_2^{(1)} = \mathbf{W}^T \mathbf{X}_2^{(1)}$ и т. д. (рис. 1.15).

Тогда средние значения для выборок спроецированных точек равны

$$m_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i^{(1)} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{W}^T \mathbf{X}_i^{(1)} = \mathbf{W}^T \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{X}_i^{(1)} = \mathbf{W}^T \mathbf{M}_1,$$

$$m_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i^{(2)} = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{W}^T \mathbf{X}_i^{(2)} = \mathbf{W}^T \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{X}_i^{(2)} = \mathbf{W}^T \mathbf{M}_2.$$

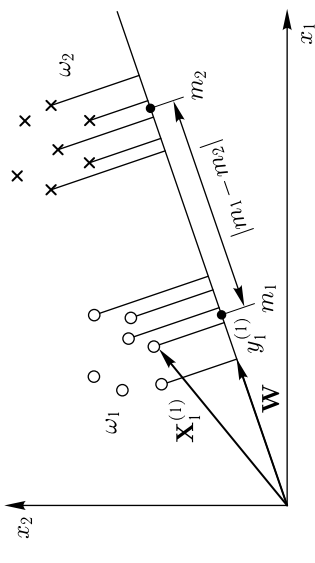


Рис. 1.15. Определение проекций объектов двух классов на прямую, продолжающую вектор $\mathbf{W}$

Очевидно, что расстояние между m_1 и m_2 будет отражать удаленность классов друг от друга после их проецирования на $\mathbf{W}$. Это расстояние, оценивающее межклассовый разброс, задается выражением

$$(m_1 - m_2)^2 = (\mathbf{W}^T \mathbf{M}_1 - \mathbf{W}^T \mathbf{M}_2)^2 = \mathbf{W}^T (\mathbf{M}_1 - \mathbf{M}_2) (\mathbf{M}_1 - \mathbf{M}_2)^T \mathbf{W} = \mathbf{W}^T \mathbf{S}_B \mathbf{W}, \quad (1.13)$$

где $\mathbf{S}_B = (\mathbf{M}_1 - \mathbf{M}_2)(\mathbf{M}_1 - \mathbf{M}_2)^T$ — матрица разброса между классами. Также можно определить выборочный разброс для двух классов s_1^2 и s_2^2 в виде

$$s_1^2 = \sum_{i=1}^{n_1} (y_i^{(1)} - m_1)^2 = \sum_{i=1}^{n_1} (\mathbf{W}^T \mathbf{X}_i^{(1)} - \mathbf{W}^T \mathbf{M}_1)^2, \\ s_2^2 = \sum_{i=1}^{n_2} (y_i^{(2)} - m_2)^2 = \sum_{i=1}^{n_2} (\mathbf{W}^T \mathbf{X}_i^{(2)} - \mathbf{W}^T \mathbf{M}_2)^2.$$

Этот показатель, как и дисперсия, является оценкой внутриклассового рассеяния — чем он больше, тем и рассеяние данных больше. По s_1^2 и s_2^2 можно определить «средний» внутриклассовый разброс в виде $s_1^2 + s_2^2$.

Для определения s_1^2 и s_2^2 через статистические параметры выборок классов определим матрицу разброса $\mathbf{S}_j$ для j -го класса (аналог корреляционной матрицы) в виде

$$\mathbf{S}_j = \sum_{i=1}^{n_j} (\mathbf{X}_i^{(j)} - \mathbf{M}_j)(\mathbf{X}_i^{(j)} - \mathbf{M}_j)^T$$

и «усредненную» матрицу разброса для двух классов в виде $\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$. Тогда

$$s_j^2 = \sum_{i=1}^{n_j} (\mathbf{W}^T \mathbf{X}_i^{(j)} - \mathbf{W}^T \mathbf{M}_j)^2 = \sum_{i=1}^{n_j} \mathbf{W}^T (\mathbf{X}_i^{(j)} - \mathbf{M}_j) (\mathbf{X}_i^{(j)} - \mathbf{M}_j)^T \mathbf{W} = \\ = \mathbf{W}^T \left[\sum_{i=1}^{n_j} (\mathbf{X}_i^{(j)} - \mathbf{M}_j) (\mathbf{X}_i^{(j)} - \mathbf{M}_j)^T \right] \mathbf{W} = \mathbf{W}^T \mathbf{S}_j \mathbf{W}, \quad j = 1, 2; \\ s_1^2 + s_2^2 = \mathbf{W}^T \mathbf{S}_1 \mathbf{W} + \mathbf{W}^T \mathbf{S}_2 \mathbf{W} = \mathbf{W}^T (\mathbf{S}_1 + \mathbf{S}_2) \mathbf{W} = \mathbf{W}^T \mathbf{S}_W \mathbf{W}. \quad (1.14)$$

Теперь можно определить линейный дискриминант Фишера — линейную функцию с максимальным отношением разброса между классами к «среднему» разбросу внутри классов. Тогда ЛДФ Фишера определяется как такой вектор $\mathbf{W}$, для которого критерий

$$J(\mathbf{W}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

максимален. Для $\mathbf{W}$, найденного по данному критерию, классы, спроецированные на направление $\mathbf{W}$, максимально удалены друг от друга.

С учетом выражений (1.13) и (1.14) вышеприведенный критерий записывается как

$$J(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}}.$$

Анализ этой формулы [1.6] показывает, что максимум $J(\mathbf{W})$ достигается при

$$\mathbf{W} = \mathbf{S}_W^{-1} (\mathbf{M}_1 - \mathbf{M}_2), \quad (1.15)$$

что почти совпадает с выражением (1.11) для нормальных классов с равными ковариационными матрицами.

Преобразование признакового пространства на основе линейного дискриминанта Фишера

Можно улучшить качество классификации в методе линейного дискриминанта Фишера, если для распознавания применять не один, а большее число признаков, найденных как ортогональные весовые векторы в пространствах меньшей размерности с помощью критерия Фишера. В работе [1.27] выведена формула для рекуррентного вычисления дополнительных ортогональных весовых векторов. Для исходного n -мерного признакового пространства формулу (1.15) можно переписать в виде

$$\mathbf{W}_n = \mathbf{S}_n^{-1} (\mathbf{m}_1 - \mathbf{m}_2). \quad (1.16)$$

Спроектируем все данные на плоскость, перпендикулярную $\mathbf{W}_n$. Тогда на этой плоскости, которая является $(n-1)$ -мерным признаковым пространством, можно найти наилучший весовой вектор, используя критерий Фишера. Очевидно, что на этой плоскости

$$\mathbf{W}_{n-1} = \mathbf{S}_{n-1}^{-1} (\mathbf{m}_1 - \mathbf{m}_2), \quad (1.17)$$

где

$$\mathbf{m}_{1n-1} = \mathbf{m}_1 - m_1 \mathbf{W}_n = \mathbf{m}_1 - \mathbf{W}_n^T \mathbf{m}_1 \mathbf{W}_n, \\ \mathbf{m}_{2n-1} = \mathbf{m}_2 - m_2 \mathbf{W}_n = \mathbf{m}_2 - \mathbf{W}_n^T \mathbf{m}_2 \mathbf{W}_n, \\ \mathbf{m}_{1n-1} - \mathbf{m}_{2n-1} = \mathbf{m}_1 - \mathbf{m}_2 - \mathbf{W}_n^T (\mathbf{m}_1 - \mathbf{m}_2) \mathbf{W}_n. \quad (1.18)$$

При вычислении $\mathbf{S}_{n-1}$ обращаемая матрица может оказаться вырожденной. В этом случае вместо вычисления явной обратной матрицы можно вычислять псевдообратную матрицу:

$$\mathbf{S}_{n-1} = \mathbf{S}_1 - \mathbf{S}_1 \mathbf{W}_n - \mathbf{S}_2 - \mathbf{S}_2 \mathbf{W}_n, \\ \mathbf{S}_{1n-1} = \sum_{x \in \omega_1} (\mathbf{X}_{n-1} - \mathbf{m}_{1n-1})(\mathbf{X}_{n-1} - \mathbf{m}_{1n-1})^T, \quad (1.19)$$

где

$$\begin{aligned} \mathbf{X}_{n-1} - \mathbf{m}_{1_{n-1}} &= (\mathbf{X}_n - \mathbf{W}_n^T \mathbf{X}_n \mathbf{W}_n) - (\mathbf{m}_{1_n} - \mathbf{W}_n^T \mathbf{m}_{1_n} \mathbf{W}_n) = \\ &= (\mathbf{X}_n - \mathbf{m}_{1_n}) - \mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n, \\ \mathbf{S}_{1_{n-1}} &= \sum_{x \in \omega_1} [(\mathbf{X}_n - \mathbf{m}_{1_n}) - \mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n] [(\mathbf{X}_n - \mathbf{m}_{1_n}) - \\ &- \mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n]^T = \sum_{x \in \omega_1} [(\mathbf{X}_n - \mathbf{m}_{1_n})(\mathbf{X}_n - \mathbf{m}_{1_n})^T - \\ &- \mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n (\mathbf{X}_n - \mathbf{m}_{1_n})^T - \\ &- (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n^T \mathbf{W}_n (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n^T + B], \end{aligned}$$

где

$$B = \mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n \cdot [\mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n]^T.$$

Обозначим $\alpha = \mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n})$. Это скаляр. Отметим, что также $\alpha = (\mathbf{X}_n - \mathbf{m}_{1_n})^T \mathbf{W}_n$. Тогда $B = \alpha \mathbf{W}_n \alpha \mathbf{W}_n^T$ и

$$\begin{aligned} \mathbf{S}_{1_{n-1}} &= \sum_{x \in \omega_1} [(\mathbf{X}_n - \mathbf{m}_{1_n})(\mathbf{X}_n - \mathbf{m}_{1_n})^T - \\ &- \alpha \mathbf{W}_n (\mathbf{X}_n - \mathbf{m}_{1_n})^T - \alpha (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n^T + B]. \quad (1.20) \end{aligned}$$

Так как

$$\begin{aligned} \sum_{x \in \omega_1} [-\alpha \mathbf{W}_n (\mathbf{X}_n - \mathbf{m}_{1_n})^T - \alpha (\mathbf{X}_n - \mathbf{m}_{1_n}) \mathbf{W}_n^T] &= \\ = \sum_{x \in \omega_1} [(\alpha \mathbf{W}_n \mathbf{m}_{1_n}^T - \alpha \mathbf{W}_n \mathbf{X}_n^T) + (\alpha \mathbf{m}_{1_n} \mathbf{W}_n^T - \alpha \mathbf{X}_n \mathbf{W}_n^T)] &= \\ = (\alpha N_1 \mathbf{W}_n \mathbf{m}_{1_n}^T - \alpha N_1 \mathbf{W}_n \mathbf{m}_{1_n}^T) + (\alpha N_1 \mathbf{m}_{1_n} \mathbf{W}_n^T - \alpha N_1 \mathbf{m}_{1_n} \mathbf{W}_n^T) &= 0, \end{aligned}$$

где N_1 — число элементов в множестве ω_1 , то из (1.19) следует

$$\mathbf{S}_{1_{n-1}} = \sum_{x \in \omega_1} (\mathbf{X}_n - \mathbf{m}_{1_n})(\mathbf{X}_n - \mathbf{m}_{1_n})^T + \sum_{x \in \omega_1} B = \mathbf{S}_{1_n} + \sum_{x \in \omega_1} B. \quad (1.21)$$

Так как

$$\begin{aligned} \sum_{x \in \omega_1} B &= \sum_{x \in \omega_1} \alpha \mathbf{W}_n \alpha \mathbf{W}_n^T = \\ &= \sum_{x \in \omega_1} \mathbf{W}_n^T (\mathbf{X}_n - \mathbf{m}_{1_n})(\mathbf{X}_n - \mathbf{m}_{1_n})^T \mathbf{W}_n (\mathbf{W}_n \mathbf{W}_n^T) = \\ &= \mathbf{W}_n^T \left[\sum_{x \in \omega_1} (\mathbf{X}_n - \mathbf{m}_{1_n})(\mathbf{X}_n - \mathbf{m}_{1_n})^T \right] \mathbf{W}_n (\mathbf{W}_n \mathbf{W}_n^T) = \\ &= \mathbf{W}_n^T \mathbf{S}_{1_n} \mathbf{W}_n (\mathbf{W}_n \mathbf{W}_n^T), \end{aligned}$$

то

$$\mathbf{S}_{1_{n-1}} = \mathbf{S}_{1_n} + \mathbf{W}_n^T \mathbf{S}_{1_n} \mathbf{W}_n (\mathbf{W}_n \mathbf{W}_n^T)$$

и

$$\begin{aligned} \mathbf{S}_{n-1} &= \mathbf{S}_{1_{n-1}} + \mathbf{S}_{2_{n-1}} = \mathbf{S}_{1_n} + \mathbf{S}_{2_n} + \mathbf{W}_n^T (\mathbf{S}_{1_n} + \mathbf{S}_{2_n}) \mathbf{W}_n (\mathbf{W}_n \mathbf{W}_n^T) = \\ &= \mathbf{S}_n + \mathbf{W}_n^T \mathbf{S}_n \mathbf{W}_n (\mathbf{W}_n \mathbf{W}_n^T). \quad (1.22) \end{aligned}$$

Окончательно из (1.17) с учетом (1.18) получаем

$$\mathbf{W}_{n-1} = \mathbf{S}_{n-1}^{-1} [(\mathbf{m}_{1_n} - \mathbf{m}_{2_n}) - \mathbf{W}_n^T (\mathbf{m}_{1_n} - \mathbf{m}_{2_n}) \mathbf{W}_n], \quad (1.23)$$

или, подставляя в (1.23) выражение (1.22):

$$\begin{aligned} \mathbf{W}_{n-1} &= [\mathbf{S}_n + \mathbf{W}_n^T \mathbf{S}_n \mathbf{W}_n (\mathbf{W}_n \mathbf{W}_n^T)]^{-1} \times \\ &\times [(\mathbf{m}_{1_n} - \mathbf{m}_{2_n}) - \mathbf{W}_n^T (\mathbf{m}_{1_n} - \mathbf{m}_{2_n}) \mathbf{W}_n]. \quad (1.24) \end{aligned}$$

В $(n-2)$ -мерном признаковом пространстве получаем соответственно

$$\begin{aligned} \mathbf{W}_{n-2} &= [\mathbf{S}_{n-1} + \mathbf{W}_{n-1}^T \mathbf{S}_{n-1} \mathbf{W}_{n-1} (\mathbf{W}_{n-1} \mathbf{W}_{n-1}^T)]^{-1} \times \\ &\times [(\mathbf{m}_{1_{n-1}} - \mathbf{m}_{2_{n-1}}) - \mathbf{W}_{n-1}^T (\mathbf{m}_{1_{n-1}} - \mathbf{m}_{2_{n-1}}) \mathbf{W}_{n-1}] \end{aligned}$$

и так далее.

В результате найдено рекуррентное выражение для последовательного вычисления добавочных признаков.

Для примера выберем два множества 3-мерных данных $f1$ и $f2$, приведенных на с. 28. Столбцы этих матриц — признаки, а строки — отдельные объекты. Известно, что данные множества линейно разделимы в 3-мерном пространстве. Применим к ним метод главных компонент, линейный дискриминант Фишера и линейный дискриминант Фишера с одним добавочным признаком, полученным по формуле (1.24).

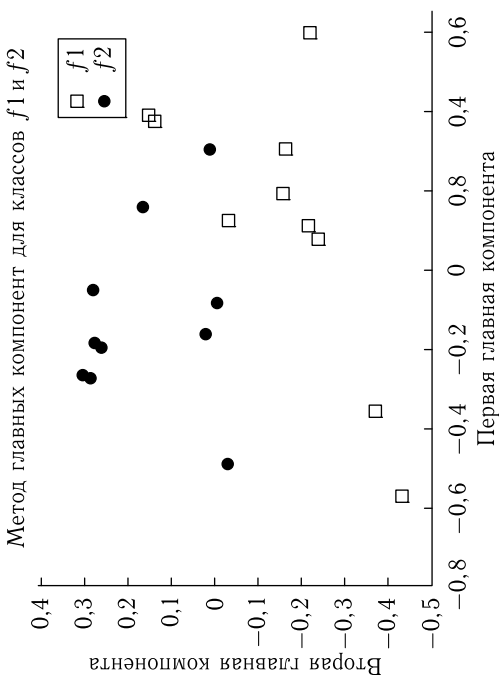


Рис. 1.16. Анализ множеств $f1$ и $f2$ по методу главных компонент

f_1	f_2				
0,7	0,3	1,2	0,4	0,2	0,8
0,5	0,7	1,0	0,2	0,2	0,7
0,4	1,0	0,4	0,9	0,3	0,5
0,7	0,7	1,0	0,8	0,3	0,6
0,6	0,6	1,5	0,5	0,6	0,4
0,6	0,6	1,2	0,6	0,5	0,7
0,6	0,5	1,0	0,4	0,4	1,2
0,4	0,9	0,6	0,6	0,3	1,0
0,5	0,6	1,1	0,3	0,2	0,6
0,8	0,3	1,2	0,5	0,5	0,8

Результат вычисления первых двух главных компонент показан на рис. 1.16. Из него видно, что метод главных компонент не обеспечивает линейную разделимость классов.

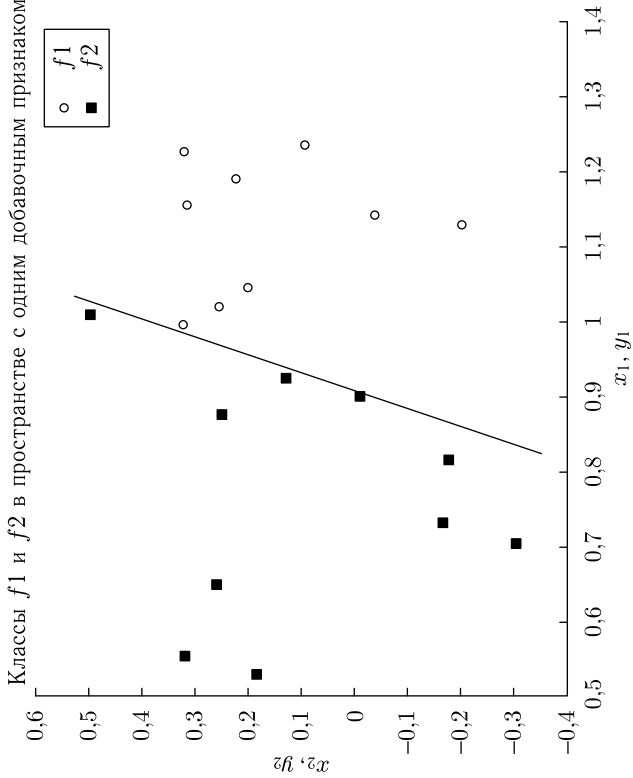


Рис. 1.17. Анализ множеств f_1 и f_2 методом линейного дискриминанта Фишера с одним добавочным признаком. Ось абсцисс — вектор $\mathbf{W}_1$, ось ординат — вектор $\mathbf{W}_2$

Реализация метода линейного дискриминанта Фишера с одним добавочным признаком на данных множествах показана на рис. 1.17.

Из рис. 1.17 видно, что использованием только одного весового вектора $\mathbf{W}_1$, найденного по критерию Фишера, линейной разделимости достичь не удастся. Добавочный же признак (весовой вектор $\mathbf{W}_2$) обеспечивает полную линейную разделимость классов f_1 и f_2 . Такой же результат получен при нахождении весового вектора с помощью простого однослойного персептрона.

На рис. 1.18 изображена область пересечений двух классов ирисов Фишера: виргинского и разноцветного (справа) [1.28] в сокращенном пространстве, образованном двумя весовыми векторами $\mathbf{W}_1$ и $\mathbf{W}_2$.

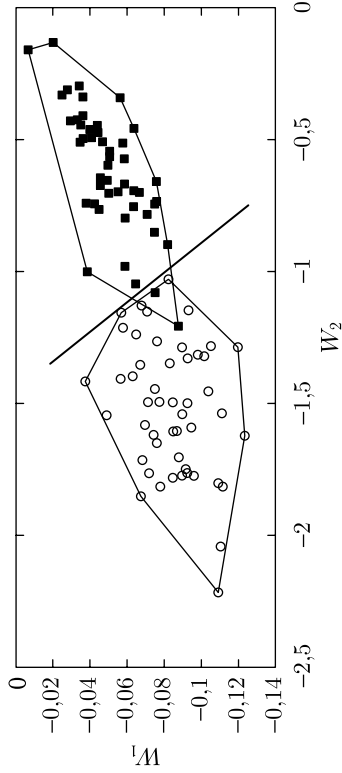


Рис. 1.18. Анализ ирисов Фишера — виргинского (слева) и разноцветного (справа) — в редуцированном пространстве методом линейного дискриминанта Фишера с одним добавочным признаком

Из рис. 1.18 видно, что только в двумерном пространстве можно достичь нулевой ошибки классификации виргинских ирисов при минимальной ошибке классификации разноцветных ирисов.

Проведенные эксперименты с другими данными показали, что добавочный признак улучшает разделимость классов. Это может оказаться важным для некоторых задач, особенно требующих полной разделимости классов. В предложенном подходе могут использоваться и другие критерии расстояния между классами.

1.4. Дискриминантный анализ многоклассовой задачи с использованием критерия Фишера

Сложнее решается задача распознавания сигналов в том случае, когда число классов больше двух (на практике такие задачи возникают довольно часто). Так, в системах кардиологического наблюдения необходимо надежно обнаруживать опасные аритмии в момент их внешнего возникновения, но не менее важно распознавать нарушения — предвестники тяжелых состояний пациента. В этой связи возникает

задача построения решающих функций для многих классов электрокардосигналов.

При наличии c классов необходимо построить $(c - 1)$ решающую функцию. Решение данной задачи возможно путем сведения ее к набору двухклассовых задач или при помощи множественного дискриминантного анализа. Последний подход основан на применении того же критерия оптимизации J и обобщения определений для матриц разброса между классами и внутри классов. Критерий J , оценивающий степень разделения заданных классов сигналов, в общем виде задается следом матрицы

$$J = \text{tr}(\mathbf{S}_W^{-1} \mathbf{S}_B), \quad (1.25)$$

где $\mathbf{S}_B$ — матрица рассеяния между классами; $\mathbf{S}_W$ — обобщенная матрица рассеяния внутри классов [1.6]. Его также нужно *максимизировать*.

Для случая двух классов нахождение линейного преобразования $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$, максимизирующего значение J , сводится к определению лишь одного собственного значения матрицы $(\mathbf{S}_W^{-1} \mathbf{S}_B)$. Отсюда следует возможность определения $\mathbf{W}$ с максимальным отношением разброса между классами к разбросу внутри классов в виде

$$\mathbf{W} = \mathbf{S}_W^{-1} (\mathbf{M}_1 - \mathbf{M}_2),$$

где $\mathbf{S}_W = 0.5(\Sigma_1 + \Sigma_2)$.

В случае c классов при переходе из L -мерного пространства признаков в $(c - 1)$ -мерное пространство проекции объектов найдутся с помощью того же матричного преобразования $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$, но $\mathbf{W}$ — это уже матрица размером $L \times (c - 1)$. Нахождение ее связано с представлением в формуле (1.25) матрицы $\mathbf{S}_B$ в следующем виде [1.14]:

$$\mathbf{S}_B = \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j (\mathbf{M}_i - \mathbf{M}_j)(\mathbf{M}_i - \mathbf{M}_j)^T = \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j \mathbf{S}_B^{(i,j)},$$

где n_i и n_j — частоты появления объектов, образующих классы ω_i и ω_j . Матрица $\mathbf{S}_W$ определяется в том же виде: $\mathbf{S}_W = \sum_{j=1}^c n_j \mathbf{S}_j$, где $\mathbf{S}_j$ — матрица разброса для j -го класса.

Ранг матрицы $\mathbf{S}_B$ равен $(c - 1)$, и это значение определяет число собственных значений и собственных векторов, задаваемых матрицей $\mathbf{W}$.

Недостаток применения выражения (1.25) связан с тем, что при увеличении числа классов критерий J ориентирован в основном на большие межгрупповые расстояния и слабо отражает взаимное расположение близко расположенных классов. А это в свою очередь ограничивает возможность применения метода для классификации трудно-разделимых групп объектов.

Оптимизировать процедуру построения решающих правил можно путем сведения ее к набору задач попарной классификации с весовыми коэффициентами $a_{i,j}$, усиливающими влияние на критерий J близко расположенных классов [1.14]. В этом случае обобщенное выражение для критерия J принимает вид

$$J = \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j a_{i,j} \text{tr}[(\mathbf{W}^T \mathbf{S}_W \mathbf{W})^{-1} \cdot (\mathbf{W}^T \mathbf{S}_B^{(i,j)} \mathbf{W})]. \quad (1.26)$$

Весовую функцию $a_{i,j}$ можно связать с ценой ошибки распознавания классов ω_i и ω_j . В указанной работе предлагается использовать веса в виде функции ошибок $\text{erf}\left(\frac{\eta - t}{\sigma}\right)$, где t — граница решающего правила, а η и σ — параметры распределений. Эта функция вычисляется для заданных групп объектов (i и j) исходя из предположений о нормальном законе распределения и равенстве ковариационных матриц. В этом случае критерий J может быть приближен к оценке достоверности распознавания объектов, представленных с классами, путем суммирования вероятностей правильного решения при их попарной классификации.

Нахождение элементов матрицы $\mathbf{W}$ сводится к задаче определения собственных значений матрицы

$$\mathbf{S}_W^{-1} \cdot \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j a_{i,j} \cdot \mathbf{S}_B^{(i,j)}.$$

Они могут быть найдены как корни характеристического уравнения $|\mathbf{S}_B - \lambda \mathbf{S}_W| = 0$, а затем получены $d = c - 1$ собственных векторов $\mathbf{W}_i$ как решение системы уравнений $(\mathbf{S}_B - \lambda_i \mathbf{S}_W) \cdot \mathbf{W}_i = 0$. В новом пространстве размерности d значение критерия J определяется в виде $J = \sum_{i=1}^d \lambda_i$.

Задачу выбора вида $a_{i,j}$ в (1.26) можно решить следующим образом. Рассмотрим в двумерном пространстве (x_1, x_2) два класса объектов ω_i и ω_j с нормальным законом распределения и единичными матрицами ковариации (рис. 1.19). Если расстояние между центрами этих классов обозначить как $\Delta_{i,j} = |\mathbf{m}_i - \mathbf{m}_j|$, где $\mathbf{m}_i$ и $\mathbf{m}_j$ — векторы средних значений, то, проецируя эти классы на новое направление $\mathbf{V}$, получим изменение расстояния между ними в зависимости от угла α между направлением, соединяющим центры классов, и вектором $\mathbf{V}$. Эту зависимость можно представить как $\Delta_{i,j}^{(\mathbf{V})} = \Delta_{i,j} \cos \alpha = \Delta_{i,j} \sin \beta$, где β — дополнительный угол. При равных априорных вероятностях

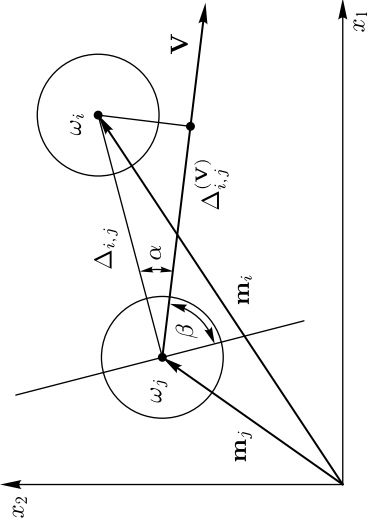


Рис. 1.19. Двумерное пространство (x_1, x_2) с двумя классами объектов ω_i и ω_j появления объектов обоих классов вероятностно правильное распознавание определяется в виде

$$\gamma_{i,j} = \frac{1}{2} + \gamma'_{i,j} = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\Delta_{i,j}^{(V)}}{2\sqrt{2}} \right],$$

где $\operatorname{erf}(\cdot)$ — функция ошибок.

Тогда для случая с классами с одинаковыми распределениями критерий $J^{(\gamma)}$, оценивающий среднюю точность распознавания, можно записать в виде

$$J^{(\gamma)} = \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j \gamma_{i,j}, \quad (1.27)$$

а критерий J , оценивающий степень расхождения классов (1.26), примет вид

$$J = \sum_{i=1}^{c-1} \sum_{j=i+1}^c n_i n_j a_{i,j} \operatorname{tr}[(\mathbf{V}^T \mathbf{S}_B^{(i,j)} \mathbf{V})]. \quad (1.28)$$

Сравнивая (1.27) и (1.28), веса $a_{i,j}$ можно задать в виде

$$a_{i,j} = \gamma'_{i,j} / \operatorname{tr}(\mathbf{V}^T \mathbf{S}_B^{(i,j)} \mathbf{V})$$

для случая наилучшего взаимного расположения двух классов (ω_i, ω_j) , что соответствует совпадению направлений векторов $\mathbf{V}$ и $\mathbf{m}_{i,j} = (\mathbf{m}_i - \mathbf{m}_j)$. При этом $\alpha = 0$; $\beta = \pi$; $\operatorname{tr}(\mathbf{V}^T \mathbf{S}_B^{(i,j)} \mathbf{V}) = (\Delta_{i,j})^2$, и метр $a_{i,j}$ определится следующим образом:

$$a_{i,j} = \gamma'_{i,j} / (\Delta_{i,j})^2 = \frac{1}{2(\Delta_{i,j})^2} \operatorname{erf} \left(\frac{\Delta_{i,j}}{2\sqrt{2}} \right).$$

Такой вид задания весовых функций предлагается в работе [1.14]. В области малых значений $\Delta_{i,j}$, например, при распознавании объектов по параметрам нормированного спектра или иным нормированным

параметрам, веса $a_{i,j}$ можно задать, используя приближение функции ошибок полиномиальной функцией

$$a_{i,j} \approx \frac{1}{8\sqrt{\pi} \cdot x_{i,j}} \left(1 - \frac{x_{i,j}^2}{3} + \frac{x_{i,j}^4}{215} \right), \quad (1.29)$$

где $x_{i,j} = \left(\frac{\Delta_{i,j}}{2\sqrt{2}} \right)$, $\Delta_{i,j} \leq \sqrt{2}$, $x_{i,j} \leq 0.5$.

Этот способ нахождения весовых функций $a_{i,j} = a(\Delta_{i,j})$ можно применить и для многоклассовой задачи [1.15, 1.16], предполагая, что каждый из c классов имеет матрицу внутригруппового рассеяния, равную $\mathbf{S}_W$. Тогда для каждой пары классов $(\omega_i, \omega_j; i, j = 1, \dots, c; i \neq j)$ в исходном L -мерном пространстве признаков необходимо найти евклидово расстояние $\Delta_{i,j}$ между их центрами и определить веса $a_{i,j}$, используя выражение (1.29). Максимизация критерия J (1.26) позволит перейти к процедуре нахождения собственных векторов $\mathbf{W}_i$, $i = 1, \dots, c-1$, и анализу групп объектов в пространстве признаков пониженной размерности.

Применение метода для распознавания опасных аритмий

Эффективность применения рассмотренных процедур оценивалась по результатам экспериментов, выполненных на реальных данных, включающих записи ЭКС длительностью более 20 минут [1.16]. Все реализации ЭКГ получены из стандартной компьютерной базы ЭКГ-сигналов MIT-BIH.

В качестве примера на рис. 1.20 приведены результаты отображения объектов на плоскость, заданную координатами (y_1, y_2) в пространстве собственных векторов $\mathbf{W}_1, \mathbf{W}_2$, которые получены при распознавании следующих трех классов сигналов: ω_1 — опасные аритмии (желудочковая фибрилляция и трепетание желудочков); ω_2 — пароксизмальная тахикардия и фоновый ритм, представленный разными формами экстрасистолии; ω_3 — пируэтная тахикардия, являющаяся переходной формой опасного нарушения ритма. Оптимизация пространства проведена по критерию J (1.26) с применением весовых функций.

В качестве исходного описания объектов, представленных фрагментами электрокардиосигнала длительностью 2 с, дискретизированными с частотой 250 Гц, использован упорядоченный набор 28 спектральных признаков. Они получены для частотной области, ограниченной частотой 15 Гц, с применением перекрывающихся сегментов. Некоррированные оценки спектральной плотности мощности вычислены при ширине спектрального окна $\Delta f = 0.976$ Гц, но при этом шаг по частотной оси выбран вдвое меньшим этой величины и составил 0.488 Гц. В этом случае удается сохранить особенности формы спектра анализируемых сигналов при относительной устойчивости получаемых оценок спектральной плотности мощности (СПМ).

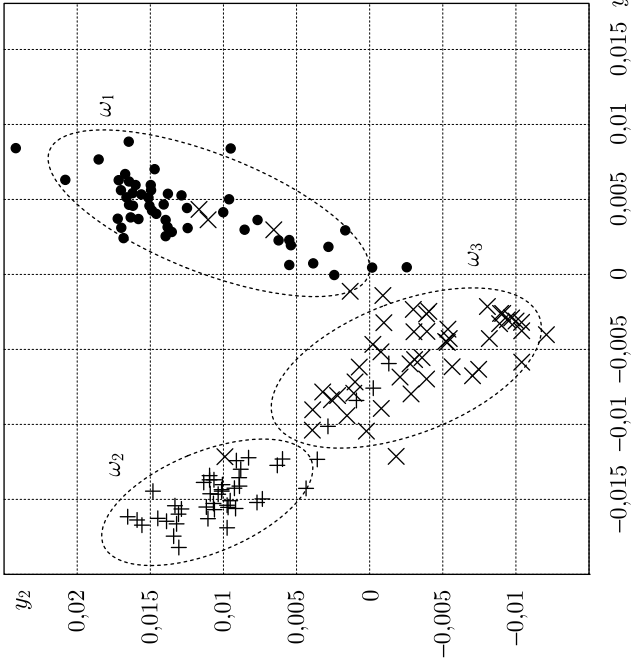


Рис. 1.20. Представление объектов классов ω_1 , ω_2 , ω_3 в пространстве векторов $\mathbf{W}_1$, $\mathbf{W}_2$, полученных с применением весовых функций

На рисунке выделены области нахождения представленных групп объектов, причем видно, что классы частично пересекаются. Соотношение расстояний между центрами классов в исходном пространстве составило $\Delta_{1,2} = 0,42$; $\Delta_{1,3} = 0,36$; $\Delta_{2,3} = 0,22$, откуда следует, что исходно классы ω_2 и ω_3 расположены значительно ближе друг к другу, чем к ω_1 . Вычисление весов по формуле (1.29), а именно $a_{1,2} = 0,623$; $a_{1,3} = 0,707$; $a_{2,3} = 1,185$, обеспечило преимущественное влияние расстояния $\Delta_{2,3}$ на критерий J . В результате в пространстве координат (y_1, y_2) анализируемые группы объектов ($\omega_1, \omega_2, \omega_3$) расположились относительно друг друга на расстояниях $\Delta_{1,2} = 0,36$; $\Delta_{1,3} = 0,34$; $\Delta_{2,3} = 0,30$, что обеспечило более равномерную картину распределения центров классов.

В ходе экспериментов были построены разделяющие функции, определены границы областей решений и найдены ошибки классификации. Распределение объектов в двумерном пространстве $d = 2$ было получено для традиционного подхода на основе критерия (1.25) и с использованием весовых функций (1.26). Как показал результат линейного дискриминантного анализа, при использовании весовых функций средняя ошибка классификации может быть уменьшена с 8,2 до 4,6 %, что служит показателем эффективности применения этой процедуры оптимизации.

В результате анализа объектов, попавших в зону пересечения полученных областей решений, установлено, что они спорны в плане классификации, поскольку могут быть отнесены к одному из альтернативных классов. Это касается в основном пересечения областей ω_1 и ω_2 с промежуточным классом ω_3 . Важен результат безошибочного распознавания классов ω_1 и ω_2 , гарантирующий надежное обнаружение желудочковой фибрилляции в стадии устойчивого ее проявления.

Рассмотренный метод позволяет в первую очередь сократить размерность пространства признаков с учетом наилучшего разделения имеющихся классов объектов. Такой переход в пространство меньшей размерности упрощает интерпретацию данных и процедуру построения разделяющих функций.

1.5. Непараметрические методы обучения (алгоритм персептрона)

Рассмотрим задачу нахождения линейной разделяющей границы для двух классов объектов ω_1 и ω_2 . Предположим, что для обучения имеется N объектов обоих классов. Обозначим это множество объектов через Ψ . Множество Ψ состоит из двух обучающих выборок Ψ_1 (объекты первого класса ω_1) и Ψ_2 (объекты второго класса ω_2). Будем предполагать, что существует линейная функция, разделяющая Ψ_1 и Ψ_2 (т. е. Ψ_1 и Ψ_2 линейно разделимы).

Линейное решающее правило, разделяющее в n -мерном пространстве Ψ_1 и Ψ_2 , как было показано ранее, имеет вид (1.2)

$$\sum_{k=1}^n w_k x_k + w_{n+1} = 0,$$

где $x_1, x_2, \dots, x_n$ — компоненты вектора-объекта $\mathbf{X} = (x_1, x_2, \dots, x_n)$. Добавим к ранее введенному вектору-объекту $\mathbf{X}$ $(n+1)$ -ю компоненту, равную 1, и обозначим такой *расширенный* вектор-объект через $\mathbf{Y} = (y_1, y_2, \dots, y_{n+1})$, где $y_i = x_i$ для $i = 1, \dots, n$; $y_{n+1} = 1$. Тогда обучающим выборкам Ψ_1 и Ψ_2 будут соответствовать обучающие векторы ϑ_1 и ϑ_2 — множества соответствующих расширенных векторов. Их объединение обозначим через ϑ .

В этом случае решающее правило будет иметь вид

$$\sum_{k=1}^{n+1} w_k y_k = 0,$$

или в векторной форме

$$\mathbf{W}^T \mathbf{Y} = 0.$$

Предположение о линейной делимости Ψ_1 и Ψ_2 означает, что существует вектор $\mathbf{W}$, называемый решающим весовым вектором, такой, что

$$\begin{aligned}\mathbf{W}^T \mathbf{Y} &> 0 \text{ для каждого } \mathbf{Y} \in \vartheta_1, \\ \mathbf{W}^T \mathbf{Y} &< 0 \text{ для каждого } \mathbf{Y} \in \vartheta_2.\end{aligned}$$

Если все образы из ϑ_2 умножить на -1 , то вышеприведенное условие запишется в виде

$$\mathbf{W}^T \mathbf{Y} > 0 \text{ для каждого } \mathbf{Y} \in \vartheta.$$

При таком представлении классификатор на два класса можно изобразить в виде рис. 1.21. Это схема персептрона. Обучение такого классификатора на обучающем множестве ϑ заключается в подборе таких значений компонент весового вектора $\mathbf{W} = (w_1, w_2, \dots, w_{n+1})$, чтобы при предъявлении на входе любого $\mathbf{Y} \in \vartheta_1$ на выходе появилось бы $+1$, а при появлении любого $\mathbf{Y} \in \vartheta_2$ на выходе было бы -1 .

Каждому $\mathbf{Y} \in \vartheta$ можно поставить в соответствие плоскость (плоскость объекта) в этом же пространстве, которая проходит через начало координат и перпендикулярна $\mathbf{Y}$. Ответ порогового логического блока

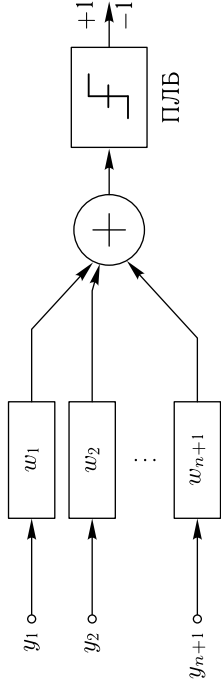


Рис. 1.21. Классификатор на 2 класса (персептрон)

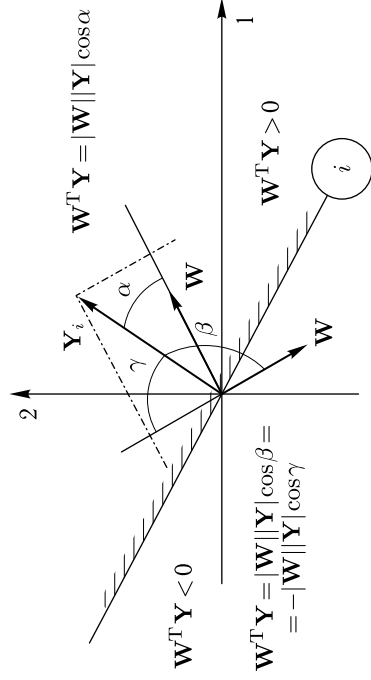


Рис. 1.22. Определение плоскости объекта для i -го объекта

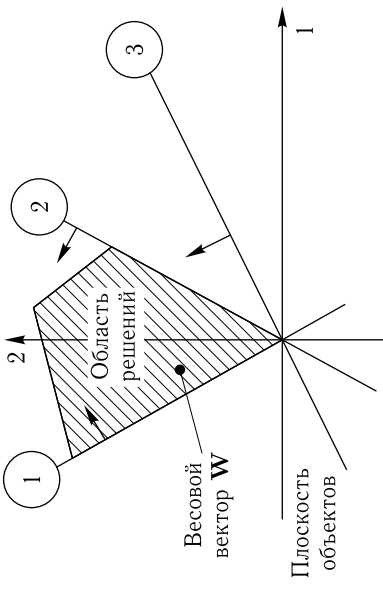


Рис. 1.23. Двумерное пространство весов с тремя плоскостями объектов

зависит от того, с какой стороны от плоскости объекта находится точка $\mathbf{W}$, как показано на рис. 1.22.

Будем считать, что при правильной классификации объектов из обучающей выборки ϑ (с учетом изменения знака на противоположный у всех объектов из ϑ_2) ответ ПЛБ должен быть всегда положительным. Тогда для N объектов $\mathbf{Y} \in \vartheta$ можно построить N плоскостей и по ним определить некоторую область решений, как показано на рис. 1.23. Для любого $\mathbf{W}$ из этой области решений для всех объектов $\mathbf{Y} \in \vartheta$ будет выполняться

$$\mathbf{W}^T \mathbf{Y} > 0.$$

Процедуры последовательного обучения с коррекцией ошибок

Можно по N обучающим объектам найти область решений и взять $\mathbf{W}$ внутри этой области. Мы рассмотрим иную процедуру последовательного обучения [1.1, 1.8, 1.17], когда обучающие объекты предъявляются последовательно, по одному, а весовой вектор $\mathbf{W}$ уточняется с каждым новым входным вектором $\mathbf{Y}$. В этом случае если после предъявления объекта $\mathbf{Y} \in \vartheta$ выполняется $\mathbf{W}^T \mathbf{Y} > 0$, то $\mathbf{W}$ остается неизменным. Если же $\mathbf{W}^T < 0$, то вычисляется новое значение $\mathbf{W}$, равное $\mathbf{W}'$, по формуле

$$\mathbf{W}' = \mathbf{W} + c\mathbf{Y},$$

где $c > 0$ называется коэффициентом коррекции. Из этого выражения видно, что новое значение весового вектора выбирается на нормали к плоскости объекта, где $\mathbf{W}^T \mathbf{Y} > 0$, как показано на рис. 1.24.

Объекты предъявляются либо циклически, либо в случайном порядке, до тех пор пока не повторится результат испытания каждого из них. Начальное значение $\mathbf{W}$ произвольно.

Рассмотрим три варианта алгоритма обучения в зависимости от процедуры выбора величины c .

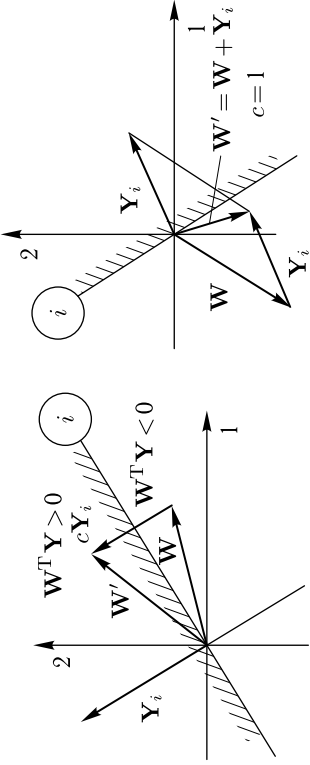


Рис. 1.24. Последовательное обучение с коррекцией ошибок

Рис. 1.25. Алгоритм с постоянным коэффициентом

Если выбирается $c = 1$, то $W' = W + Y$. Этот алгоритм с постоянным коэффициентом в зависимости от отношения величин $W^T Y$ и c исправляет или не исправляет ошибку, произошедшую на данном объекте, как показано на рис. 1.25, так как W перемещается всегда на фиксированную величину, которая может быть больше или меньше расстояния W до плоскости объекта.

Для алгоритма с абсолютной коррекцией величина c выбирается заново на каждом шаге представления объекта Y , причем значение c выбирается как наименьшее целое число, которое делает величину $W^T Y$ большей нуля, т.е. выбирается из условия

$$W^T Y = (W + cY)^T Y > 0.$$

При этом W перемещается перпендикулярно плоскости объекта на величину, чуть большую, чем расстояние от W до этой плоскости. Если Y расположен на плоскости объекта, то из вышеприведенного выражения следует, что

$$\begin{aligned} (W + cY)^T &= W^T Y + cY^T Y = 0, \\ W^T Y &= -cY^T Y, \\ |W^T Y| &= |cY^T Y| = c|Y|^2, \\ c &= \frac{|W^T Y|}{|Y|^2}. \end{aligned}$$

Поэтому c должно быть наименьшим целым числом, при этом больше, чем

$$\frac{|W^T Y|}{Y^T Y} = \frac{|W^T Y|}{|Y|^2}.$$

После корректировки весов по этому правилу на данном шаге классификатор будет верно распознавать предъявленный объект Y (рис. 1.26).

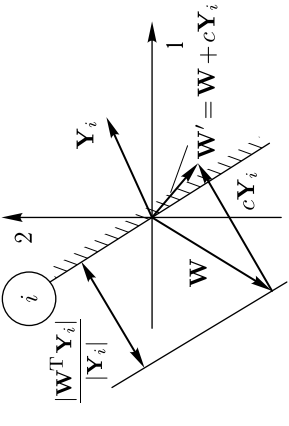


Рис. 1.26. Алгоритм с абсолютной коррекцией, где c — наименьшее целое, большее, чем $\frac{|W^T Y_i|}{|Y_i|^2}$; $|cY_i| = \frac{|W^T Y_i|}{|Y_i|^2}$

Очевидно, что этот алгоритм приводит к тем же результатам, что и предыдущий алгоритм с постоянным коэффициентом при $c = 1$, если при использовании этого алгоритма объект представляется повторно до тех пор, пока он не будет правильно распознан.

Алгоритм с частичной коррекцией перемещает весовой вектор на расстояние, пропорциональное расстоянию от точки до плоскости, т.е. величине $\frac{|W^T Y|}{|Y|}$. Если новую точку выбрать на нормали к плоскости на расстоянии, пропорциональном этой величине, то

$$|W' - W| = \lambda \frac{|W^T Y|}{|Y|},$$

где $\lambda > 0$ — коэффициент пропорциональности. Тогда

$$|Y| \cdot |W' - W| = \lambda \cdot |W^T Y|.$$

Так как векторы Y и $W' - W$ коллинеарны, то из последнего равенства следует

$$|Y^T (W' - W)| = \lambda \cdot |W^T Y|.$$

Так как $W' - W = cY$, то $|W' - W| = |cY|$ и поэтому $|Y||cY| = \lambda |W^T Y|$ и

$$\lambda \frac{|W^T Y|}{|Y|} = c|Y|,$$

откуда

$$c = \lambda \frac{|W^T Y|}{|Y|^2} = \frac{|W^T Y|}{Y^T Y}.$$

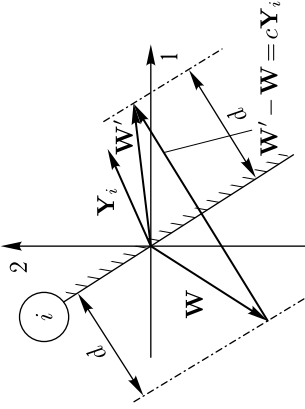


Рис. 1.27. Алгоритм с частичной коррекцией

$$(d = \frac{|W'^T Y_i|}{|Y_i|}, c = \lambda \frac{|W'^T Y_i|}{|Y_i|^2}, \lambda = 2)$$

Коэффициент λ выбирается в диапазоне $0 < \lambda < 2$. В случае $\lambda = 0$ весовой вектор вообще не изменяет положения, для $\lambda = 1$ вектор W переводится в W' , расположенный непосредственно на плоскости объекта, а при $\lambda = 2$ весовая точка переводится по другую сторону плоскости на расстояние, равное исходному расстоянию W до плоскости (рис. 1.27).

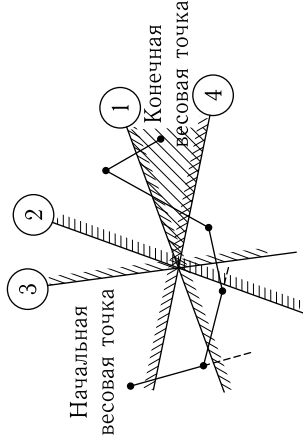


Рис. 1.28. Процесс обучения с исправлением ошибок

В общем случае алгоритм обучения с коррекцией ошибок заключается в последовательном испытании множества плоскостей объектов. Для каждой плоскости выясняется, лежит ли весовая точка по нужную сторону. Если это расположение неправильно, то весовая точка перемещается перпендикулярно плоскости по направлению к ней. Процесс продолжается до тех пор, пока не будет получен разделяющий весовой вектор (рис. 1.28). В теории доказывается [1.8], что все три алгоритма сходятся за конечное число шагов.

1.6. Построение линейных разделяющих функций с применением методов линейного программирования

Проблема нахождения разделяющей гиперплоскости, возникающая в задачах медицинской диагностики, в частности при обнаружении некоторых классов биомедицинских сигналов, может быть сформулирована и решена в виде задачи линейного программирования (ЛП) [1.18].

Определим распознаваемый объект как точку в m -мерном евклидовом пространстве признаков. Для множества объектов $(j = 1, 2, \dots)$ положение соответствующих точек задается вектором $P_j = (p_{1j}, p_{2j}, \dots, p_{mj})$, представляющим собой матрицу-столбец. Предполагается, что для обучающей последовательности известна принадлежность объектов к одному из классов ω_1 ($j \in \omega_1$) или ω_2 ($j \in \omega_2$).

Отнесение объекта P_j к одному из классов осуществляется в соответствии с правилом

$$\begin{cases} \Lambda^T P_j \geq \theta, & j \in \omega_1; \\ \Lambda^T P_j < \theta, & j \in \omega_2, \end{cases} \quad (1.30)$$

где $\theta > 0$, а вектор $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ представлен в виде матрицы-строки.

Как видно из (1.30), по знаку линейной функции

$$\Lambda^T P_j - \theta = \lambda_1 p_{1j} + \lambda_2 p_{2j} + \dots + \lambda_m p_{mj} - \theta$$

можно судить о принадлежности объекта к одному из классов.

Геометрически это означает построение разделяющей гиперплоскости $\Lambda^T P - \theta = 0$. Задачу нахождения элементов вектора Λ , удовлетворяющего системе линейных неравенств (1.30), можно решить с использованием методов ЛП.

Для этого рассмотрим задачу нахождения двух параллельных гиперплоскостей $\Lambda^T P = \bar{\lambda} + \theta$ и $\Lambda^T P = -\bar{\lambda} + \theta$, для которых выполняются следующие условия:

$$\Lambda^T P_j \geq \bar{\lambda} + \theta, \quad j \in \omega_1; \quad (1.31)$$

$$\Lambda^T P_j \leq -\bar{\lambda} + \theta, \quad j \in \omega_2; \quad (1.32)$$

$$\bar{\lambda} + \frac{\bar{\lambda}}{|\Lambda|} \rightarrow \max. \quad (1.33)$$

Таким образом, ставится задача максимизации расстояния между двумя полупространствами (1.31) и (1.32), содержащими все точки из множеств ω_1 и ω_2 соответственно. Если окажется, что искомым максимум (1.33) отрицательный, то очевидно, что множества ω_1 и ω_2 неразделимы.

Пример нахождения по критерию (1.33) вектора Λ и параллельных гиперплоскостей H_1 и H_2 , задающих области решений в одной из задач

1.7. Выравнивание символьных последовательностей

Наиболее важной задачей анализа биологических последовательностей, распознавания биомедицинских сигналов по лингвистическому описанию является их выравнивание. Для двух заданных последовательностей оно означает выполнение ряда преобразований некоторой цепочки символов, обеспечивающее совмещение и сравнение ее с другой цепочкой. При этом длины обеих цепочек становятся равными.

В молекулярной биологии выравнивание используется для анализа межнуклеотидных последовательностей в молекуле РНК, что позволяет решать задачу распознавания стабильной «вторичной» ее структуры. В задачах автоматического анализа биомедицинских сигналов применение структурных методов обработки данных позволяет представить фрагменты сигнала в виде цепочек символов конечного алфавита, а последующее выравнивание становится важным этапом процедуры классификации биосигналов.

Рассмотрим задачу выравнивания двух упорядоченных последовательностей символов T_1 и T_2 . Вначале паре сравниваемых цепочек сопоставим сеть $G(T_1, T_2)$, которая строится следующим образом [1.20]. Составим прямоугольную сетку, вертикальные линии которой соответствуют символам цепочки T_1 , а горизонтальные — T_2 . С каждой ячейкой, представляющей собой квадрат с ребром единичной длины, свяжем пару индексов (i, j) , где i — порядковый номер символа в T_1 , а j — в цепочке T_2 . Выделим в сети только те ячейки, для которых символы, стоящие на горизонтальной и вертикальной осях, совпадают. В каждой такой ячейке проведем диагональ, связывающую ее левую верхнюю и правую нижнюю вершины. Все вертикальные и диагональные ребра ориентируем стрелкой вниз, а горизонтальные — стрелкой слева направо. Стрелки указывают на все возможные перемещения по узлам сетки.

Множество вершин полученной решетки есть множество вершин графа сети $G(T_1, T_2)$, число их равно $q = (m+1)(n+1)$, где m — число символов в T_1 , n — число символов в T_2 . Множество ориентированных ребер — множество дуг этого графа. Истоком S_0 сети является левая верхняя вершина, а стоком S_k — правая нижняя вершина.

Пусть $T_1 = \langle a a b b c a a \rangle$ и $T_2 = \langle a b c b \rangle$, тогда соответствующая этой паре цепочек сеть $G(T_1, T_2)$ примет вид, приведенный на рис. 1.30. Каждый путь S , ведущий из S_0 в S_k , порождает цепочку элементарных преобразований, переводящих T_1 в T_2 . При этом движение по левой вертикали ячейки (i, j) означает вставку в соответствующую позицию цепочки T_1 j -го символа цепочки T_2 . Движение по верхней горизонтальной означает удаление i -го символа в T_1 . Движение по диагонали означает сохранение i -го символа.

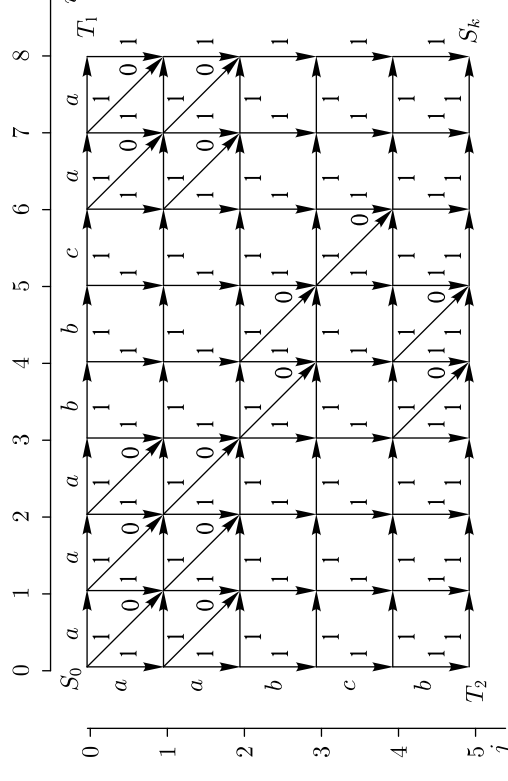


Рис. 1.30. Результат построения сети для двух сравниваемых цепочек T_1 и T_2

Для сети $G(T_1, T_2)$ зададим систему весов: каждой вертикальной и горизонтальной дуге определим вес, равный 1, а каждой диагональной дуге — вес, равный 0. Длину пути $r(S)$ будем вычислять как сумму весов дуг, лежащих на этом пути.

Представленная в виде ориентированного графа сеть (рис. 1.30) показывает множество путей $\Phi(T_1, T_2)$ возможных трансформаций T_1 в T_2 . Длина каждого пути равна числу необходимых при этом элементарных преобразований и, следовательно, значению меры их различия. Наибольший интерес представляют наиболее короткие пути преобразований. Именно они позволяют оценить степень сходства анализируемых цепочек. В качестве меры сходства цепочек T_1 и T_2 принимают длину кратчайшего из множества путей $\Phi(T_1, T_2)$:

$$r^*(T_1, T_2) = \min_{S \in \Phi(T_1, T_2)} r(S).$$

Задача нахождения кратчайшего пути может быть решена с использованием метода динамического программирования (ДП).

Прежде чем описать процедуру нахождения наилучшей траектории, дадим формальное определение оптимального выравнивания упорядоченных последовательностей символов. Под выравниванием $\tilde{U} = u_1, u_2, \dots, u_m$ и $\tilde{V} = v_1, v_2, \dots, v_n$ будем понимать последовательность элементарных преобразований, переводящих их в пару цепочек $\tilde{U}$ и $\tilde{V}$ длиной $l \geq \max(m, n)$, в которых нет пустых символов в одинаковых позициях. При этом используются следующие операции редактирования: замена, удаление и вставка символа.

Пусть для любой пары символов α, β определена стоимость $\delta(\alpha \rightarrow \beta)$ замены α на β . Замена $\alpha \rightarrow \varepsilon$ соответствует удалению символа α , а $\varepsilon \rightarrow \beta$ — вставке символа β . Стоимостью $W(\tilde{U}, \tilde{V})$ выравнивания будем называть величину

$$W(\tilde{U}, \tilde{V}) = \sum_{i=1}^l \delta(\tilde{u}_i \rightarrow \tilde{v}_i).$$

Тогда каждая последовательность преобразований получает свою цену. Выравнивание называют *оптимальным*, если его стоимость *минимальна*.

Суть рассматриваемой процедуры заключена в вычислении стоимости оптимального выравнивания между всевозможными начальными фрагментами (префиксами) двух сравниваемых слов U и V . При этом задача решается в порядке возрастания длин фрагментов.

Выравнивание префиксов может быть выполнено одним из трех способов:

- 1) сопоставлением префиксов длины $(i-1)$ и $(j-1)$, а также символов u_i и v_j , что соответствует оставлению символов u_i и v_j в анализируемых последовательностях;
- 2) сопоставлением префиксов длины $(i-1)$ и j , а также символов u_i и ε , что соответствует удалению символа u_i ;
- 3) сопоставлением префиксов длины i и $(j-1)$, а также символов ε и v_j , что соответствует вставке символа v_j .

Алгоритм расчета оптимального выравнивания основан на вычислении стоимости редактирования префиксов U и V длины i и j с использованием значений уже известной стоимости, рассчитанной для более коротких префиксов. Обозначим через $W_\delta(i, j)$, $0 \leq i \leq m$, $0 \leq j \leq n$, стоимость оптимального выравнивания $(u_1, u_2, \dots, u_i)$ и $(v_1, v_2, \dots, v_j)$. Тогда на каждом шаге продвижения по сети можно использовать следующие рекуррентные соотношения:

$$W_\delta(0, 0) = 0; \quad (1.43)$$

$$W_\delta(i+1, 0) = W_\delta(i, 0) + \delta(u_{i+1} \rightarrow \varepsilon); \quad (1.44)$$

$$W_\delta(0, j+1) = W_\delta(0, j) + \delta(\varepsilon \rightarrow v_{j+1}); \quad (1.45)$$

$$W_\delta(i+1, j+1) = \min \begin{cases} W_\delta(i, j+1) + \delta(u_{i+1} \rightarrow \varepsilon), \\ W_\delta(i+1, j) + \delta(\varepsilon \rightarrow v_{j+1}), \\ W_\delta(i, j) + \delta(u_{i+1} \rightarrow v_{j+1}). \end{cases} \quad (1.46)$$

Здесь $W_\delta(i, j+1)$, $W_\delta(i+1, j)$, $W_\delta(i, j)$ — стоимости оптимального выравнивания для пар последовательностей меньшей длины. Оно выполнено на предыдущих шагах рассматриваемой процедуры.

В процессе последовательного продвижения по узлам сети в направлении от истока к стоку на каждом шаге производится *словная оптимизация* управления с использованием стоимости замены $\delta(u_{i+1} \rightarrow v_{j+1})$, вставки $\delta(\varepsilon \rightarrow v_{j+1})$ или удаления $\delta(u_{i+1} \rightarrow \varepsilon)$

последнего символа анализируемого фрагмента слова U . Как следует из (1.43), стоимость выравнивания пустых слов задается равной нулю. Выражение (1.44) оценивает суммарные затраты при движении по узлам верхней горизонтальной сети. Выражение (1.45) связано с оценкой затрат при движении по левой вертикали сети. Соотношение (1.46) используется для условной оптимизации во всех остальных узлах ориентированного графа. При этом для каждого узла сети выбирается то шаговое управление, которое обеспечивает минимум суммарных затрат.

Возвращаясь к задаче выравнивания цепочек T_1 и T_2 , следует отметить, что стоимости $\delta(u_{i+1} \rightarrow v_{j+1}) = 0$, $\delta(u_{i+1} \rightarrow \varepsilon) = 1$, $\delta(\varepsilon \rightarrow v_{j+1}) = 1$. Используя рекуррентные соотношения (1.43)–(1.46) в процедуре оптимального выравнивания T_1 и T_2 в направлении от вершины S_0 к S_k , получим ориентированный граф (рис. 1.31). Здесь для каждого узла сетки (i, j) , соответствующего промежуточному состоянию $S_{i,j}$, стрелками указаны траектории, необходимые для оптимального выравнивания соответствующих последовательностей. Кроме того, в кружках узлов сетки указаны стоимости соответствующих суммарных затрат. Так, для промежуточного узла $(i, j) = (4, 2)$ соответствующего выравниванию префиксов $T_1 - \langle a a b \rangle$ и $T_2 - \langle a a \rangle$, одна из оптимальных траекторий задается последовательностью вершин графа $\{S_{0,0}, S_{1,0}, S_{2,1}, S_{3,2}, S_{4,2}\}$. При этом в цепочке $\langle a a b \rangle$ вначале удаляется символ a , последующие два символа сохраняются, а последний символ b также подлежит удалению. Напротив, в цепочке $\langle a a \rangle$

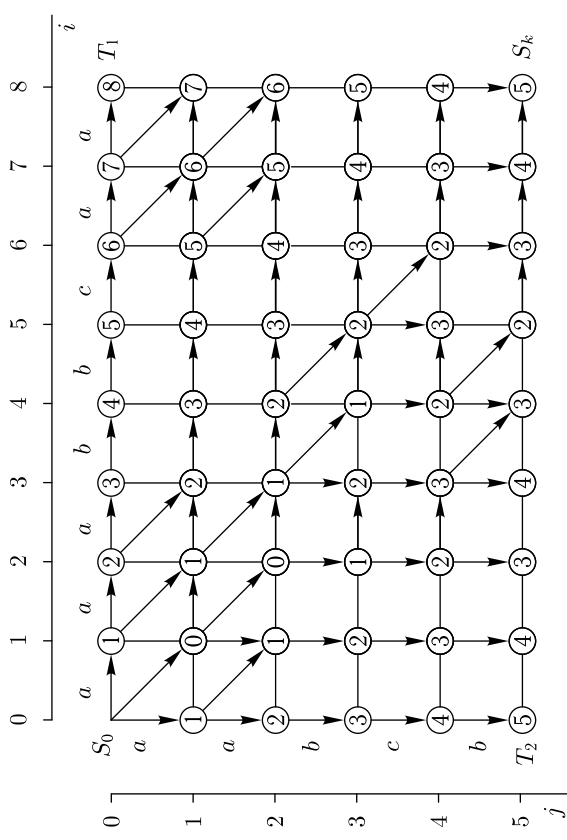


Рис. 1.31. Результат условной оптимизации выравнивания цепочек T_1 и T_2

вначале добавляется символ a , последующие два символа сохраняются, а в конце цепочки вставлен символ b . В результате соответствующие префиксы принимают вид $\tilde{T}_1 = (-, a, a, -)$ и $\tilde{T}_2 = (a, a, a, b)$. Здесь типе обозначают пустые символы, а запятые служат разделителями.

Суммарные затраты такого преобразования составляют 2 единицы, что отмечено в кружке конечной для префикса вершины $S_{4,2}$. Как следует из рисунка, оптимальная траектория для сравниваемых префиксов оказалась не единственной, но все эквивалентные пути требуют одинаковых затрат.

Затем в соответствии со схемой решения задач ДП строится оптимальная траектория из S_0 в S_k . Она обеспечивает выравнивание полных последовательностей T_1 и T_2 с минимальными затратами. Для этого нужно построить непрерывную цепочку стрелок перехода из начального состояния в конечное. Результат такой *безусловной оптимизации* показан на рис. 1.32.

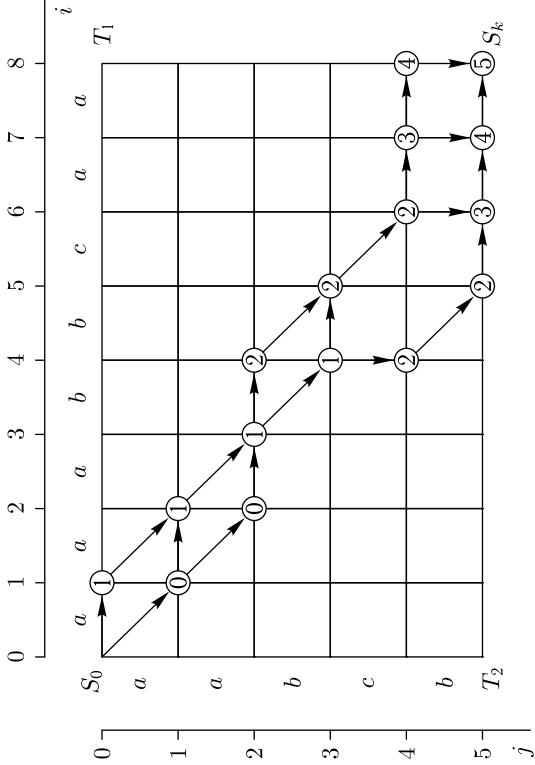


Рис. 1.32. Результат безусловной оптимизации выравнивания цепочек T_1 и T_2

Как следует из рисунка, возможно целое множество альтернативных траекторий выравнивания слов T_1 и T_2 , суммарные затраты при этом $W(\tilde{T}_1, \tilde{T}_2) = 5$. Эта величина определяет число элементарных преобразований, переводящих цепочку T_1 в T_2 .

Выделим два способа выравнивания, которые соответствуют верхней и нижней траекториям на сети $G(T_1, T_2)$. Результаты такого редактирования приведены в табл. 1.1, а также проиллюстрированы на рис. 1.33.

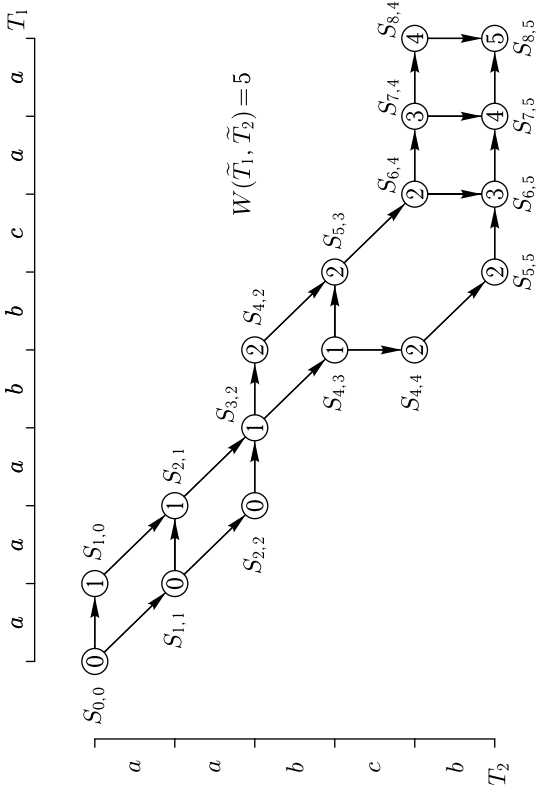


Рис. 1.33. Множество оптимальных траекторий перехода из S_0 в S_k

Таблица 1.1

Путь из S_0 в S_k	Траектория	Результат выравнивания $T_1 = \langle a a a b c a a \rangle$ и $T_2 = \langle a a b c b \rangle$
Верхний путь	$S_{0,0}; S_{1,0}; S_{2,1}; S_{3,2}; S_{4,2}; S_{5,3}; S_{6,4}; S_{7,4}; S_{8,4}; S_{8,5}$	$\tilde{T}_1 = (-, a, a, -, b, c, -, -, b)$ $\tilde{T}_2 = (a, a, a, b, b, c, a, a, -)$
Нижний путь	$S_{0,0}; S_{1,1}; S_{2,2}; S_{3,2}; S_{4,3}; S_{4,4}; S_{5,5}; S_{6,5}; S_{7,5}; S_{8,5}$	$\tilde{T}_1 = (a, a, -, b, c, b, -, -, -)$ $\tilde{T}_2 = (a, a, a, b, -, b, c, a, a)$

Во втором столбце таблицы указана вся последовательность промежуточных шагов, которая приводит к оптимальному выравниванию. Третий столбец содержит упорядоченные последовательности символов, полученные в результате преобразования $T_1 = \langle a a a b c a a \rangle$ и $T_2 = \langle a b c b \rangle$.

Выравнивание цепочки T_1 по верхней траектории ориентированного графа включает следующую последовательность шагов (рис. 1.33):

1) удаление символа a :

$$(a \rightarrow \varepsilon); S_{0,0} \rightarrow S_{1,0}; \tilde{T}_1 = (-); W(\tilde{T}_1, \tilde{T}_2) = 1;$$

2) сохранение символа a :

$$(a \rightarrow a); S_{1,0} \rightarrow S_{2,1}; \tilde{T}_1 = (-, a); W(\tilde{T}_1, \tilde{T}_2) = 1;$$

3) сохранение символа a :

$$(a \rightarrow a); S_{2,1} \rightarrow S_{3,2}; \tilde{T}_1 = (-, a, a); W(\tilde{T}_1, \tilde{T}_2) = 1;$$

4) удаление символа b :

$$(b \rightarrow \varepsilon); S_{3,2} \rightarrow S_{4,2}; \tilde{T}_1 = (-, a, a, -); W(\tilde{T}_1, \tilde{T}_2) = 2;$$

5) сохранение символа b :

$$(b \rightarrow b); S_{4,2} \rightarrow S_{5,3}; \tilde{T}_1 = (-, a, a, -, b); W(\tilde{T}_1, \tilde{T}_2) = 2;$$

6) сохранение символа c :

$$(c \rightarrow c); S_{5,3} \rightarrow S_{6,4}; \tilde{T}_1 = (-, a, a, -, b, c); W(\tilde{T}_1, \tilde{T}_2) = 2;$$

7) удаление символа a :

$$(a \rightarrow \varepsilon); S_{6,4} \rightarrow S_{7,4}; \tilde{T}_1 = (-, a, a, -, b, c, -); W = (\tilde{T}_1, \tilde{T}_2) = 3;$$

8) удаление символа a :

$$(a \rightarrow \varepsilon); S_{7,4} \rightarrow S_{8,4}; \tilde{T}_1 = (-, a, a, -, b, c, -, -); W = (\tilde{T}_1, \tilde{T}_2) = 4;$$

9) вставка символа b :

$$(\varepsilon \rightarrow b); S_{8,4} \rightarrow S_{8,5}; \tilde{T}_1 = (-, a, a, -, b, c, -, -, b), W = (\tilde{T}_1, \tilde{T}_2) = 5.$$

Выравнивание T_2 строится по той же траектории. Отличия заключаются лишь в том, что движение по верхней горизонтальной ячейки (i, j) означает вставку символа, стоящего в T_1 на i -м месте, а движение по левой вертикали — удаление j -го символа в T_2 .

Как видно из табл. 1.1, длина $\tilde{T}_1$ и $\tilde{T}_2$ в результате трансформации цепочек T_1 и T_2 стала одинаковой ($l = 9$), а пустые символы стоят в них в разных позициях.

Процедура временного выравнивания и оценка $W(\tilde{T}_1, \tilde{T}_2)$ могут быть использованы в задачах классификации биомедицинских сигналов по их символьному описанию. В том случае когда цепочка T_2 представляет собой эталонную последовательность символов, которая описывает некоторый класс сигналов, величина $W(\tilde{T}_1, \tilde{T}_2)$ отождествляется с расстоянием от анализируемой цепочки T_1 до этого эталона. При наличии нескольких классов сигналов, каждый из которых представлен своим эталоном, по этой величине может быть оценена мера принадлежности данной последовательности каждому из классов.

1.8. Медицинская диагностика на основе теории статистических решений

Постановка диагноза врачом — результат сложного процесса осмысления и сопоставления большого количества информации. Достоверность диагноза зависит от опыта, накопленного специалистом в данной

области медицины. Результаты диагностического процесса зависят также от личных качеств врача. Методы автоматической диагностики свободны от многих свойственных человеку недостатков, но уступают человеку в области творчества, интуиции. В настоящий момент для постановки диагноза с помощью компьютера применяются различные методы: вероятностные, обучения распознаванию, математической логики и др.

Вероятностные алгоритмы. Метод Байеса

В клинической практике известны симптомы, наличие которых однозначно определяет заболевание. С другой стороны, встречаются симптомы, включающие тот или иной диагноз. Однако подобная детерминированность представляет скорее исключение, чем правило. Чаще всего кардинальных, определяющих клинику симптомов нет и любой из них может встречаться с некоторой частотой при различных заболеваниях. Поэтому естественно использовать вероятностные методы для постановки диагноза.

Пусть необходимо проводить дифференциальную диагностику между заболеваниями $D_1, D_2, \dots, D_n$. Для каждого из них характерно распределение условных вероятностей $P(S | D_j)$ появления у больного того или иного симптомокомплекса: $S = \{S_1, \dots, S_k, \dots, S_n\}$, где S_i — возможные значения (градации) различных симптомов. Если бы эти распределения, а также априорные вероятности заболеваний $P(D_j)$ были заданы, то задача дифференциальной диагностики свелась бы к статистической задаче выбора гипотез, оптимальное диагностическое правило для которой нетрудно построить с помощью известной формулы Байеса. Последняя для апостериорной вероятности диагноза D_j дает выражение

$$P(D_j | S_i) = (P(D_j)P(S | D_j)) / \left(\sum_j P(D_j)P(S | D_j) \right), \quad (1.47)$$

где $P(D_j)$ — априорная вероятность заболевания с диагнозом D_j среди рассматриваемой группы болезней; $P(S | D_j)$ — вероятность появления комплекса признаков при диагнозе D_j .

Если для какого-нибудь D_j^* вероятность $P(D_j^* | S)$ значительно превосходит вероятность $P(D_j | S)$ для других $j \neq j^*$, оптимальное правило приписывает больному заболевание D_j^* . Словом «значительно превосходит» легко придать количественный смысл.

Широкому распространению такого решающего правила существенно препятствует то, что распределения $P(S | D_j)$ и вероятности $P(D_j)$ не заданы. И если вероятность заболеваний $P(D_j)$ можно с достаточной точностью оценить, вычислив на достаточно большом клиническом материале частоты, с которыми рассматриваемые болезни встречаются, то нахождение распределения $P(S | D_j)$ значительно сложнее. Последнее обстоятельство связано с тем, что для каждого диагноза

необходимо определить условную вероятность любой комбинации признаков. Но уже для 30 двоичных клинических параметров мы имеем свыше миллиарда симптомокомплексов, составленных из принятых признаков. Разумеется, собрать клинический массив для оценки условных вероятностей $P(S | D_j)$ совершенно невозможно. Единственный выход в данном случае — использование вместо $P(S | D_j)$ какой-нибудь аппроксимации, для оценки и запоминания которой не требуется столь больших масштабов. Вероятностные алгоритмы различаются, вообще говоря, способом аппроксимации $P(S | D_j)$.

Наиболее распространено предположение, что события, состоящие в появлении у больных тех или иных значений рассматриваемых нами симптомов, статистически независимы. Тогда

$$P(S | D_j) = \prod_i P(S_i | D_j).$$

Теперь для задания $P(S | D_j)$ уже нет необходимости запоминать и оценивать миллиарды чисел: достаточно иметь в памяти машины Kn чисел $P(S_i | D_j)$ и n чисел $P(D_j)$, где K — число симптомов, n — число заболеваний. Для десятков заболеваний и нескольких сотен признаков объем необходимой памяти не превысит нескольких десятков тысяч чисел, что вполне доступно для современных ЭВМ. Кроме того, $P(S_i | D_j)$ легко оценить, подсчитав частоту появления значений симптома S_i при заболевании D_j на достаточно большом количестве историй болезни.

Поскольку такой алгоритм, позволяющий получить достаточно хорошие результаты, широко распространен, остановимся на нем подробнее.

Медицинские сведения используются в виде диагностической таблицы, содержащей вероятности появления признаков для данной группы заболеваний. Представим для простоты, что диагностическая таблица составлена всего для трех заболеваний и содержит только 4 признака.

Обозначим условно диагнозы: D_1 — тетрада Фалло; D_2 — дефект межпредсердной перегородки; D_3 — незарощенный артериальный проток, и признаки: S_1 — цианоз; S_2 — усиление легочного рисунка; S_3 — акцент II тона во II межреберье слева; S_4 — правограмма (ЭКГ).

Признак S_p будет абсолютно достоверным, если при данном заболевании он встречается в 100 % случаев. Вероятность такого признака принимается за единицу. В общем случае вероятность признака при заболевании D_j равна

$$P(S_i | D_j) = n_{ij} / (n_j - n_{-ij}),$$

где n_{ij} — число больных с диагнозом D_j , имеющих признак S_i ; n_j — общее число больных с данным заболеванием; n_{-ij} — число больных с диагнозом D_j , не обследованных на признак S_i .

Величина $P(S_i | D_j)$ определяется на основании данных медицинской статистики, результатов обработки архивного материала и литературных данных, причем будет тем более достоверна, чем больше n .

Предположим [1.2, 1.10], что по данным статистики при D_1 признак S_1 встречается в 90 % случаев, S_2 не встречается (0 %), S_3 встречается в 5 %, S_4 в 60 % случаев; при D_2 признак S_1 встречается в 15 %, S_2 — в 80 % случаев; при D_3 признак S_1 встречается в 100 %, S_2 — в 95 %, S_3 — в 90 %, S_4 — в 10 % случаев.

Будем также считать, что априорная вероятность заболевания D_1 равна 35 %, D_2 — 15 % и D_3 — 50 %.

В диагностической таблице эти данные могут быть представлены, как в табл. 1.2.

Таблица 1.2

D_j	$P(D_j)$	$P(S_1 D_j)$	$P(S_2 D_j)$	$P(S_3 D_j)$	$P(S_4 D_j)$
D_1	0.35	0.90	0	0.05	0.60
D_2	0.15	0.15	0.80	0.80	0.80
D_3	0.50	0.10	0.95	0.90	0.10

Апостериорная вероятность диагноза D_j при наличии комплекса признаков S определяется по приведенной ранее формуле Байеса. Предполагая признаки независимыми, имеем

$$P(S | D_j) = P(S_1 | D_j)P(S_2 | D_j)P(S_3 | D_j)P(S_4 | D_j); \quad (1.48)$$

$$P(S) = \sum_{j=1}^3 P(D_j)P(S_1 | D_j)P(S_2 | D_j)P(S_3 | D_j)P(S_4 | D_j). \quad (1.49)$$

Вычислим вероятности диагноза сначала в том случае, когда у больного проявились все четыре признака. Тогда по формулам (1.47)–(1.49) находим вероятности диагнозов $P(D_1 | S) = 0$; $P(D_2 | S) = 0.73$; $P(D_3 | S) = 0.27$. При диагностике по методу Байеса предполагается, что у больного имеется одно из заболеваний, содержащихся в диагностической таблице. Таким образом, наиболее вероятным оказывается диагноз D_2 .

Теперь рассмотрим случай, когда у больного отсутствует признак S_1 (цианоз), но имеются все остальные признаки. Вероятность отсутствия признака S_1 равна

$$P(\bar{S}_1 | D_j) = 1 - P(S_1 | D_j).$$

Расчет проводится точно так же, но вероятность $P(S_1 | D_j)$ в формуле (1.47) заменяется на $1 - P(S_1 | D_j)$. В результате получим

$$P(D_1 | \bar{S}_1) = 0; \quad P(D_2 | \bar{S}_1) = 0.63; \quad P(D_3 | \bar{S}_1) = 0.37.$$

В табл. 1.3 приведены вероятности диагнозов при наличии всех признаков и при отсутствии одного из них.

Таблица 1.3

D_j	Признаки					
	S_1, S_2, S_3, S_4	$\overline{S}_1, S_2, S_3, S_4$	$S_1, \overline{S}_2, S_3, S_4$	$S_1, S_2, \overline{S}_3, S_4$	$S_1, S_2, S_3, \overline{S}_4$	$S_1, S_2, S_3, \overline{S}_4$
D_1	0	0	0,75	0	0	0
D_2	0,73	0,63	0,23	0,86	0,07	0,07
D_3	0,27	0,37	0,02	0,14	0,93	0,93

Из табл. 1.3 видно, что наиболее четкая картина в данном примере получается в том случае, когда у больного отсутствует признак S_4 , но имеются все остальные признаки: $P(D_3) = 93\%$.

В реальных условиях при наличии 100–200 признаков для 30–80 диагнозов подобные расчеты проводятся на ЭВМ с выдачей вероятностей каждого диагноза.

Полученные вероятности сравниваются с некоторым пороговым значением T_j , устанавливаемым в процессе обучения. Если $P(D_j) > T_j$, то делается вывод о наличии у больного диагноза (обычно $T_j > 0,9$). В случае когда $P(D_j) < T_j$, требуется проведение дополнительных обследований.

Метод Вальда (последовательный статистический анализ)

Для дифференциальной диагностики может быть использован метод Вальда. Он представляет собой последовательную процедуру обследований, при которой достигается выбранный уровень вероятности диагноза. Сущность метода состоит в следующем. Предположим, что требуется установить один из двух возможных диагнозов D_1 и D_2 . Сначала проводится обследование по признаку S_1 . Предположим далее, что при диагнозе D_1 признак S_1 имеет частоту встречаемости $P(S_1 | D_1)$, а при диагнозе D_2 — соответственно $P(S_1 | D_2)$. Если у больного отмечается наличие признака S_1 , а при диагнозе D_2 он встречается значительно чаще, чем при D_1 , то можно сделать вывод в пользу диагноза D_2 .

Таким образом, принимается диагноз D_2 , если

$$\frac{P(S_1 | D_2)}{P(S_1 | D_1)} > A,$$

где A — верхняя граница, необходимая для принятия решения. В противоположном случае, когда признак S_1 значительно чаще встречается при диагнозе D_1 :

$$\frac{P(S_1 | D_2)}{P(S_1 | D_1)} < B,$$

где B — нижняя граница отношения, принимается решение в пользу диагноза D_1 . Если же отношение вероятностей, называемое *отношением правдоподобия*, занимает промежуточное положение:

$$B < \frac{P(S_1 | D_2)}{P(S_1 | D_1)} < A,$$

то требуется провести дополнительное обследование. Тогда проводится обследование по признаку S_1 и, например при его отсутствии, вычисляется отношение

$$\frac{P(\bar{S}_2 | D_2)}{P(\bar{S}_2 | D_1)}.$$

При

$$\frac{P(S_1 | D_2)}{P(S_1 | D_1)} \frac{P(\bar{S}_2 | D_2)}{P(\bar{S}_2 | D_1)} > A$$

ставится диагноз D_2 . Подобным образом используется нижняя граница B и область неопределенности. Если признаки зависимы, то следует определять отношение правдоподобия с учетом результатов предыдущих испытаний (т.е. определять вероятность отсутствия признака S_2 при условии, что имеется признак S_1).

В последовательном анализе границы принятия решения имеют определенный физический смысл:

$$A = \frac{1 - \beta}{\alpha}; \quad B = \frac{\beta}{1 - \alpha},$$

где α — вероятность ошибки, при которой больному с диагнозом D_1 устанавливается диагноз D_2 ; β — вероятность ошибки, при которой больному с диагнозом D_2 устанавливается диагноз D_1 . Для сокращения объема обследований рекомендуется проводить сначала обследование по наиболее информативным признакам (т.е. по признакам с наибольшей диагностической ценностью). Уменьшение величины α и β приводит к уменьшению числа ошибок диагностической системы, но одновременно увеличивает число неопределенных ответов, что на практике весьма нежелательно.

Субъективная мера вероятности

На практике часто встречаются случаи, когда вероятности каких-либо событий из-за отсутствия или очень малого объема статистических данных не могут быть оценены как соответствующие частоты. В таких случаях вполне уместно применение субъективных вероятностей [1.21], т.е. оценок, данных человеком на основании его интуиции и опыта. В жизни от врачей мы порой слышим: «вероятность летального исхода для данного пациента равна 80%» или «для данного результата теста вероятность наличия данного заболевания составляет 30%». Мнения разных врачей могут различаться. Возможно привлечение группы экспертов, которые оценят значения субъективных вероятностей более точно.

Если таким субъективным образом оценить $P(D_j)$ и $P(S | D_j)$ то, используя теорему Байеса, можно вычислить $P(D_j | S)$ для разных заболеваний. При допущении независимости симптомов

$$P(D_j | S) = \frac{P(S | D_j)P(D_j)}{P(S)}.$$

1.9. Таблицы решений и блок-схемы алгоритмов в медицинской диагностике

Медицинская диагностика на основе логических систем

При попытке автоматизации диагностического процесса возникает необходимость использовать знания, накопленные врачами. В большинстве случаев эти знания выражены детерминированно (без привлечения вероятностных категорий) в виде логических категорий, что приводит к алгоритмическому представлению диагностического процесса. Такие диагностические алгоритмы могут быть описаны разными способами. К наиболее часто используемым описаниям алгоритмов относятся формулы алгебры логики, блок-схемы, деревья решений, таблицы решений и список последовательно выполняемых команд. Все эти представления алгоритмов эквивалентны, поэтому из любого описания алгоритма можно получить любое другое.

Для описания диагностической процедуры часто используют метод таблиц решений (ТР). Эти таблицы являются идеальными структурами для описания многочисленных решающих правил. Основная структура ТР (табл. 1.4) включает матрицу условий, состоящую из знаков «Д» (Да), «Н» (Нет) и «—» (Безразлично), и матрицу действий, состоящую из знаков «×» и «—». Знаки матрицы условий соответствуют ограничениям на условия, формируемые слева в этой же строке, а матрица действий с помощью знака «×» указывает, какое или какие действия надлежит выполнять, если выполнены все условия данного правила (с учетом сформулированных ограничений).

Например, если для двух условий и трех действий вектор-столбец какого-либо правила имеет вид

Д	—	Н	—	×	×
---	---	---	---	---	---

то это правило формулируется так: «Если выполняется условие 1 и не выполняется условие 3, то нужно выполнить действие 2, а затем действие 3».

Проверка правил может проходить независимо, а может иметь последовательный характер: правило 1, правило 2 и т. д. При последовательном характере выполнения правил, если ни одно условие данного правила не выполняется, переходят к следующему правилу и так далее, пока не будет найдено первое выполняющееся правило. Тогда проверка

Таблица 1.4

	Наименование таблицы	Правило 1	Правило 2	...	Правило M
ЕСЛИ	Условие 1				
и	Условие 2				
и					
:	...				
и	Условие				
ТОГДА	Действие 1				
и	Действие 2				
и					
:	...				
и	Действие K				

правил заканчивается и переходят к выполнению действий, соответствующих этому выполняющемуся правилу.

Как примеры таблиц решений (с последовательным характером проверки правил), применяющихся в системе автоматической интерпретации ЭКГ, рассмотрим табл. 1.5 и 1.6 [1.23].

Они являются частью решающих правил для диагностики гипертрофии желудочков, которая в свою очередь — часть процедуры контурного анализа ЭКГ. Входом в программу контурного анализа ЭКГ служат выходные данные программ распознавания зубцов и анализа аритмий, Р-зубца и электрической оси сердца. Данные о наличии синдрома WPW и блокады ножек пучка Гиса (RBVB и LBVB) получают на предыдущих этапах контурного анализа. Допустим, что на данном этапе анализа данные о пациенте имеют вид, указанный ниже.

WPW. Нет
RBVB. Нет
Возраст ≥ 40 лет. Да
RDv1 ≥ 40 мс. Нет
Отклонение оси QRS влево на 80°. Нет
RA/SAv1 ≥ 1. Нет
LBVB. Нет

Индекс v1 означает, что параметры RA, RD и SA получены в отведении v1. Непосредственной проверкой можно установить, что полученный вектор данных пациента (Н, Н, Д, Н, Н, Н) не удовлетворяет правилу 1, но удовлетворяет правилу 2 таблицы ГЖ, часть I (табл. 1.5). Согласно этому правилу осуществляется переход к таблице решений ГЖ, часть II (табл. 1.6). Для этой второй таблицы строится свой вектор

Таблица 1.5

ЭКГ-диагностика	Гипертрофия желудочков (ГЖ), часть I	Номер правила				
		1	2	3	4	Иначе
Синдром WPW (WPW)		Н	Н	Н	—	—
Блокада правой ножки пучка Гиса (RBBB)		Н	Н	—	—	—
Возраст ≥ 40 лет		Д	—	—	—	—
RDv1 ≥ 40 мс		Д	—	—	—	—
Отклонение оси QRS влево на 80°		Д	—	—	—	—
RA/SAv1 $\geq 1,0$		Д	—	—	—	—
Блокада левой ножки пучка Гиса (LBBB)		—	—	Н	Д	—
	(Возможно) гипертрофия правого желудочка или перенесенный инфаркт	×	—	—	—	—
	Перейти к таблице ГЖ, часть II	—	×	—	—	—
	Перейти к таблице ГЖ, часть III	×	—	—	—	—
	Перейти к таблице ГЖ, часть IV	—	—	×	—	—
	Перейти к таблице ГЖ, часть V	—	—	—	×	—
	Перейти к таблице ГЖ, часть VI	—	—	—	—	×

Таблица 1.6

ЭКГ-диагностика	Гипертрофия желудочков (ГЖ), часть II	Номер правила					
		1	2	3	4	5	Иначе
Возраст ≥ 16 лет		Д	Д	Д	Д	Д	—
$RA/SA_{V1} \geq 1,0$		Д	—	—	—	—	—
$RA/SA'1 \geq 1,0$		—	Д	—	—	—	—
Отклонение оси вправо $\geq 110^\circ$		—	—	Д	—	—	—
Отклонение оси влево $\geq -31^\circ$		—	—	—	Н	Н	—
$RA/SA_{V5} \geq 1,0$		—	—	—	Д	—	—
$RA/SA_{V6} \geq 1,0$		—	—	—	—	Д	—
(Возможно) гипертрофия правого желудочка		×	×	×	×	×	—
Перейти к таблице ГЖ, часть III		×	×	×	×	×	×

данных. Пусть он равен (Д, Н, Н, Н, Д, Н). Тогда он удовлетворяет правилу 4, что вызывает печать сообщения «(Возможно) гипертрофия правого желудочка» и переход к таблице решений ГЖ, часть III, и т. д. В момент окончания всей программы печатается сообщение «Интерпретация должна быть проведена кардиологом».

Теория таблиц решений

Рассмотрим таблицы с простыми (ограниченными) входами, условия которых являются булевыми функциями с двумя значениями: «истина» и «ложь». Они фигурируют в условии соответственно как Д (Да) и Н (Нет). Будем считать, что проверка правил может проводиться в произвольной последовательности. Для таких таблиц число возможных исходов (число правил) равно 2^n , где n — число условий. Например, для трех условий (C1, C2, C3) число правил равно $2^3 = 8$ и матрица условий будет иметь указанный ниже вид.

C1	Н	Н	Н	Н	Д	Д	Д	Д
C2	Н	Н	Д	Д	Н	Н	Д	Д
C3	Н	Д	Н	Д	Н	Д	Н	Д
	D1	D2	D3	D4	D5	D6	D7	D8

Под матрицей условий изображены действия (диагнозы), соответствующие столбцу условий. Данная таблица может быть также представлена следующими восемью уравнениями алгебры логики:

D1 = $\overline{C1} \cdot \overline{C2} \cdot \overline{C3}$,
D2 = $\overline{C1} \cdot \overline{C2} \cdot C3$,
D3 = $\overline{C1} \cdot C2 \cdot \overline{C3}$,
D4 = $\overline{C1} \cdot C2 \cdot C3$,
D5 = $C1 \cdot \overline{C2} \cdot \overline{C3}$,
D6 = $C1 \cdot \overline{C2} \cdot C3$,
D7 = $C1 \cdot C2 \cdot \overline{C3}$,
D8 = $C1 \cdot C2 \cdot C3$.

Таблица решений с n условиями называется формально полной, если для каждой из всех 2^n ситуаций предусмотрено по одному решению правилу, поэтому вышеприведенная таблица является полной.

Таблица решений может оказаться *двусмысленной*. В случае когда одновременно могут удовлетворяться 2 правила или более, приводящие к различным действиям, таблица называется *противоречивой*. Если же одновременно удовлетворяются 2 правила или более, но они ведут к одним и тем же действиям, то такая таблица называется *избыточной*. Пример противоречивой таблицы приведен ниже.

	1	2	3
C1	Н	Н	Н
C2	Н	Н	Д
C3	Н	Н	Н
Действие 1	×		
Действие 2		×	
Действие 3			×

Далее приведен пример избыточной таблицы.

	1	2	3
C1	Н	Н	Н
C2	Н	Н	Д
C3	Н	Н	Н
Действие 1	×	×	
Действие 2	×	×	
Действие 3			×

Так как даже при небольших n число правил может оказаться значительным, для сокращения записи таблиц применяются методы сокращения их описания. Первый метод заключается в применении символа «—» для обозначения того, что значение данного условия несущественно для данного правила. Например, ниже слева изображены два столбца таблицы решений.

C1	Д	Д		C1	Д
C2	Н	Н		C2	Н
C3	Н	Д		C3	—
C4	Д	Д		C4	Д
					A1
A1					
A2	×	×		A2	×
A3				A3	
A4	×	×		A4	×

Видно, что состояние условия C3 не играет роли, так как в обоих случаях будут выполнены действия A2 и A4. Используя тире, можно слить эти два столбца в один, как показано справа.

В таблице решений с ограниченным входом каждое тире означает замену двух правил решений на одно. Если в правиле решений записано n тире, то таким одним правилом решений заменено 2^n правил решений, не содержащих тире. Таким образом, использование тире не мешает проверке полноты. Ниже представлены примеры проверки полноты.

C1	Д	Д	Н	Н	Всего возможно $2 \times 2 \times 2 \times 2 = 16$ правил.	
C2	—	Н	—	Д	В таблице представлено	
C3	Н	Д	—	—	$4 + 1 + 4 + 2 = 11$ правил.	
C4	—	Н	Д	Н	Таблица неполна.	
	4	2	4	2		
C1	Д	Н	Д	—	Н	Всего возможно $2 \times 2 \times 2 = 8$ правил.
C2	Н	—	Д	Н	Д	В таблице представлено
C3	Д	Д	—	Н	Н	$1 + 2 + 2 + 2 + 1 = 8$ правил.
	1	2	2	2	1	Таблица полна.

Второй метод сокращения заключается в использовании правила ИНАЧЕ. Ниже приведен пример использования этого правила. Правила 3–8 ведут к одному и тому же действию A3 и поэтому могут быть заменены одним правилом ИНАЧЕ. Таблица решений, использующая его, по определению полна, так как явно или неявно (в правиле ИНАЧЕ) в ней выражено любое возможное правило.

	1	2	3	4	5	6	7	8		1	2	ИНАЧЕ
C1	Н	Н	Н	Н	Д	Д	Д	Д	C1	Н	Н	—
C2	Н	Н	Д	Д	Н	Н	Д	Д	C2	Н	Н	—
C3	Н	Д	Н	Д	Н	Д	Н	Д	C3	Н	Д	—
A1	×								A1	×		
A2									A2	×		
A3									A3	×		

Третий часто применяемый способ сокращения ТР — это использование последовательного характера проверок правил. В таких таблицах проверки правил осуществляются последовательно, а переход к проверке следующего правила происходит, только если данное правило не удовлетворяется.

Рассмотрим простой пример. Дана следующая ТР:

	1	2
C1	Н	—
C2	—	Д
		A1 ×
A2		×

Если в этой таблице рассмотреть произвольную последовательность проверки правил и перейти от нее к формально полной ТР, то получим таблицу

	1	2	3	4
C1	Н	Н	Н	Д
C2	Д	Н	Д	Д
A1	×	×		
A2			×	×

В ней правила 1 и 3 совпадают, что ведет к неопределенности. Однако если допустить последовательный характер проверки правил, то мы приходим к формулам

$$A1 = \overline{C1} = \overline{C1} \cdot C2 + \overline{C1} \cdot \overline{C2},$$
$$A2 = \overline{A1} \cdot C2 = C1 \cdot C2,$$

которые дают следующую формально полную ТР:

	1	2	3
C1	Н	Н	Д
C2	Д	Н	Д
A1	×	×	
A2			×

ТР с последовательным характером проверки правил широко используются на практике, в частности для медицинской диагностики. При составлении таких таблиц легко допустить неопределенность или противоречивость. Не останавливаясь на проблемах полноты и непротиворечивости, будем считать, что неопределенных и противоречивых случаев у нас нет.

Обработка ТР методом маски

Обработку таблиц решений удобно производить формальными методами. Один из наиболее распространенных — метод маски [1.22, 1.24] заключается в вычислении по заданной ТР двух двоичных матриц: табличной матрицы Т и матрицы масок А. Матрица Т получается из матрицы условий ТР заменой символов Д на 1 и помещения нулей в остальные места. Матрица А содержит 0 вместо «—», а в остальных местах — 1. Например, для ТР, приведенной в табл. 1.7, матрицы Т и А будут иметь вид табл. 1.8 и 1.9 соответственно.

Таблица 1.7

ТР	П1	П2	П3	П4
У1	Д	Н	Д	Н
У2	Н	—	Д	Н
У3	Д	Д	—	Н

Таблица 1.8

Т	П1	П2	П3	П4
У1	1	0	1	0
У2	0	0	1	0
У3	1	1	0	0

Таблица 1.9

А	П1	П2	П3	П4
У1	1	1	1	1
У2	1	0	1	1
У3	1	1	1	1

Обработка ТР проводится последовательной проверкой возможности применения какого-либо правила к входному вектору С (вектору данных пациента), который является результатом проверки условий данной ТР и кодируется так же, как и табличная матрица Т. В начале обработки первый столбец матрицы А умножается на вектор-столбец С для приведения к нулю незначимых для первого правила составляющих С, а затем полученный вектор сравнивается с первым столбцом

матрицы Т. В случае совпадения совершается переход к соответствующим действиям; при несовпадении второй столбец матрицы А умножается на С, а результат сравнивается со вторым столбцом матрицы Т, и т.д. Допустим, что для ТР, приведенной в табл. 1.7, вектор-столбец данных пациента равен $C = (0, 1, 1)$. Первый же вектор-столбец матрицы А равен $A1 = (1, 1, 1)$. В этом случае

$$A1 \cdot C = (0, 1, 1) \neq (1, 0, 1) = T1;$$
$$A2 \cdot C = (0, 0, 1) = (0, 0, 1) = T2,$$

что означает необходимость использования правила П2.

Переход от ТР к блок-схемам алгоритмов (ТР с независимыми проверками)

Дана следующая ТР:

C1	Д	Н	Н	—
C2	—	Д	Н	—
C3	—	Д	—	—
А	В	С	З	

Выбираем произвольно для первой проверки С1 (рис. 1.34).



Рис. 1.34. Первый шаг перехода ТР к блок-схеме алгоритма

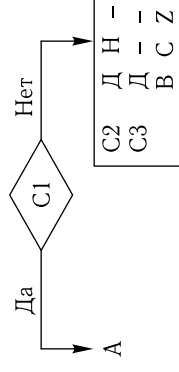


Рис. 1.35. Второй шаг перехода ТР к блок-схеме алгоритма

Рис. 1.36. Третий шаг перехода ТР к блок-схеме алгоритма

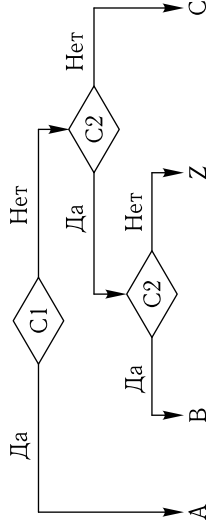


Рис. 1.37. Окончание процесса перехода ТР к блок-схеме алгоритма

Если C1 принимает значение Да, то вне зависимости от значений условий 2 и 3 должно быть выполнено действие А. Однако при C1, равном Нет, остается часть ТР, которую надо обрабатывать дальше. Эта стадия преобразований изображена на рис. 1.35.

Теперь в качестве очередной проверки в оставшейся части ТР возьмем условие C2 (рис. 1.36). Преобразование ТР завершается проверкой условия C3 в оставшейся части ТР (рис. 1.37).

Переход от ТР к блок-схемам алгоритмов (ТР с последовательной проверкой)

Рассмотрим диагностику 5 заболеваний D_1, D_2, D_3, D_4, D_5 по 7 признакам (симптомам) $S_1, S_2, S_3, S_4, S_5, S_6, S_7$. Диагностический процесс может быть неформально описан в виде 5 тестов следующим образом.

Тест 1. Вначале проверяем наличие заболевания D_1 . Диагноз D_1 ставится, если есть признаки S_3, S_4, S_5, S_6 и нет признаков S_1, S_2 .

Тест 2. При отсутствии заболевания D_1 проверяем наличие заболевания D_2 . Диагноз D_2 ставится, если нет признаков S_1, S_2 .

Тест 3. При отсутствии заболеваний D_1 и D_2 проверяем наличие заболевания D_3 . Диагноз D_3 ставится, если нет признаков S_1 и S_7 .

Тест 4. При отсутствии заболеваний D_1, D_2, D_3 проверяем наличие заболевания D_4 . Диагноз D_4 ставится, если есть признак S_7 .

Тест 5. При отсутствии заболеваний (отрицательных диагнозах) D_1, D_2, D_3, D_4 ставится диагноз D_5 .

Эта диагностическая процедура наглядно изображена в виде таблицы решений (табл. 1.10).

В столбцах этой таблицы расположены диагностические тесты и их результат. Буква Д означает Да, т. е. наличие соответствующего признака. Буква Н (Нет) означает отсутствие соответствующего признака. Проверк (—) означает, что наличие или отсутствие данного признака безразлично. В нижней строке таблицы расположены диагнозы, которые принимаются, если входные данные удовлетворяют условиям соответствующего теста. Тесты применяются строго последовательно слева направо, и процедура останавливается при выдаче первого диагноза. Первый тест (первый столбец) показывает, что если нет (Н) признаков S_1 и S_2 , но есть (Д) признаки S_3, S_4, S_5, S_6 , то принимается

Таблица 1.10

Признак	Диагностические тесты				
	1	2	3	4	5
S_1	Н	Н	Н	—	—
S_2	Н	Н	—	—	—
S_3	Д	—	—	—	—
S_4	Д	—	—	—	—
S_5	Д	—	—	—	—
S_6	Д	—	—	—	—
S_7	—	—	Н	Д	—
Диагноз	D_1	D_2	D_3	D_4	D_5

диагноз D_1 . Второй тест, применяемый при отрицательном результате первого, показывает, что если нет признаков S_1, S_2 , то принимается диагноз D_2 , и т. д.

Теперь по столбцам таблицы решений можно легко составить формулы алгебры логики, которые описывают диагностику данных заболеваний:

$$D_1 = S_3 \wedge S_4 \wedge S_5 \wedge S_6 \wedge \neg S_1 \wedge \neg S_2,$$
$$D_2 = \neg D_1 \wedge (\neg S_1 \wedge \neg S_2),$$

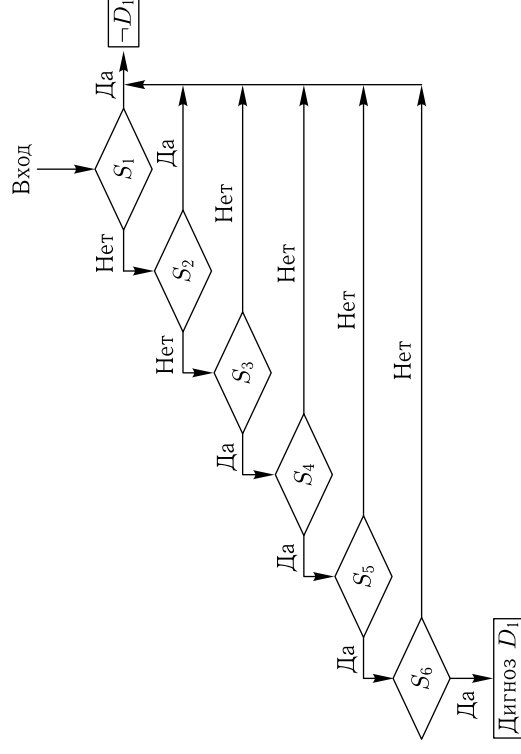


Рис. 1.38. Блок-схема алгоритма для диагностики заболевания D_1

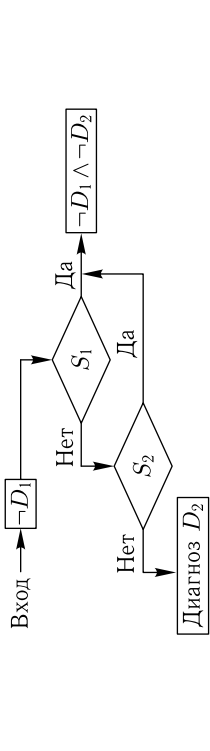


Рис. 1.39. Блок-схема для диагностики заболевания D_2

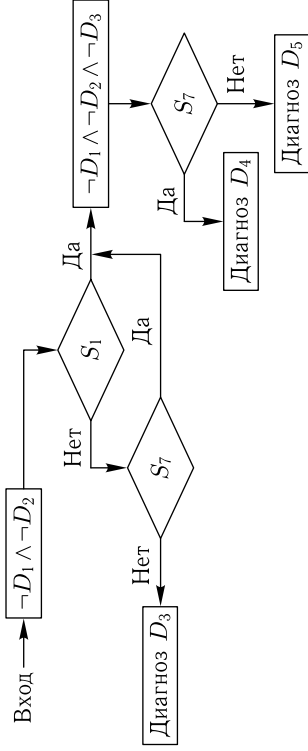


Рис. 1.40. Блок-схема алгоритма для диагностики заболеваний D_3, D_4, D_5

$$D_3 = \neg D_1 \wedge \neg D_2 \wedge (\neg S_1 \wedge \neg S_7),$$

$$D_4 = \neg D_1 \wedge \neg D_2 \wedge \neg D_3 \wedge (S_7),$$

$$D_5 = \neg D_1 \wedge \neg D_2 \wedge \neg D_3 \wedge \neg D_4.$$

Конечно, используя таблицу решений или логические выражения, легко перейти к блок-схеме алгоритма диагностики, однако она будет содержать некоторое число одинаковых проверок (задача минимизации проверок может решаться если перейти от данной ТР к ТР с независимыми проверками). Блок-схема алгоритма диагностики приведена ниже. Первое правило ТР (первый столбец) дает следующую блок-схему, показанную на рис. 1.38.

Второй столбец ТР (второе правило) изображен в виде блок-схемы, приведенной на рис. 1.39. Третий и четвертый столбцы (третье и четвертое правила) изображены в виде схемы, приведенной на рис. 1.40. Можно объединить вышеприведенные схемы в блок-схему диагностического алгоритма.

1.10. Характеристики точности и цены диагностики

Пусть мы используем бинарный тест для установления факта наличия или отсутствия некоторого заболевания в определенной популяции населения. Пусть A обозначает случай наличия у пациента определенной патологии (заболевания), а N — отсутствие у него этой

патологии (норма). Допустим, априорные вероятности $P(A)$ и $P(N)$ соответственно обозначают доли страдающих данным заболеванием пациентов и нормальных пациентов в тестовой популяции. Пусть T^+ представляет собой положительный результат теста (наличие заболевания), а T^- — отрицательный результат (отсутствие заболевания). Существуют следующие возможности.

- о *Истинно положительный* результат представляет собой случай, когда тест положителен для пациента, имеющего заболевание (известный также как *правильная диагностика* или *чувствительность*). Этот показатель определяется как $P(T^+ | A)$, или

$$S^+ = \frac{\text{число положительных результатов теста}}{\text{число субъектов с заболеванием } A}.$$

Чувствительность теста отражает его способность обнаруживать наличие искомого заболевания.

- о *Истинно отрицательный* результат представляет собой случай, когда тест отрицателен для пациента, не страдающего заболеванием. Этот показатель, называемый *специфичностью* S^- , задается как $P(T^- | N)$, или

$$S^- = \frac{\text{число отрицательных результатов теста}}{\text{число субъектов без заболевания}}.$$

Специфичность теста показывает точность определения отсутствия исследуемого заболевания.

- о *Ложно отрицательный* результат (*пропуск цели*)¹ — это случай, когда тест отрицателен для пациента, у которого имеется исследуемое заболевание; т. е. данный тест пропустил этот случай. Вероятность такой ошибки равна $\beta = P(T^- | A)$.
- о *Ложно положительный* результат (*ложная тревога*) определяется как случай, когда результат теста положителен, а пациент, который тестируется, не имеет данного заболевания. Вероятность этого типа ошибки равна $\alpha = P(T^+ | N)$.

В матрице результатов классификации (табл. 1.11) сведены все возможные результаты классификации.

Таблица 1.11

Истинный диагноз	Предсказанный диагноз	
	Отсутствие заболевания	Наличие заболевания
Отсутствие заболевания	S^-	$\alpha = P(T^+ N)$
Наличие заболевания	$\beta = P(T^- A)$	S^+

¹ Термины «пропуск цели» и «ложная тревога» взяты из задачи радиолокационного обнаружения самолета противника — одной из первых задач автоматической системы принятия решений.

Отметим, что

- $P(T^+ | A) + P(T^- | A) = 1$,
- $P(T^+ | N) + P(T^- | N) = 1$,
- $S^- = 1 - P(T^+ | N) = P(T^- | N)$
- и $S^+ = 1 - P(T^- | A) = P(T^+ | A)$.

Суммарный показатель точности может быть определен как

$$\text{точность} = S^+ P(A) + S^- P(N),$$

где $P(A)$ — доля исследуемой популяции, которая в действительности имеет данное заболевание (распространенность этого заболевания), а $P(N)$ — доля исследуемой популяции, которая в действительности не страдает этим заболеванием.

При разработке нового теста или метода диагностики для подтверждения присутствия или отсутствия некоторого заболевания необходимо использовать другой предварительно созданный метод в качестве референтного. Такой референтный метод часто называют *золотым стандартом*. При испытании метода, основанного на компьютерной обработке, чаще всего в качестве золотого стандарта используют диагностику, которая была выполнена специалистом в данной области. Для этой же цели могут использоваться хорошо опробованные лабораторные или исследовательские процедуры. Термин «истинный диагноз» в табл. 1.12 показывает результат использования золотого стандарта, а термин «предсказанный диагноз» относится к результату проведенного теста. Может быть также интересной информация о вероятности того, что пациент, у которого результат тестирования оказался положительным, в действительности имеет данное заболевание: эта информация дается условной вероятностью $P(A | T^+)$. На данный вопрос можно ответить с помощью теоремы Байеса, используя которую мы можем получить

$$P(A | T^+) = \frac{P(A)P(T^+ | A)}{P(T^+)} = \frac{P(A)P(T^+ | A) - P(N)P(T^+ | N)}{P(T^+)}.$$

Отметим, что $P(T^+ | A) = S^+$ и $P(T^+ | N) = 1 - S^-$. Для того чтобы определить значения апостериорной вероятности, необходимо знать чувствительность и специфичность теста, а также априорные вероятности отрицательных и положительных случаев (степень распространенности данного заболевания).

Для того чтобы охарактеризовать общую ценовую эффективность теста или метода диагностики, может быть определена матрица стоимости диагностики (табл. 1.12).

Стоимость получения истинно отрицательного решения характеризуется величиной C_N : это цена того, что нормальный пациент получит предсказанный диагноз об отсутствии данного заболевания. Стоимость получения истинно положительного решения показывает величина C_A : она может включать стоимость дальнейшего лечения, отслеживания и т. д., которые являются вторичными по отношению к самому тесту,

Таблица 1.12

Истинный диагноз	Предсказанный диагноз	
	Отсутствие заболевания	Наличие заболевания
Отсутствие заболевания	C_N	C^+
Наличие заболевания	C^-	C_A

но входят в программу диагностики и поддержания здоровья. Величина C^+ показывает стоимость «ложной тревоги»: она представляет собой стоимость того, что пациент, у которого нет заболевания, будет ошибочно подвергнут дальнейшей диагностике или терапии. Стоимость C^- — это цена «пропуска цели»: присутствовавшее заболевание у пациента не было диагностировано, ситуация ухудшается со временем, пациент сталкивается с дальнейшими трудностями данного заболевания и система здравоохранения или пациент должны нести расходы на дальнейшие тесты и задержанную терапию.

Коэффициент проигрыша, связанный с неправильной классификацией, может быть определен как

$$Q = \alpha \cdot C^+ + \beta \cdot C^-.$$

Тогда общая стоимость программы скрининга может быть оценена как

$$C_S = S^+ \cdot C_A + S^- \cdot C_N + \alpha \cdot C^+ + \beta \cdot C^-.$$

Рабочие характеристики

Характеристики суммарной правильной диагностики, выраженные в процентах, имеют ограниченные возможности для оценки точности какого-либо диагностического метода. Некоторое улучшение анализа может быть достигнуто за счет использования отдельных оценок правильности диагностики, таких как чувствительность и специфичность для каждой категории. К сожалению, эти характеристики не показывают зависимости результатов диагностики от решающего порога.

Было бы желательно иметь тест для скрининга или диагностики, который одновременно был бы и высокочувствительным, и высокоспецифичным. Однако в реальности такой тест обычно недостижим. Как правило, таким тестам свойственен компромисс между чувствительностью и специфичностью. Соотношение между чувствительностью и специфичностью иллюстрируется *рабочей характеристикой* (ROC-кривой, от *англ.* receiver operating characteristics, рабочие характеристики пользователя), которая позволяет осуществить более полный анализ точности классификации диагностического метода.

Рассмотрим случай, показанный на рис. 1.41. Допустим, что для данного диагностического теста с решающей переменной x мы имеем предварительно определенную функцию плотности вероятности для нормальных случаев, обозначенную как $f(x | N)$, и для аномальных

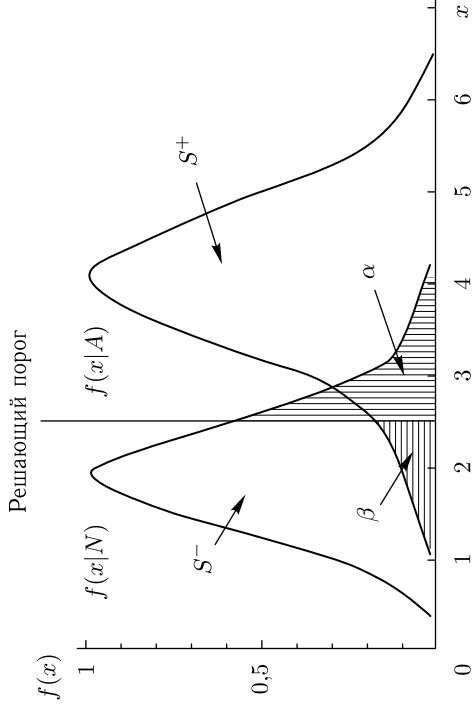


Рис. 1.41. Зависимость ошибок диагностики от решающей переменной x . Вертикальная линия показывает решающий порог. $f(x|N)$ и $f(x|A)$ — функции плотности вероятности для нормальных и аномальных случаев

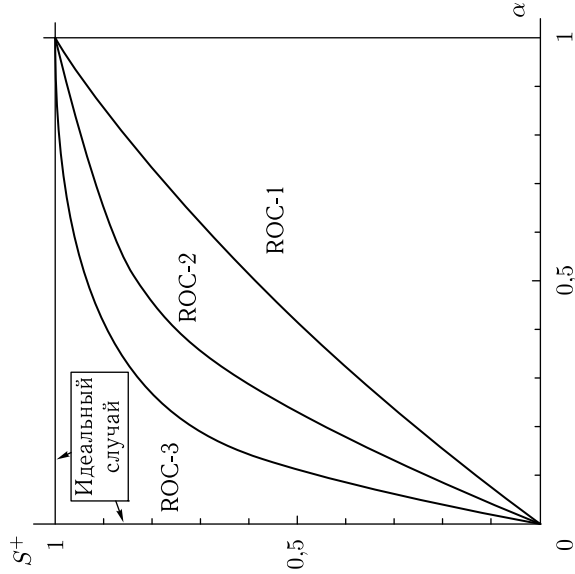


Рис. 1.42. Примеры рабочих характеристик (ROC-кривых)

случаев — обозначенную как $f(x|A)$. Как показано на рис. 1.41, эти две функции будут почти всегда перекрываться, поскольку ни один из методов не может быть совершенным. Пользователь должен так определить решающий порог (показанный вертикальной линией),

чтобы достигнуть компромисса между чувствительностью и специфичностью. Понижение решающего порога будет увеличивать S^+ за счет увеличения α . (Замечание: S^- и β могут быть легко получены из α и S^+ соответственно.)

ROC-кривая — это график, который показывает точки (α, S^+) , полученные для некоторого диапазона значений решающего порога (рис. 1.42). Меняя решающий порог, мы получаем различные решения в пределах диапазона $(0, 1)$. ROC-кривая описывает свойства обнаружения, присущие какому-либо тесту или методу: пользователь может выбрать для работы любую точку на кривой. ROC-кривая не зависит от распространенности исследуемого заболевания. Так как в принципе возможна ситуация, когда все случаи будут отмечены как отрицательные или, наоборот, как положительные, то ROC-кривая проходит через точки $(0, 0)$ и $(1, 1)$.

Суммарной характеристикой эффективности теста является площадь под ROC-кривой S_x . Из рис. 1.42 видно, что ее величина ограничена диапазоном $(0, 1)$. Тест, дающий большую площадь под ROC-кривой, означает лучший метод, чем дающий меньшую площадь: на рис. 1.42 метод, соответствующий кривой ROC-3, лучше, чем метод, соответствующий кривой ROC-2; и оба лучше, чем метод, представленный кривой ROC-1, при которой $S_x = 0.5$. Идеальный метод будет иметь ROC-кривую, которая следует по вертикальной линии от точки $(0, 0)$ до точки $(0, 1)$, далее по горизонтальной линии от точки $(0, 1)$ до точки $(1, 1)$ и имеет $S_x = 1$. Такой идеальный метод имеет $S^+ = 1$ и $\alpha = 0$. Для этого требуется, чтобы кривые, представленные на рис. 1.41, не перекрывались.

1.11. Базы данных физиологических сигналов *physionet.org*

Разработчики и производители электронной медицинской аппаратуры, в которой используются различные методы математической обработки и анализа биомедицинских сигналов, постоянно сталкиваются с проблемой получения выборок реальных данных, необходимых как для поддержки работ по созданию алгоритмов и методов анализа, так и для оценки качества разработанных методов.

Сбор достаточного объема экспериментальных данных может оказаться очень длительной и дорогостоящей работой, так как требует привлечения квалифицированных специалистов, использования специальной аппаратуры, решения ряда организационных и этических проблем. Кроме того, некоторые из видов патологий встречаются редко и получение соответствующих записей сигналов может потребовать неопределенного времени.

Для того чтобы помочь исследователям и производителям медицинской техники в решении перечисленных выше проблем, в 1999 г.

был введен в действие специализированный интернет-портал PhysioNet Resource [1.25], доступный по адресу physionet.org [1.26]. Данный интернет-ресурс финансируется двумя подразделениями американского Национального института здоровья (National Institutes of Health, NIH): Национальным институтом биомедицинских изображений и биоинженерии (National Institute of Biomedical Imaging and Bioengineering, NIBIB) и Национальным институтом общемедицинских наук (National Institute of General Medical Sciences, NIGMS).

Примерно начиная с 1970-х гг. специалисты из Массачусетского технологического института (Massachusetts Institute of Technology, MIT, Бостон, США), работая над созданием первых микропроцессорных кардиологических мониторов, осознали необходимость формирования общедоступной базы данных аннотированных записей ЭКГ, которая могла бы послужить основой для разработки, совершенствования и объективного сравнения между собой алгоритмов автоматического анализа сердечного ритма по ЭКГ. Пятилетние усилия в этом направлении завершились в 1980 году созданием базы данных аритмий (MIT-BIN Arrhythmia Database), которая за это короткое время завоевала репутацию общепринятого стандарта. Вслед за этим одна за другой начали появляться новые базы данных ЭКГ и других физиологических сигналов. К 1999 г. сотрудниками MIT был распространен комплект оптических дисков, содержащих уже 11 подобных коллекций сигналов. Создание портала PhysioNet Resource позволило не только сделать эти базы данных общедоступными, но и открыть возможность для расширения их состава всеми желающими исследователями.

Портал включает три основных компонента.

1. Архив записей сигналов PhysioBank, представляющий собой большой и постоянно расширяющийся набор тщательно аннотированных цифровых записей физиологических сигналов, временных последовательностей и различного вида сопровождающей информации. В настоящее время (2016 г.) PhysioBank содержит более 60 коллекций кардио-пульмонарных, неврологических и других сигналов как от здоровых пациентов, так и от пациентов, страдающих различными видами заболеваний. В частности, направленность данных коллекций охватывает такие социально значимые медицинские проблемы, как внезапная кардиологическая смерть, застойная сердечная недостаточность, эпилепсия, нарушения походки, апноэ сна и старение.

2. Библиотека программ PhysioToolkit является обширным и постоянно расширяющимся набором программных средств, предназначенных для обработки и анализа физиологических сигналов, обнаружения значимых с точки зрения физиологии событий, интерактивного отображения и аннотирования сигналов, создания новых баз данных, моделирования физиологических и других сигналов, количественной оценки и сравнения методов анализа сигналов, а также для анализа

нестандартных сигналов. Все программное обеспечение свободно доступно в форме исходных кодов программ на условиях общественной лицензии «GNU» (General Public License, GPL).

3. Виртуальная лаборатория PhysioNetWorks обеспечивает возможность сотрудничества с коллегами со всего мира в направлении создания, оценки, совершенствования, документирования и подготовки новых данных и программных средств для публикации на портале PhysioNet Resource. В отличие от остальных двух компонентов портала, доступ в виртуальную лабораторию защищен паролем (который легко может быть получен в течение нескольких минут). Для участников лаборатории доступны надежные средства сохранения данных, удобные программы для просмотра и аннотирования сигналов, а также возможность интерактивного взаимодействия более чем с 3000 других исследователей.

Все базы данных, размещаемые на портале PhysioNet Resource, должны быть подготовлены в соответствии со специально разработанным форматом, который предусматривает три вида файлов.

1. *Заголовочные файлы*. Представляют собой небольшие текстовые файлы, которые описывают сигналы (имя или URL файла сигнала, формат хранения, количество и типы сигналов, частота дискретизации, информация о калибровке, характеристики аналого-цифрового преобразования, продолжительность записи и начальное время).

2. *Файлы аннотаций*. Содержат наборы меток, каждая из которых описывает свойства одного или нескольких сигналов в заданные моменты времени. Метки должны следовать в файле строго в порядке нарастания времени. Информация в файлах аннотаций упакована с использованием специального формата. Поэтому доступ к ней возможен только с помощью специальных программ из библиотеки PhysioToolkit.

3. *Файлы сигналов*. Содержат в бинарном виде цифровые отсчеты сигналов, которые могут быть представлены либо непосредственно значениями отсчетов, либо в одном из предусмотренных форматов упаковки, служащих для сокращения занимаемого объема памяти. Объем файлов сигналов может быть очень велик (до десятков и сотен мегабайт).

Все три вида файлов, относящихся к одной записи сигнала, должны иметь одинаковое имя, но разные расширения. Заголовочные файлы имеют расширение «hea», файлы данных — расширение «dat», а файлы аннотаций могут иметь произвольные расширения, определяемые создателями и пользователями конкретной базы данных. Причем одной и той же записи сигнала могут соответствовать несколько видов файлов аннотаций, что позволяет осуществлять аннотирование записей с различных точек зрения.

В качестве примера рассмотрим описание одной из двухканальных записей ЭКГ из базы данных аритмий MIT-BIN Arrhythmia Database.

Эта запись в базе данных имеет имя 119. Рассмотрим содержимое заголовочного файла 119.head:

```
119 2 360 650000
119.dat 212 200 11 1024 825 2943 0 MLI
119.dat 212 200 11 1024 930 -26388 0 V1
# 51 F 1129 654 x2
# Pronestyl
# The PVCs are uniform.
```

В первой строке указано имя записи (119), число каналов сигнала (2), частота дискретизации (360 Гц) и общее число отсчетов сигнала в каждом из каналов (650 000).

Во второй и третьей строках содержится информация, относящаяся к сигналам первого и второго каналов соответственно. Например, данные для первого канала, перечисленные во второй строке заголовочного файла, имеют следующий смысл:

119.dat — имя файла, в котором содержится сигнал;

212 — обозначение используемого формата упаковки отсчетов сигнала;

200 — масштаб сигнала (число уровней квантования на 1 мВ);

11 — разрядность аналого-цифрового преобразования;

1024 — уровень нулевой линии дискретизованного сигнала;

825 — значение первого отсчета сигнала по данному каналу;

2943 — контрольная сумма, служащая для проверки ошибок;

0 — размер буфера для чтения сигнала (здесь не используется);

MLI — обозначение отведения ЭКГ (здесь — мониторное отведение II).

Последние три строки заголовочного файла содержат комментарий для данной записи ЭКГ.

В файле с именем 119.dat сигнал записан в виде последовательности 11-битных отсчетов, представляющих собой целые неотрицательные числа, которые могут принимать значения от 0 до 2047. Для пересчета сигнала к общепринятому масштабу в милливольтах необходимо от каждого значения отнять величину указанного в заголовочном файле уровня нулевой линии (1024) и разделить результат на величину масштаба сигнала (200).

На рис. 1.43 показаны графики первых 10 секунд обоих каналов ЭКГ. Масштабы по времени и по ординате приведены к исходным для ЭКГ значениям. На нижнем графике жирными точками отмечены позиции меток верификации, над которыми надписаны условные идентификаторы, раскрывающие смысл каждой из отметок. Эти отметки содержатся в файле аннотаций, который для данной записи ЭКГ имеет имя 119.atg.

Ниже приведены данные из файла аннотаций, относящиеся к показанному на рисунке фрагменту сигнала:

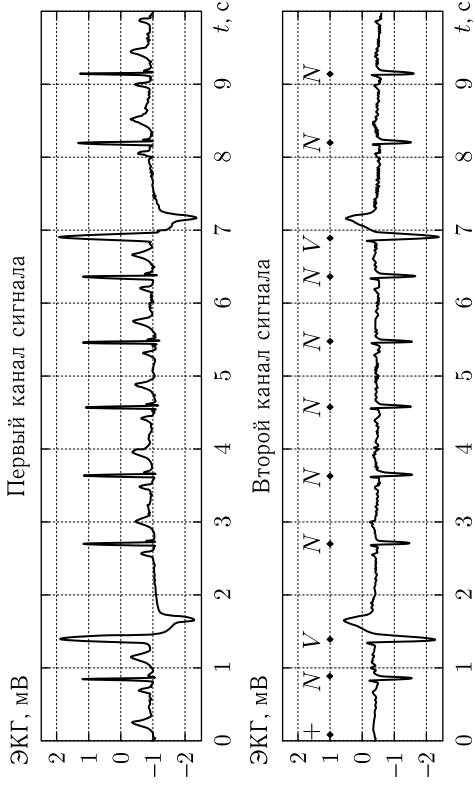


Рис. 1.43. Графики первых 10 секунд двух каналов ЭКГ записи «119» с отметками из файла аннотаций

В первом поле каждой строки указана позиция отмеченной точки сигнала, выраженная в отсчетах (для перехода к времени в секундах это значение надо разделить на частоту дискретизации 360 Гц). Во втором поле каждой строки содержится идентификатор, раскрывающий смысл данной отметки времени. В нашем случае мы имеем три варианта таких идентификаторов:

- + — признак смены ритма (в самом начале записи ставится всегда);
- N — QRS-комплекс фонового ритма (относящийся к «норме»);
- V — желудочковый комплекс (желудочковая экстрасистола).

Отметим, что приведенный пример — очень упрощенная иллюстрация. В действительности предлагаемый формат представления данных намного сложнее и обеспечивает пользователю богатый и гибкий набор возможностей. Подробное описание как форматов данных, так и специализированных программ для манипуляций с данными содержится в инструкциях, доступных на портале PhysioNet Resource.

Одна из наиболее важных возможностей, предлагаемых порталом PhysioNet Resource, — набор функций, позволяющих выполнять сравнение между собой различных алгоритмов обработки и анализа физиологических сигналов. Эта возможность поддерживается специализированным программным обеспечением, свободно доступным на данном сайте. Сравнение осуществляется путем сопоставления данных, содержащихся в двух файлах аннотаций, из которых один обычно содержит данные верификации (то есть служит эталоном), а другой — выраженные в тех же терминах и представленные в том же формате результаты работы некоего испытываемого алгоритма. Программы реализуют сравнение информации из пары сравниваемых файлов аннотаций, подсчет ошибок и формирование итоговых таблиц, в которых имеются как сводные результаты по всему используемому набору записей, так и подробная расшифровка для каждой из записей. Анализ полученных результатов позволяет как получать объективные стандартизованные оценки качества алгоритмов, так и осуществлять целенаправленную работу по совершенствованию программно-алгоритмического обеспечения систем анализа биомедицинских сигналов.

Наиболее часто решаемая задача в медицинских системах обработки информации — выявление тех или иных патологий (например, экстрасистол в случае анализа сердечных аритмий). В этом случае для оценки качества алгоритмов используются методики и показатели, описанные в разделе 1.10 настоящей монографии.

Глава 2

МЕТОДЫ АНАЛИЗА БИОМЕДИЦИНСКИХ СИГНАЛОВ

2.1. Дискретизация и квантование сигналов

Частотный подход к анализу сигналов

Наиболее распространенным подходом к обработке и анализу сигналов является *частотный подход*, при котором предполагается, что рассматриваемый процесс может быть представлен в виде суммы синусоид с различными частотами. Теоретической базой для частотного подхода служит теория преобразований и рядов Фурье [2.19].

Синусоида полностью определяется формулой $x(t) = A \sin(2\pi ft + \theta)$, где A — амплитуда; f — частота; θ — фазовый сдвиг (рис. 2.1).

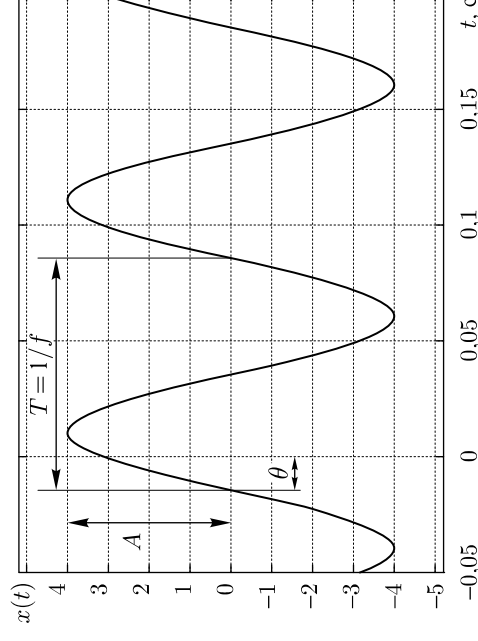


Рис. 2.1. Синусоида и ее параметры

С точки зрения частотного подхода, процедуры *цифровой фильтрации* (см. 2.2) решают задачу целенаправленного изменения амплитуды и фазы каждой из частотных составляющих в соответствии с заданной передаточной функцией. В свою очередь, задача *спектрального*

анализа (см. 2.3) состоит в количественной оценке параметров синусоид для всего диапазона исследуемых частот.

Дискретное представление непрерывного сигнала и теорема отсчетов

Для обработки и анализа непрерывных сигналов необходимо их преобразование в цифровую форму, которое выполняется при помощи аналого-цифрового преобразователя и может рассматриваться как сочетание двух базовых процедур: *дискретизации* и *квантования*.

Дискретизация — взятие измерений (*отсчетов*) в дискретные моменты времени.

Квантование — присваивание цифровому отсчету значения, соответствующего некоторому фиксированному уровню сигнала.

Компьютерная обработка возможна только для данных, представленных в форме чисел. В случае обработки сигналов это, как правило, последовательность отсчетов, полученная в результате *равномерной дискретизации*. Равномерная дискретизация обычно выполняется с использованием *аналого-цифрового преобразователя* (АЦП) и заключается в измерении сигнала через определенный интервал времени.

Пусть $x(t)$ — некоторый непрерывный сигнал. Здесь t — непрерывный аргумент, под которым подразумевается время, хотя в общем случае аргументом может быть и какая-нибудь другая величина. Тогда $x(n) = x(nT)$ представляет собой результат равномерной дискретизации непрерывного сигнала $x(t)$. Здесь T — *интервал дискретизации*, а n — порядковый номер отсчета.

На рис. 2.2 приведена иллюстрация процесса равномерной дискретизации. В данном случае интервал дискретизации $T = 10 \text{ мс} = 0,01 \text{ с}$.

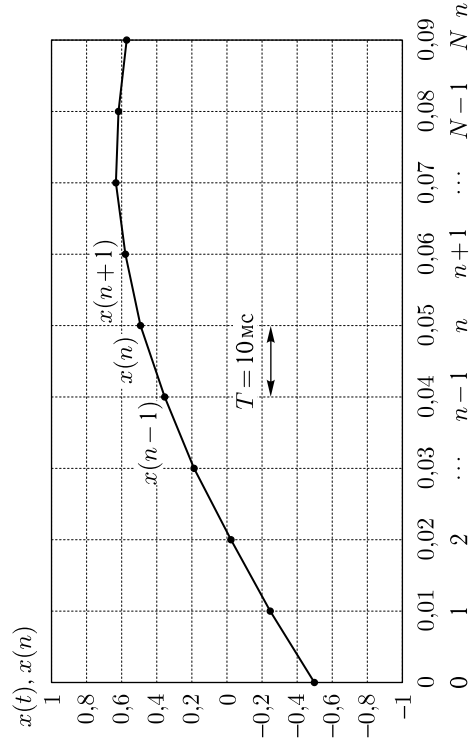


Рис. 2.2. Процесс равномерной дискретизации сигнала

Величина, обратная интервалу дискретизации ($f_d = 1/T$), называется *частотой дискретизации* (в приведенном примере $f_d = 1/T = 100 \text{ Гц}$).

Очевидно, что в дискретном сигнале $x(n)$, полученном из непрерывного сигнала $x(t)$, потеряна информация о значениях, лежащих в промежутках между точками взятия отсчетов. В этом случае возникает вопрос: насколько правомерно рассмотрение дискретного сигнала вместо непрерывного? Теоретический ответ на этот вопрос дает теорема отсчетов [2.1, 2.26]. Она упоминается в литературе как теорема Котельникова, теорема УКШ (Уиттекера–Котельникова–Шеннона) и теорема Найквиста. Это связано с тем, что ее формулировка и доказательства были даны независимо друг от друга несколькими учеными.

Формулировка теоремы отсчетов: *если функция, зависящая от времени, имеет ограниченный частотой F спектр, то она полностью определяется дискретными отсчетами ее значений, следующими с частотой $f_d = 2F$, и может быть точно реконструирована с помощью выражения*

$$x(t) = \sum_{n=-\infty}^{\infty} x(n) \frac{\sin[\pi(t - nT)/T]}{\pi(t - nT)/T}.$$

Отметим, что утверждение о возможности полной реконструкции функции — чисто теоретическое, так как основано оно на предположении о бесконечном числе отсчетов дискретного сигнала $x(n)$, что на практике неосуществимо. Тем не менее теорема отсчетов является одной из основ всей теории дискретной обработки сигналов.

Нарушение условий теоремы отсчетов проявляется в эффекте *наложения частот*, который состоит в том, что при выборе недостаточной высокой частоты дискретизации некоторые частотные составляющие становятся неразличимыми.

Классической иллюстрацией эффекта наложения служит восприятие человеком вращения колеса автомобиля, заснятого на кинолентку. При трогании автомобиля с места сначала наблюдается ожидаемое ускорение движения колес. Затем с набором скорости движение колес кажется все медленнее, далее создается впечатление, что колеса остановились совсем и начали вращаться в противоположную сторону. Такая картина объясняется тем, что угловая скорость перемещения регулярных элементов колеса оказывается слишком высокой для того, чтобы при заданной частоте кадров все фазы этого перемещения правильно фиксировались.

Природу возникновения эффекта наложения частот можно продемонстрировать с помощью несложных тригонометрических выкладок [2.26].

Пусть заданы три сигнала:

$$\begin{aligned}x(t) &= \cos(2\pi Ft), \\x_1(t) &= \cos[2\pi(F_n + F)t], \\x_2(t) &= \cos[2\pi(F_n - F)t],\end{aligned}$$

где F и F_n — некоторые частоты, причем $F < \frac{1}{2}F_n$.

Тогда в результате процедуры равномерной дискретизации с частотой F_n (и периодом $T = 1/F_n$) будут соответственно получены следующие три последовательности отсчетов:

$$\begin{aligned}x(n) &= \cos(2\pi F_n nT), \\x_1(n) &= \cos[2\pi(F_n + F)nT], \\x_2(n) &= \cos[2\pi(F_n - F)nT].\end{aligned}$$

Выполним преобразование

$$\begin{aligned}x_1(n) &= \cos[2\pi(F_n + F)nT] = \cos[2\pi(1/T + F)nT] = \\&= \cos(2\pi n + 2\pi F_n nT) = \cos(2\pi F_n nT) = x(n).\end{aligned}$$

Аналогично

$$\begin{aligned}x_2(n) &= \cos[2\pi(F_n - F)nT] = \cos[2\pi(1/T - F)nT] = \\&= \cos(2\pi n - 2\pi F_n nT) = \cos(2 - \pi F_n nT) = \cos(2\pi F_n nT) = x(n).\end{aligned}$$

В приведенных преобразованиях использовались свойства периодичности и четности функции косинус. Оказывается, что $x_1(n) = x_2(n)$. Это означает, что заведомо разные гармонические сигналы $x(t)$, $x_1(t)$ и $x_2(t)$ после их дискретизации с частотой F_n становятся невозможно отличить друг от друга.

На рис. 2.3 показан пример, иллюстрирующий приведенные выкладки. На всех трех графиках сплошными линиями показаны сигналы, дискретизованные с частотой $F_{д1} = 1000$ Гц, а пунктирными — с частотой $F_{д2} = 100$ Гц.

На верхнем, среднем и нижнем графиках соответственно показаны синусоиды с частотами $F_1 = 30$ Гц, $F_2 = F_{д2} + F = 130$ Гц и $F_3 = F_{д2} - F = 70$ Гц.

Из графиков видно, что нарушение условий теоремы отсчетов приводит к тому, что по отсчетам, взятым с частотой дискретизации $F_{д2} = 100$ Гц, заданные три синусоиды невозможно отличить друг от друга.

На рис. 2.4 показан один и тот же фрагмент ЭКГ, содержащий короткие импульсы от электрокардиостимулятора, имеющие длительность около 2 мс. На верхнем графике отсчеты взяты с частотой

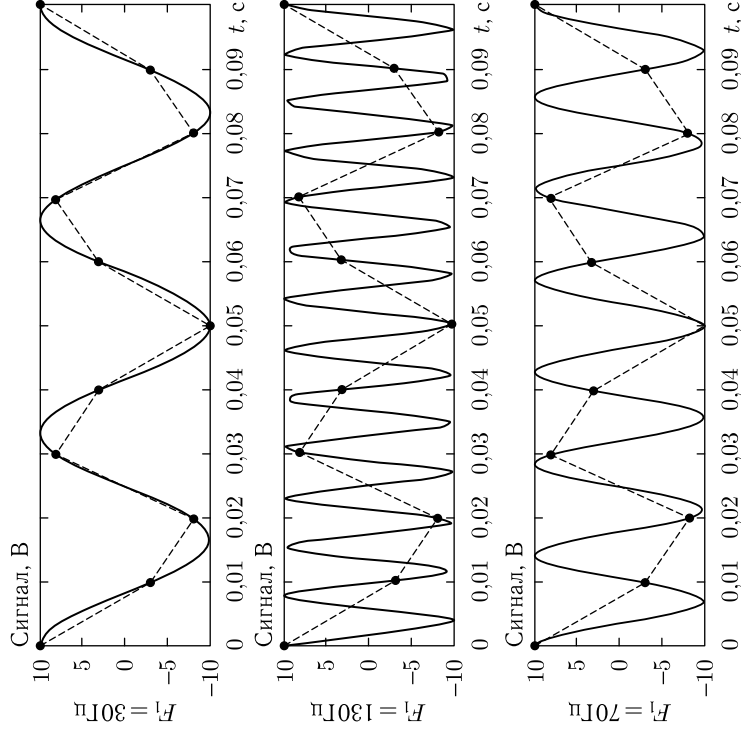


Рис. 2.3. Иллюстрация эффекта наложения частот

1000 Гц, а на нижнем — с частотой 125 Гц, которая явно недостаточна для правильного представления данного сигнала. Из графиков видно, что в данном случае наличие потери важной для медицинской информации сигнала информации, что явилось следствием нарушения условий теоремы отсчетов.

В практических задачах цифровой обработки сигналов использование теоремы отсчетов можно сформулировать в виде следующих основных рекомендаций:

- частота дискретизации должна выбираться так, чтобы она (с запасом не менее чем в 2–3 раза) превышала верхнюю частоту информативных составляющих сигнала;
- перед аналого-цифровым преобразованием полоса частот исходного аналогового сигнала должна быть ограничена с использованием аналогового фильтра нижних частот с частотой среза намного меньшей, чем половина выбранной частоты дискретизации;
- рассмотрение частотных свойств дискретизованного сигнала оправдано лишь для частот, существенно меньших половины частоты дискретизации.

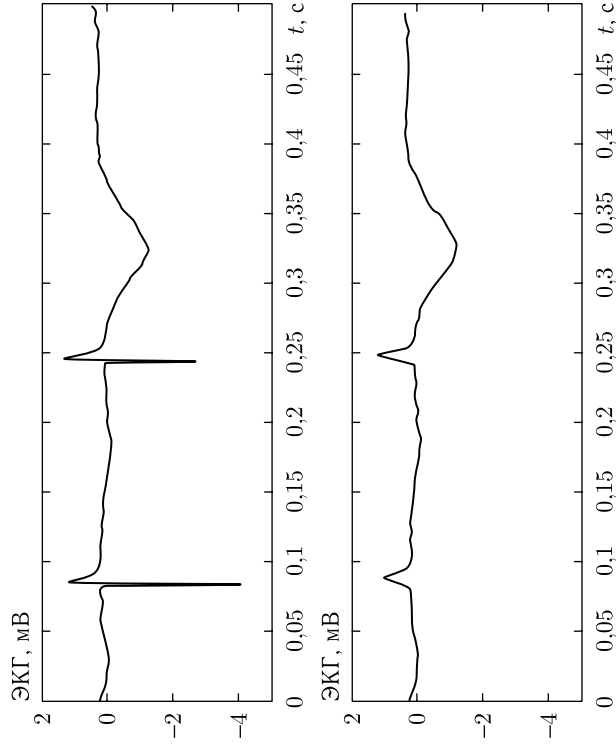


Рис. 2.4. Пример потери высокочастотных составляющих сигнала в результате использования недостаточно высокой частоты дискретизации

Квантование сигналов

Как уже отмечалось, *квантование* — это присваивание отсчету цифрового значения, соответствующего некоторому фиксированному уровню сигнала.

Точность квантования сигнала с использованием аналого-цифрового преобразователя определяется двумя параметрами:

- динамическим диапазоном, задающим пределы измерения входного сигнала;
- разрядностью, определяющей число двоичных разрядов выходного кода АЦП.

Обозначим размах динамического диапазона как $A = A_{\max} - A_{\min}$, где $A_{\max}$ и $A_{\min}$ — соответственно наибольшее и наименьшее измеряемые значения непрерывного сигнала на входе АЦП. Если далее обозначить разрядность АЦП как k , то наибольшее возможное *число уровней квантования* $K = 2^k$, а величина $a = A/K$ соответствует дистанции между соседними уровнями квантования и называется *шагом квантования*.

Шаг квантования определяет точность представления входного аналогового сигнала в виде цифровых отсчетов. Максимальная ошибка квантования зависит от принципа работы АЦП и равняется $\pm a/2$ для АЦП, присваивающего значение ближайшего уровня квантования.

Нетрудно показать [2.24], что среднеквадратичная ошибка квантования при этом будет составлять $a/\sqrt{12} \approx 0,29a$.

Для оптимального выбора динамического диапазона АЦП и его разрядности необходима информация о диапазоне возможного изменения сигнала на входе АЦП и о требуемой точности представления цифрового сигнала. Отметим, что повышение точности за счет увеличения разрядности ограничено точностью представления самого аналогового сигнала (например, уровнем собственных шумов усилителя, если на входе АЦП измеряется напряжение или ток).

Выбор параметров аналого-цифрового преобразования

При решении практических задач цифровой обработки сигналов выбор частоты и разрядности аналого-цифрового преобразования обычно осуществляется на основе компромисса между точностью представления сигнала и необходимыми вычислительными затратами (загрузкой процессора и расходом оперативной памяти).

Рассмотрим, например, случай разработки системы для электрокардиографической диагностики. При этом типичны следующие параметры АЦП: частота отсчетов 500 Гц и разрядность 12 бит при динамическом диапазоне сигнала от ± 5 до ± 10 мВ [2.8]. Эти требования определяются, с одной стороны, стандартной полосой частот усилителей электрокардиосигнала (ЭКС) (0,05–100 Гц), а с другой — необходимостью для измерения амплитудных и временных параметров ЭКГ точностью описания ЭКС.

В то же время при обработке электрокардиосигнала в приборах и системах кардиологического наблюдения (см. гл. 3) выбор параметров аналого-цифрового преобразования электрокардиосигнала определяется задачами, решаемыми конкретным видом аппаратуры. В случае если необходимы лишь текущий контроль частоты сердечных сокращений и распознавание наиболее явно выраженных нарушений сердечного ритма, достаточными могут оказаться разрядность от 8 до 10 бит и частота дискретизации от 100 до 250 Гц [2.8].

При анализе вариабельности сердечного ритма (ВСР, см. 3.8) частота дискретизации ЭКГ оказывает непосредственное влияние на точность измерения статистических и частотных показателей ВСР, так как выбор этой частоты в большой степени определяет погрешность измерения интервалов между сокращениями сердца (RR-интервалов). Международными стандартами рекомендована частота 250 Гц [2.29]. При этом допускается использование и более низких значений — до 128 Гц, что связано с существованием обширных архивов холтеровских записей ЭКС, регистрировавшихся именно с этой частотой. В то же время многие авторы утверждают, что для достоверного анализа показателей ВСР частота отсчетов должна составлять не менее 1000–2000 Гц.

Еще одним обстоятельством, оказывающим влияние на выбор частоты дискретизации ЭКГ, является необходимость анализа сигнала при наличии искусственной электрокардиостимуляции. Так как артефакт стимулирующего импульса обычно имеет длительность не более 1–2 мс, то для его надежного обнаружения цифровым способом необходима частота отсчетов не ниже 1000 Гц (в отдельных системах используют частоту до 8000 Гц). Часто используется также следующая последовательность этапов: ввод сигнала с достаточно высокой частотой, обнаружение артефакта электрокардиостимуляции на этапе предварительной обработки сигнала и последующее прореживание сигнала до частоты 200–250 Гц, достаточной для дальнейших этапов анализа ЭКС. В других случаях прибегают к аппаратному выделению импульса электрокардиостимулятора до аналого-цифрового преобразования сигнала.

2.2. Цифровая фильтрация сигналов

Основные понятия и определения

Традиционно в технике под фильтрацией понимают некоторую операцию, изменяющую частотный состав сигнала. Для дискретных последовательностей наиболее широко распространена *линейная цифровая фильтрация*, определяемая выражением [2.1, 2.24, 2.26]

$$y(n) = \sum_{i=1}^I b_i x(n-i) + \sum_{j=1}^J a_j y(n-j),$$

где $x(n)$ и $y(n)$ — отсчеты соответственно входной и выходной последовательностей, а b_i , a_j — коэффициенты фильтра. Приведенная формула называется *обобщенным разностным уравнением* цифрового фильтра и полностью определяет его.

Если $a_j = 0$ для любых значений j , то фильтр называется *нерекурсивным*, или *КИХ-фильтром* (фильтром с конечной импульсной характеристикой). В противном случае фильтр называется *рекурсивным*, или *БИХ-фильтром* (фильтром с бесконечной импульсной характеристикой).

Фильтры, все коэффициенты которых константы, называются фильтрами с *постоянными коэффициентами*. Фильтры, коэффициенты которых могут изменяться во времени, называются *адаптивными*.

Импульсной характеристикой фильтра называется его отклик на единичный импульс, а *переходной характеристикой* — отклик на единичное ступенчатое воздействие.

Частотные свойства ЦФ определяются его *амплитудно-частотной* и *фазово-частотной* характеристиками (АЧХ и ФЧХ). АЧХ показывает зависимость коэффициента передачи (усиления) фильтра от частоты, а ФЧХ — зависимость вносимого фильтром фазового сдвига (задержки) от частоты.

Классификация цифровых фильтров по виду амплитудно-частотных характеристик

Важнейшая характеристика цифрового фильтра (ЦФ) — его АЧХ, представляющая собой модуль комплексной *передаточной функции* и показывающая зависимость коэффициента передачи фильтра от частоты. Передаточная функция рассчитывается с использованием специализированного математического аппарата, называемого *z-преобразованием*. ЦФ принято классифицировать по виду их АЧХ.

Диапазон частот АЧХ, где ее значения близки к единице, принято называть *полосой пропускания* фильтра. Диапазон частот, где АЧХ приближается к нулю, называют *полосой задержки*, а область частот в промежутке между полосами пропускания и задержки называют *переходной полосой*. Фильтр может иметь по несколько полос каждого из перечисленных видов. Точки, где АЧХ пересекает уровень $\sqrt{2}/2$ (примерно 0,7, что соответствует передаче половины мощности сигнала), называются *частотами среза* фильтра. На рис. 2.5 приведен

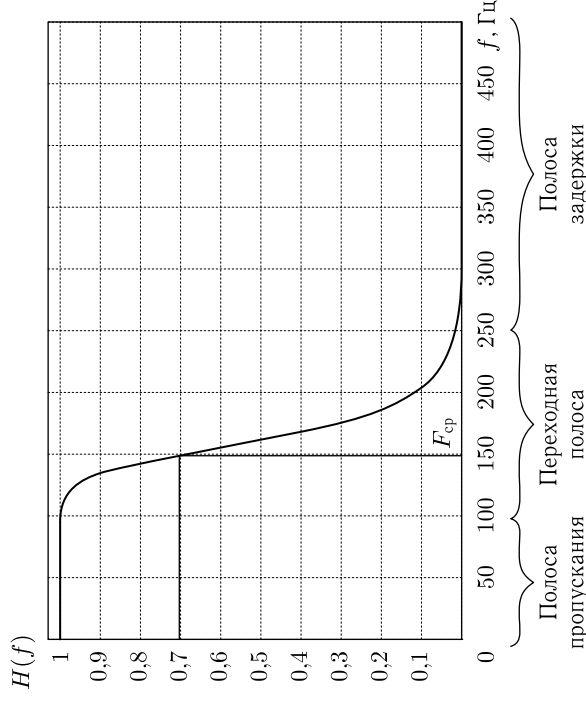


Рис. 2.5. Пример АЧХ цифрового фильтра (фильтра нижних частот). Отмечены характерные полосы частот, а также частота среза $F_{ср}$

пример амплитудно-частотной характеристики фильтра, на которой отмечены перечисленные характерные элементы АЧХ.

На рис. 2.6 приведены примеры амплитудно-частотных характеристик нескольких наиболее распространенных типов фильтров.

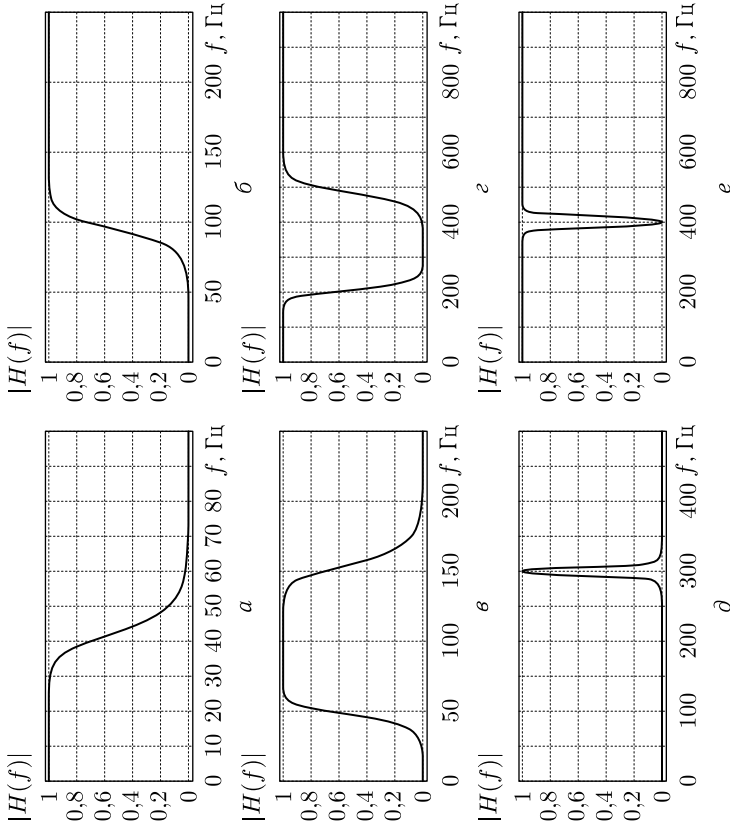


Рис. 2.6. Примеры АЧХ наиболее распространенных типов фильтров: a — фильтр нижних частот, ФНЧ; b — фильтр верхних частот, ФВЧ; c — полосовой пропускающий фильтр; d — полосовой заграждающий фильтр; e — резекторный фильтр

Фильтры, АЧХ которых приведены на рис. 2.6, решают задачу пропускания одних частотных составляющих сигнала и подавления других. Помимо таких ЦФ широкое применение в цифровой обработке сигналов находят фильтры, моделирующие операции дифференцирования и интегрирования (см. ниже). В идеальном случае амплитудно-частотная характеристика дифференциатора должна иметь вид линейно возрастающей функции, а интегратора — гиперболы [2.20]. Характеристики реализуемых на практике цифровых дифференциаторов и интеграторов отличаются от идеальных и приближаются к ним только в определенном диапазоне частот, который называют областью дифференцирования или областью интегрирования (рис. 2.7).

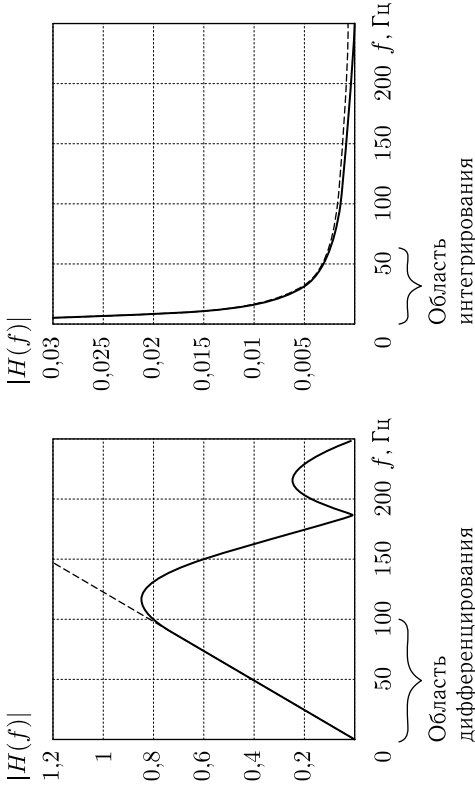


Рис. 2.7. Примеры АЧХ дифференцирующего (слева) и интегрирующего (справа) ЦФ. Пунктирными линиями показаны идеальные характеристики, сплошными — реальные. Фильтры рассчитаны на частоту дискретизации $f_d = 500$ Гц

Методы анализа цифровых фильтров

Выше линейный цифровой фильтр был определен его обобщенным разностным уравнением

$$y(n) = \sum_{i=1}^I b_i x(n-i) + \sum_{j=1}^J a_j y(n-j), \quad (2.1)$$

где $x(n)$ и $y(n)$ — отсчеты соответственно входной и выходной последовательностей, а b_i и a_j — коэффициенты фильтра. Как можно видеть, процедура цифровой фильтрации представляет собой комбинацию из трех элементарных операций: сложения, умножения и задержки (сдвига). Очень удобной и наглядной формой описания ЦФ является структурная схема, основу которой составляют графические обозначения перечисленных операций (см. рис. 2.8, где a — сложение двух отсчетов; b — умножение отсчета на некоторую величину k ; v — сдвиг (задержка) отсчета на m тактов).

Рассмотрим для примера два простых фильтра, заданные уравнениями

$$\begin{aligned} y(n) &= 1/(8x(n)) + 1/(4x(n-1)) + 1/(4x(n-2)) + \\ &\quad + 1/(4x(n-3)) + 1/(8x(n-4)), \\ y(n) &= y(n-1) + 1/(2x(n)) + 1/(2x(n-1)). \end{aligned}$$

Первый из них является КИХ-фильтром нижних частот и известен как фильтр Хемминга, а второй — БИХ-фильтром и представляет собой

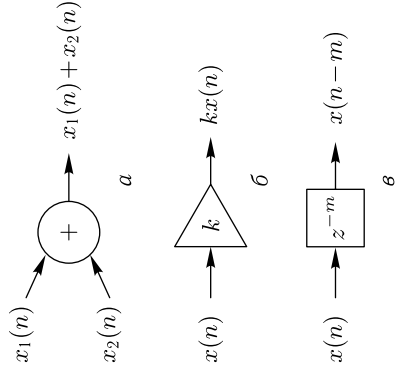


Рис. 2.8. Графические обозначения элементарных операций, используемые в структурных схемах ЦФ

численный интегратор по правилу трапеций [2.20]. На рис. 2.9 показаны структурные схемы обоих фильтров. Как видим, главная отличительная черта БИХ-фильтра — наличие обратной связи, т. е. рекурсии, с чем и связано его название «рекурсивный». Структурные схемы ЦФ не только являются наглядной формой их описания, но могут также использоваться для эквивалентных преобразований фильтров.

О воздействии ЦФ на сигнал можно судить по его импульсной и переходной характеристикам, которые представляют собой отклики соответственно на единичный импульс и на единичное ступенчатое

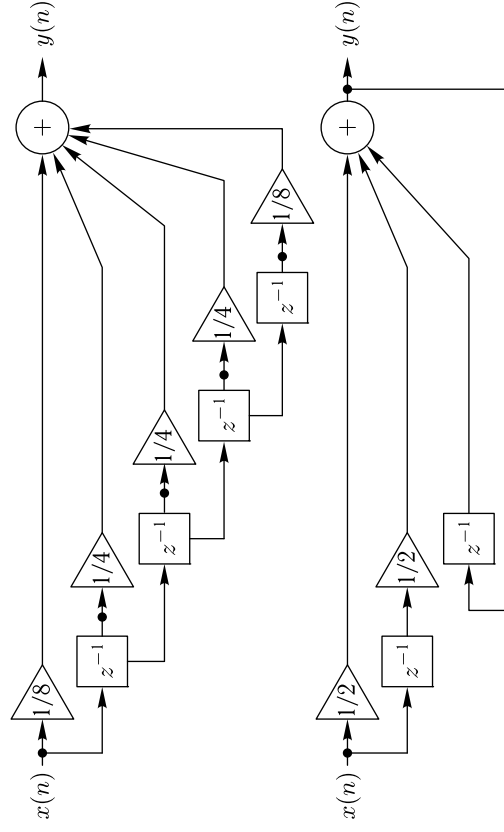


Рис. 2.9. Структурные схемы нерекурсивного (вверху) и рекурсивного (внизу) цифровых фильтров

воздействие. На рис. 2.10 показаны примеры импульсных и переходных характеристик рассмотренных ЦФ. Обратим внимание на то, что импульсная характеристика КИХ-фильтра повторяет набор его коэффициентов (рис. 2.10, б), а импульсная характеристика БИХ-фильтра бесконечна (рис. 2.10, в), с чем и связано название этого фильтра. Ряд графиков на рис. 2.10, а показывает соответственно единичное импульсное и ступенчатое воздействия, подаваемые на входы фильтров.

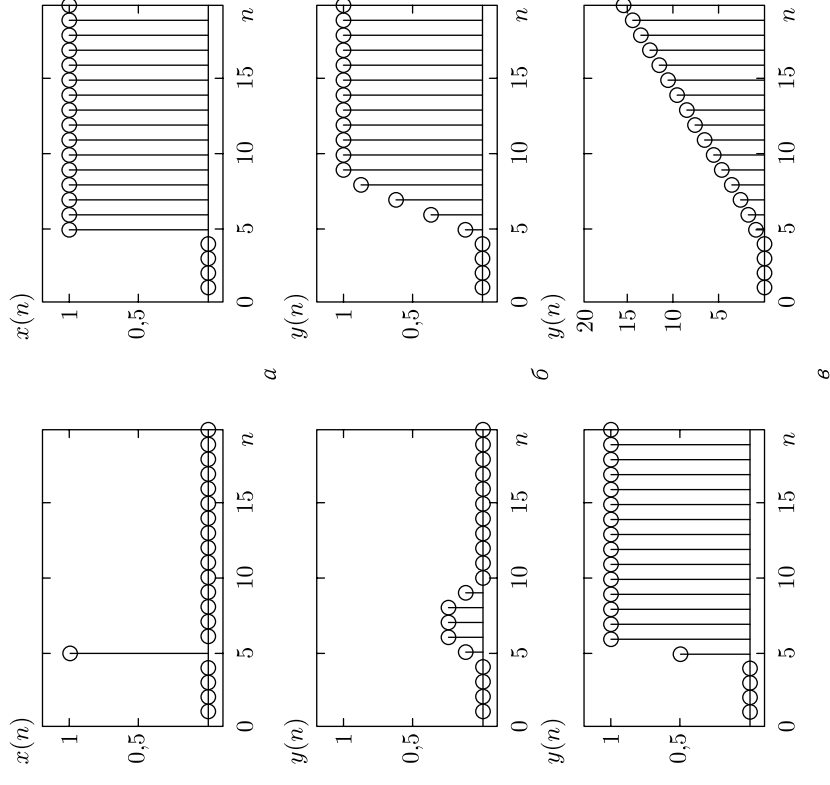


Рис. 2.10. Пример импульсных (слева) и переходных (справа) характеристик рассмотренных в настоящем разделе фильтров

Частотные свойства ЦФ описываются его *передаточной функцией*, которая позволяет определить АЧХ и ФЧХ. АЧХ показывает зависимость коэффициента передачи (усиления), а ФЧХ — вносимого фильтром фазового сдвига (задержки) от частоты.

Передаточная функция ЦФ может быть рассчитана по его разностному уравнению с использованием математического аппарата z -преобразования, которое широко используется при анализе импульсных

и дискретных систем [2.1, 2.24, 2.26]. Смысл z -преобразования можно кратко описать следующим образом.

Пусть задана дискретная последовательность $x(n)$, $n = 0, 1, 2, \dots, N$. Тогда z -преобразованием этой последовательности называется выражение

$$X(z) = \sum_{n=0}^N x(n) z^{-n}.$$

Для ЦФ, заданного разностным уравнением вида (2.1), z -преобразование его передаточной функции имеет вид

$$H(z) = \sum_{i=0}^I b_i z^{-i} / \left(1 - \sum_{j=1}^J a_j z^{-j} \right).$$

Аналитическое выражение для зависимости передаточной функции от частоты $H(\omega)$ может быть получено путем подстановки $z = e^{j\omega T}$, где ω — угловая частота в радианах, а $T = 1/f_n$ — интервал дискретизации в секундах (здесь f_n — частота дискретизации). Функция $H(\omega)$ — комплексная. Ее модуль $|H(\omega)| = \sqrt{[\operatorname{Re} H(\omega)]^2 + [\operatorname{Im} H(\omega)]^2}$ и аргумент $\arg H(\omega) = \operatorname{arctg}[\operatorname{Im} H(\omega) / \operatorname{Re} H(\omega)]$ дают выражения для АЧХ и ФЧХ фильтра соответственно. Для перехода к более удобной в практических задачах циклической частоте f следует выполнить подстановку $\omega = 2\pi f$.

Рассчитаем передаточные функции для рассмотренных выше ЦФ. Выражение для z -преобразования для КИХ-фильтра имеет вид

$$H(z) = \frac{1}{8} z^0 + \frac{1}{4} z^{-1} + \frac{1}{4} z^{-2} + \frac{1}{4} z^{-3} + \frac{1}{8} z^{-4}.$$

Подставим $z = e^{j\omega T}$:

$$\begin{aligned} H(\omega) &= \frac{1}{8} + \frac{1}{4} e^{-j\omega T} + \frac{1}{4} e^{-2j\omega T} + \frac{1}{4} e^{-3j\omega T} + \frac{1}{8} e^{-4j\omega T} = \\ &= \frac{1}{4} e^{-2j\omega T} \left(\frac{1}{2} e^{2j\omega T} + e^{j\omega T} + 1 + 4e^{-1j\omega T} + \frac{1}{2} e^{-2j\omega T} \right) = \\ &= \frac{1}{4} e^{-2j\omega T} [\cos(2\omega T) + 2\cos(\omega T) + 1]. \end{aligned}$$

Последнее преобразование выполнено на основании одной из формул Эйлера, связывающих тригонометрические функции синус и косинус с комплексной экспонентой:

$$\cos x = \frac{e^{jx} + e^{-jx}}{2}, \quad \sin x = \frac{e^{jx} - e^{-jx}}{2j}.$$

Тогда выражения для модуля и аргумента передаточной функции (АЧХ и ФЧХ фильтра) соответственно будут иметь вид

$$\begin{aligned} |H(\omega)| &= \frac{1}{4} [\cos(2\omega T) + 2\cos(\omega T) + 1]; \\ \arg H(\omega) &= -2\omega T. \end{aligned}$$

С целью упрощения математических преобразований их обычно выполняют для нормированной частоты $\omega = 2\pi$ радиан, или $f = 1$ Гц, что соответствует $T = 1$ с. По окончании преобразований переходят обратно к реальным частотам, учитывая частоту дискретизации f_n . В этом случае выражения для АЧХ и ФЧХ для циклической частоты принимают вид

$$\begin{aligned} |H(f)| &= \frac{1}{4} [\cos(4\pi f / f_n) + 2\cos(2\pi f / f_n) + 1]; \\ \arg H(f) &= 4\pi f / f_n. \end{aligned}$$

Выполним аналогичные преобразования для рассматриваемого БИХ-фильтра:

$$\begin{aligned} H(z) &= 0,5 \frac{1 + z^{-1}}{1 - z^{-1}}, \\ H(\omega) &= 0,5 \frac{1 + e^{-j\omega T}}{1 - e^{-j\omega T}} = 0,5 \frac{e^{-j\omega T/2} (e^{j\omega T/2} + e^{-j\omega T/2})}{e^{-j\omega T/2} (e^{j\omega T/2} - e^{-j\omega T/2})} = \\ &= 0,5 \frac{\cos(\omega T/2)}{j \sin(\omega T/2)} = -0,5j \operatorname{ctg}(\omega T/2). \end{aligned}$$

Здесь в преобразованиях использована вторая формула Эйлера (для синуса). Тогда окончательные выражения для АЧХ и ФЧХ фильтра примут вид

$$|H(f)| = 0,5 \operatorname{ctg}(\pi f / f_n); \quad \arg H(f) = -\frac{\pi}{2}.$$

На рис. 2.11 приведены графики АЧХ и ФЧХ обоих фильтров для случая $f_n = 200$ Гц (рис. 2.11, а и б соответственно). Как видно из приведенных характеристик, первый фильтр является фильтром нижних частот (сглаживающим фильтром), в то время как АЧХ второго на начальном участке частотного диапазона близка к характеристике идеального интегратора. Важно отметить также, что в соответствии с теоремой отсчетов преобразования с дискретным сигналом корректны только в пределах диапазона частот от 0 Гц до $f_n/2$, поэтому в данном примере оси частот ограничены полосой от 0 до 100 Гц.

Нетрудно заметить, что числитель и знаменатель выражения для передаточной функции ЦФ являются полиномами переменной z . Для каждого из них могут быть найдены корни, которые для числителя носят название *нулей*, а для знаменателя — *полюсов*. Нули и полюсы передаточной функции представляют собой комплексные числа

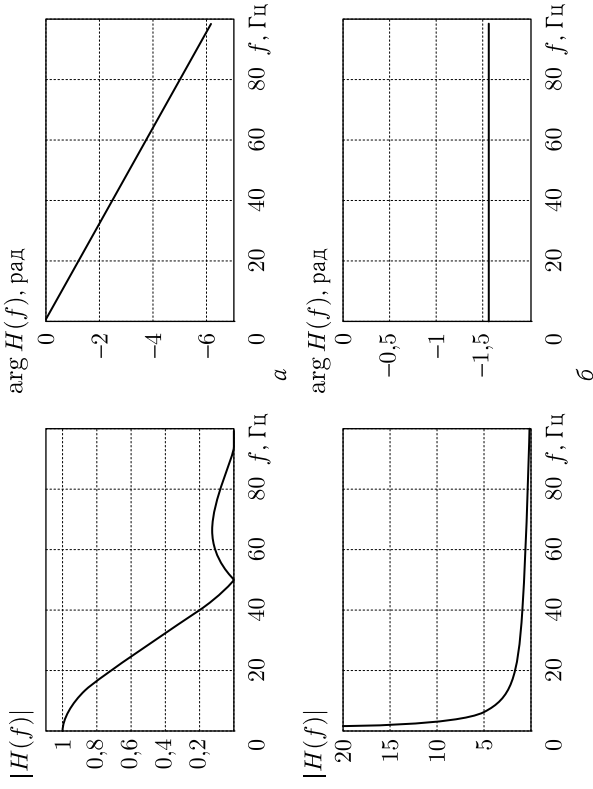


Рис. 2.11. Графики АЧХ (слева) и ФЧХ (справа) КИХ- и БИХ-фильтров

и равняются таким значениям переменной z , при которых передаточная функция обращается соответственно в нуль и в бесконечность.

Передаточная функция для КИХ-фильтра содержит только выражение для числителя, поэтому для фильтров этого типа существуют только нули — корни полинома для числителя. Формально для КИХ-фильтра также определен один полюс со значением $z = 0$, но его обычно не рассматривают. В приведенном примере КИХ-фильтра полином в числителе передаточной функции имеет четвертую степень. Решение уравнения дает следующие значения нулей:

$$z_1^0 = j; \quad z_2^0 = -j; \quad z_{3,4}^0 = -1.$$

Для БИХ-фильтров существуют и нули, и полюсы. В рассмотренном примере БИХ-фильтр имеет один полюс и один нуль, которые соответственно принимают значения

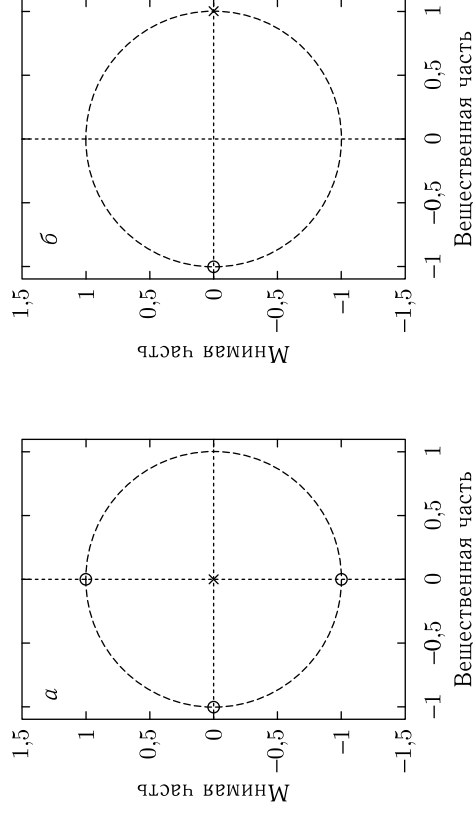
$$z_1^* = 1, \quad z_1^0 = -1.$$

Широкое применение при анализе цифровых систем находят графики, показывающие расположение нулей и полюсов на комплексной плоскости (*карты нулей и полюсов*). Нули на графиках принято обозначать кружками, а полюсы — звездочками. На графиках всегда отображают *единичную окружность*, очерчиваемую проведенным из начала координат вектором с длиной, равной единице (*единичным вектором*). Единичная окружность определяется уравнением $z = e^{j\omega T}$.

Угол поворота вектора против часовой стрелки начиная от положительного направления вещественной оси можно интерпретировать как нормированную угловую частоту ω . Если, учитывая фактически используемую частоту дискретизации сигнала f_d , перейти к реальным частотам (как показано выше для АЧХ и ФЧХ), то полному обороту вектора соответствует изменение частоты от 0 до f_d . Нетрудно продемонстрировать [2.20], что при приближении конца вектора к какому-нибудь из нулей модуль передаточной функции уменьшается, а при приближении к полюсам — возрастает. Благодаря отмеченным обстоятельствам визуальный анализ карты нулей и полюсов позволяет, не выполняя сложных аналитических преобразований, не только составить представление о виде АЧХ и ФЧХ ЦФ, но также приблизительно рассчитать их графики, пользуясь циркулем, транспортиром и линейкой.

Одним из важнейших применений карты нулей и полюсов служит оценка *устойчивости* ЦФ. Устойчивым называется ЦФ, дающий конечный отклик на конечное входное воздействие. Фильтры, не обладающие свойством устойчивости, не имеют практического применения, так как порождают процессы, способные вывести из строя всю систему, в которой они используются. Существует следующее правило: *фильтр устойчив, если все его полюсы располагаются в пределах единичной окружности*.

На рис. 2.12 приведены карты нулей и полюсов: *а* — для КИХ-фильтра; *б* — для БИХ-фильтра. Сопоставляя их с АЧХ фильтров (см. рис. 2.11) и учитывая описанную интерпретацию вращения единичного вектора, можно легко проследить причины возникновения максимумов и минимумов АЧХ на определенных частотах. Анализ

Рис. 2.12. Карты нулей и полюсов КИХ- (*а*) и БИХ-фильтров (*б*)

устойчивости с помощью приведенного выше правила показывает, что оба фильтра — устойчивые. Отметим, что КИХ-фильтры устойчивы всегда, что относится к важнейшим их достоинствам.

В рассмотренных примерах ЦФ аналитические выражения для АЧХ и ФЧХ удалось вывести из передаточной функции относительно просто, используя формулы Эйлера. Однако на практике часто используются фильтры с таким набором коэффициентов, что вывод аналитических формул оказывается чрезвычайно громоздкой процедурой. В настоящее время эта проблема без труда преодолевается благодаря наличию специальных программ расчета АЧХ и ФЧХ цифровых систем.

Одно из основных достоинств КИХ-фильтров — относительная простота обеспечения линейности фазовой характеристики (т.е. постоянства временной задержки для составляющих всех частот), что, в частности, может быть очень важным при анализе электрофизиологических сигналов, так как нелинейность фазовой характеристики способна приводить к изменению взаимного расположения элементов сигнала или даже к появлению «лишних» колебаний.

Линейность фазовых характеристик КИХ-фильтров обеспечивается при соблюдении одного из четырех условий симметрии импульсной характеристики [2.1, 2.26]:

- симметричность при нечетном числе коэффициентов;
- симметричность при четном числе коэффициентов;
- антисимметричность при нечетном числе коэффициентов;
- антисимметричность при четном числе коэффициентов.

В табл. 2.1 сравниваются характеристики КИХ- и БИХ-фильтров. В практических задачах для выбора какого-либо из этих двух

Таблица 2.1

Характеристики	Тип цифровых фильтров	
	КИХ	БИХ
Достоинства	Всегда устойчивы. Невозможно накопление ошибок округления. Простота обеспечения линейности ФЧХ	Относительно небольшое число коэффициентов для обеспечения заданного вида передаточной функции
Недостатки	Большое число коэффициентов для обеспечения заданного вида передаточной функции	Возможна неустойчивость. Возможно накопление ошибок округления. Сложность или иногда невозможность обеспечения линейности ФЧХ

видов ЦФ необходимо исходить из конкретных требований к ним и условий их применения.

Фильтры Баттерворта

Основное достоинство рекурсивных фильтров состоит в возможности получения крутого спада АЧХ при использовании относительно небольшого числа коэффициентов фильтра. Поэтому, несмотря на наличие серьезных недостатков (возможная неустойчивость, накопление ошибок округления, нелинейность фазово-частотной характеристики), рекурсивные фильтры применяются очень часто. Из наиболее распространенных рекурсивных фильтров можно назвать так называемые фильтры Баттерворта [2.1, 2.26]. Их отличительной особенностью является монотонное изменение АЧХ в переходной полосе частот и практически полное отсутствие пульсаций в полосах пропускания и задержки. В некоторых применениях это очень важно, так как позволяет гарантировать постоянно коэффициент передачи фильтра во всей рабочей полосе частот и обеспечить надежное подавление устраняемых фильтром частотных составляющих.

Вид амплитудно-частотной характеристики фильтра нижних частот Баттерворта определяется из условия

$$|H(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^2},$$

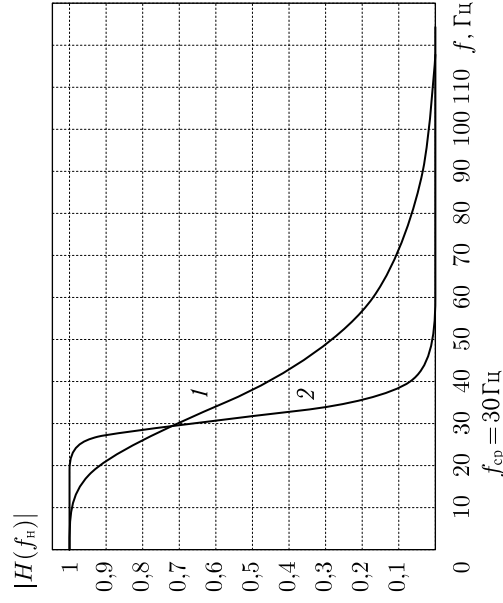


Рис. 2.13. Амплитудно-частотные характеристики фильтров нижних частот Баттерворта 2-го порядка (1) и 8-го порядка (2), имеющих одну и ту же частоту среза 30 Гц. Частота дискретизации 250 Гц

где ω_c — частота среза фильтра в радианах (уровень $\sqrt{2}/2$ на АЧХ), а N — порядок фильтра, определяющий число коэффициентов в рекурсивной и нерекурсивной частях разностного уравнения, которое в этом случае будет иметь вид

$$y(n) = \sum_{i=0}^N b_i x(n-i) + \sum_{j=1}^N a_j y(n-j).$$

Чем выше порядок фильтра Баттерворта, тем более крутой спад имеет его АЧХ в переходной полосе. В рекурсивных фильтрах достаточно крутой спад может быть достигнут при относительно небольшом числе коэффициентов. Однако крутой спад АЧХ не всегда можно рассматривать как положительное свойство фильтра. Рассмотрим пример фильтрации фрагмента ЭКГ с использованием фильтров Баттерворта. На рис. 2.13 приведены АЧХ двух фильтров нижних частот, имеющих одинаковую частоту среза 30 Гц: с порядком 2 и с порядком 8. Последний фильтр имеет АЧХ, быстро спадающую в переходной полосе.

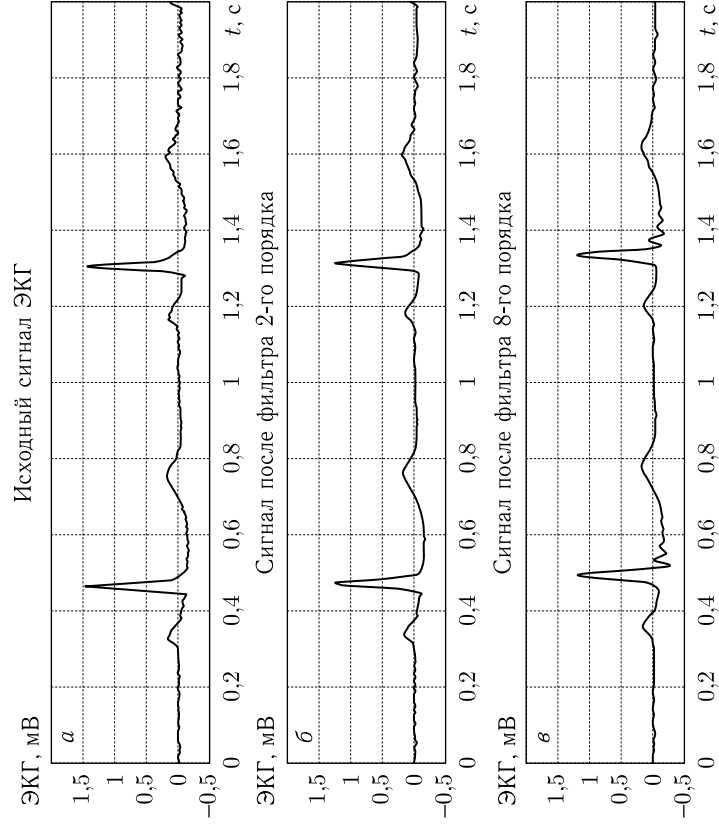


Рис. 2.14. Иллюстрация возникновения «звона» в результате фильтрации: a — исходный фрагмент ЭКГ; b — результат обработки БИХ-фильтром 2-го порядка; $в$ — результат обработки БИХ-фильтром 8-го порядка (в сигнале наблюдается «звон»)

На графике « a » на рис. 2.14 показан фрагмент ЭКГ длительностью 2 с, частота дискретизации сигнала здесь составляет 250 Гц. Графики « b » и « $в$ » представляют тот же сигнал, прошедший соответственно через ФНЧ 2-го и 8-го порядков, АЧХ которых приведены на рис. 2.13. На нижнем из графиков отчетливо наблюдаются так называемый «звон» — высокочастотные затухающие колебания, следующие сразу за участком сигнала, содержащим резкий всплеск (в данном случае — QRS-комплекс ЭКГ). При последующем анализе ЭКГ этот «звон» может быть ошибочно интерпретирован как S-волна ЭКГ, что может привести к неверному диагностическому заключению.

Одним из наиболее существенных недостатков БИХ-фильтров является почти всегда имеющая место нелинейность фазово-частотной характеристики, что означает непостоянство вносимой фильтрами задержки для разных частот. Это может вызывать фазовые искажения сигналов, что в некоторых приложениях неприемлемо. Один из вариантов выхода из данного затруднения — использование КИХ-фильтров, для которых нетрудно обеспечить линейность ФЧХ за счет выбора симметричного набора коэффициентов [2.1, 2.24, 2.26]. Однако

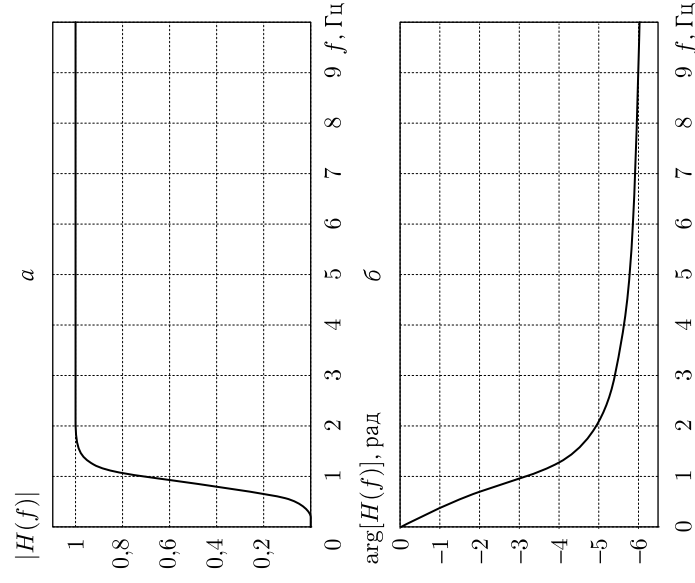


Рис. 2.15. Характеристики фильтра верхних частот Баттерворта 4-го порядка с частотой среза 1 Гц: a — амплитудно-частотная характеристика; b — фазово-частотная характеристика (нелинейная)

существует решение, позволяющее избежать фазовых искажений и при использовании БИХ-фильтров.

Рассмотрим задачу устранения низкочастотной помехи (дрейфа изо-электрической линии) из ЭКГ. Так как все полезные составляющие электрокардосигнала лежат в диапазоне частот выше 1 Гц, низкочастотную помеху можно существенно ослабить фильтром верхних частот с частотой среза около 1 Гц, не затрагивая при этом саму ЭКГ. Пример АЧХ и ФЧХ такого фильтра показан на рис. 2.15. В данном случае предлагается использовать фильтр Баттерворта 4-го порядка. Как видно из графика «б» на рис. 2.15, фазово-частотная характеристика фильтра нелинейна.

График «а» на рис. 2.16 показывает фрагмент ЭКГ, содержащий значительную низкочастотную помеху. Результат использования

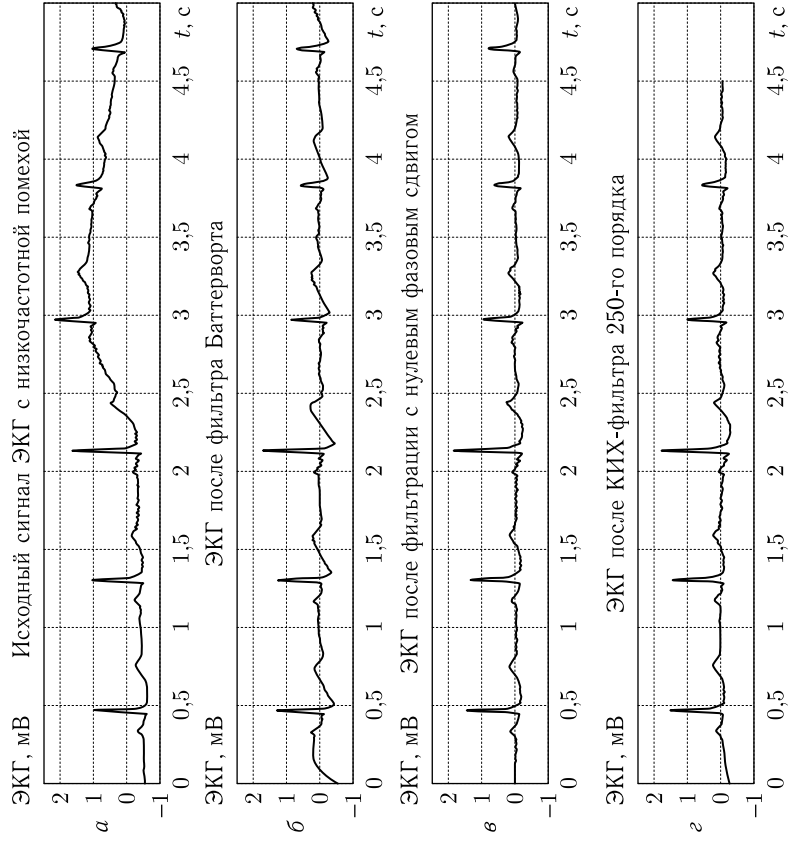


Рис. 2.16. Иллюстрация возникновения фазовых искажений в результате фильтрации: а — исходный фрагмент ЭКГ с низкочастотной помехой; б — результат обработки БИХ-фильтром 4-го порядка (в сигнале наблюдаются искажения, вызванные нелинейностью ФЧХ); в — результат фильтрации с нулевым фазовым сдвигом; г — результат обработки КИХ-фильтром 250-го порядка

предложенного выше ФВЧ показан на графике «б». Дрейф изолинии был успешно устранен фильтром, но в то же время на графике хорошо видны искажения формы кардиосигнала, причиной которых являются фазовые искажения, вносимые фильтром. Данный недостаток может быть устранен за счет применения так называемой фильтрации с нулевым фазовым сдвигом. Суть этого приема заключается в том, что фрагмент сигнала фильтруется одним и тем же фильтром сперва в прямом, а затем в обратном направлении. Возникающие при этом фазовые искажения взаимно компенсируются и результирующий фазовый сдвиг равен нулю для всех частот. На графике «в» представлен результат такой фильтрации. Как можно видеть, заметных искажений сигнала при этом не наблюдается.

При использовании фильтрации с нулевым фазовым сдвигом следует учитывать два важных обстоятельства:

1) так как один и тот же фильтр применяется дважды, эквивалентная амплитудно-частотная характеристика всей процедуры в целом определяется как квадрат АЧХ исходного фильтра. Частота среза при этом смещается в сторону верхних частот, а спад в переходной полосе становится круче (см. рис. 2.17). Для получения желаемого результата фильтрации необходимо так выбирать АЧХ фильтра, чтобы результирующая АЧХ соответствовала требуемой;

2) фильтрация с нулевым фазовым сдвигом может быть выполнена только в тех случаях, когда весь сигнал известен заранее. Это означает, что данный прием невозможно использовать при обработке сигналов в реальном масштабе времени.

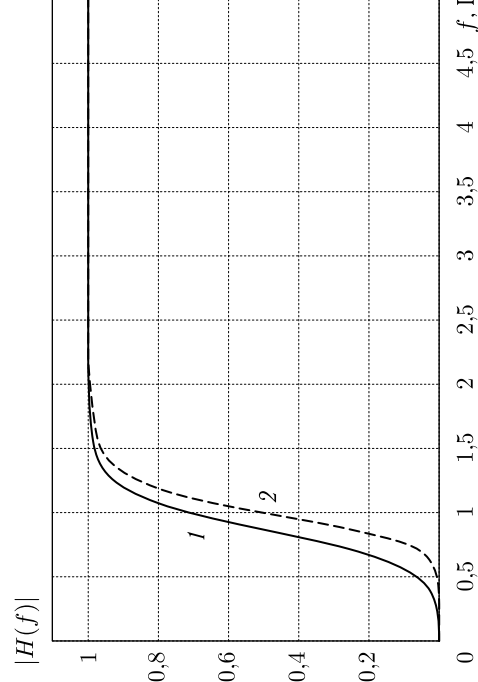


Рис. 2.17. Иллюстрация изменения вида АЧХ фильтра при использовании фильтрации с нулевым фазовым сдвигом: исходная АЧХ (1) и эквивалентная АЧХ при двукратной фильтрации (2)

Практически такой же результат может быть получен при использовании КИХ-фильтра с линейной фазово-частотной характеристикой. На рис. 2.18 приведен пример амплитудно-частотной характеристики и импульсной характеристики (набора коэффициентов) КИХ-фильтра 250-го порядка, имеющего примерно такую же АЧХ, что и рассмотренный ранее ФВЧ Баттерворта. Результат использования этого фильтра показан на графике «г» рис. 2.16. Так же как и в случае использования фильтрации с нулевой фазой, низкочастотная помеха успешно устраняется, а существенные искажения сигнала отсутствуют. Однако использование КИХ-фильтров часто ограничивается вычислительной мощностью применяемого компьютера (данный фильтр имеет 250 коэффициентов, в то время как примерно эквивалентный БИХ-фильтр — всего 10 коэффициентов). Кроме того, КИХ-фильтр вносит существенную задержку в обрабатываемый сигнал (в данном примере она составляет 125 интервалов дискретизации), что не всегда приемлемо в системах анализа сигналов в реальном масштабе времени.

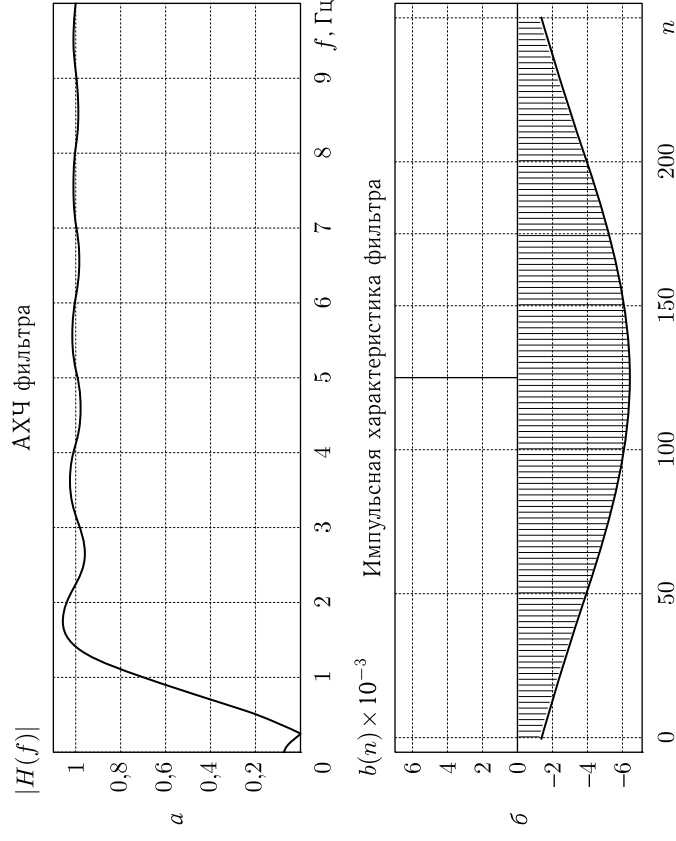


Рис. 2.18. Характеристики КИХ-фильтра верхних частот 250-го порядка с частотой среза 1 Гц: *a* — амплитудно-частотная характеристика; *б* — импульсная характеристика (набор коэффициентов); здесь верхняя часть импульса, отображающего центральный коэффициент, равный 0.9936, не помещалась на графике в выбранном масштабе

Цифровое дифференцирование и интегрирование

Дифференцирование и интегрирование — важнейшие математические операции как для решения теоретических проблем, так и при практическом использовании математических методов. Для случаев когда входные данные представлены последовательностью чисел (то есть дискретны), были разработаны методы численной математики, позволяющие, в частности, реализовать операции дифференцирования и интегрирования с любой заданной точностью. Многие из этих методов можно интерпретировать как варианты цифровой фильтрации [2.20, 2.24, 2.26], в связи с чем представляет интерес их рассмотрение с точки зрения частотного подхода. В настоящем разделе рассмотрены некоторые наиболее часто используемые практические формулы, приведенные их частотные характеристики, а также представлены примеры их применения в задачах цифровой обработки биомедицинских сигналов.

Для рассмотрения операции дифференцирования с позиций частотного подхода необходимо исследовать ее воздействие на гармонический (синусоидальный) сигнал, который в наиболее общем виде описывается комплексной экспонентой:

$$x(t) = e^{j\omega t} = \cos \omega t + j \sin \omega t,$$

где ω — угловая частота в радианах.

Производная от гармонического сигнала определяется выражением

$$x'(t) = de^{j\omega t} / dt = j\omega e^{j\omega t}.$$

Рассмотрим некоторую функцию $H(\omega) = x'(t)/x(t) = j\omega$, которая показывает связь между входом и выходом операции дифференцирования и может интерпретироваться как передаточная функция дифференциатора по отношению к гармоническому сигналу. Модуль и аргумент данной функции являются амплитудно-частотной и фазово-частотной характеристиками и определяются выражениями

$$|H(\omega)| = \omega; \quad \arg H(\omega) = \frac{\pi}{2}.$$

Это означает, что усиление дифференциатора по отношению к непрерывному гармоническому сигналу прямо пропорционально частоте, а вносимый фазовый сдвиг всегда равен $\pi/2$.

Как известно, операция вычисления производной непрерывной функции $x(t)$ определяется выражением

$$x'(t) = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t},$$

где Δt — некоторый интервал времени. Если входные данные представлены последовательностью равноотстоящих отсчетов, следующих

с интервалом T , то можно предположить, что аналогом вышеприведенной формулы для случая дискретного сигнала является выражение

$$y(n) = \frac{x(n) - x(n-1)}{T} = \frac{1}{T} x(n) - \frac{1}{T} x(n-1),$$

где $x(n)$ и $y(n)$ — соответственно отсчеты входной и выходной последовательностей, а n — целочисленный порядковый номер отсчета. Нетрудно заметить, что данная формула, известная как первая разность отсчетов, представляет собой рекурсивный цифровой фильтр с двумя коэффициентами, равными $1/T$ и $-1/T$. Используя z -преобразование, получим выражение для передаточной функции данного фильтра

$$H(z) = (1 - z^{-1})/T;$$

$$H(\omega) = (1 - e^{-j\omega})/T = e^{-j\omega/2}(e^{j\omega/2} - e^{-j\omega/2})/T = 2je^{-j\omega/2} \sin(\omega/2)/T.$$

Соответственно выражения для АЧХ и ФЧХ будут иметь вид

$$|H(\omega)| = 2 \sin(\omega/2)/T;$$

$$\arg H(\omega) = \pi/2 - \omega/2.$$

Кривая 2 на рис. 2.19 показывает вид АЧХ этого фильтра в диапазоне нормированной частоты F_n от 0 до 0,5, а пунктирной линией (кривая 1) показана АЧХ идеального дифференциатора. Под нормированной частотой понимается частота, деленная на используемую частоту дискретизации:

$$F_n = F/F_n = F \cdot T.$$

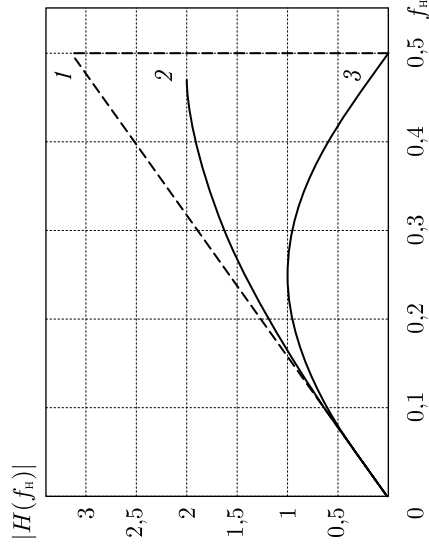


Рис. 2.19. Амплитудно-частотные характеристики идеального дифференциатора (1), фильтра первой разности отсчетов (2) и фильтра центральной первой разности отсчетов (3)

Использование нормированной частоты — общепринятый способ упрощения математических выкладок на этапе анализа свойств цифровых фильтров и эквивалентно предположению, что интервал дискретизации $T = 1$ с. Тогда выражение для модуля передаточной функции принимает вид

$$|H(\omega)| = 2 \sin(\omega/2).$$

При этом значению $F_n = 0,5$ соответствует частота свертывания (частота Найквиста) $F_c = F_n/2$. Графики обрываются при значениях $F_n = 0,5$, так как в соответствии с теоремой отсчетов цифровая обработка сигнала на более высоких частотах некорректна из-за эффекта наложения частот.

Можно видеть, что фильтр первой разности отсчетов обладает свойствами дифференциатора только в диапазоне нормированной частоты от 0 до примерно 0,1. Этот диапазон называют областью дифференцирования. На более высоких частотах кривая существенно отклоняется от идеальной. Фактически это приводит к усилению высокочастотных составляющих сигнала, в связи с чем часто говорят, что за пределами области дифференцирования этот фильтр усиливает шум. Так как данное свойство обычно считается недостатком, фильтр первой разности отсчетов на практике применяют редко.

Значительно большее распространение получила формула центральной первой разности отсчетов

$$y(n) = \frac{x(n) - x(n-2)}{2T} = \frac{1}{2T} x(n) - \frac{1}{2T} x(n-2).$$

Выражения для АЧХ и ФЧХ в этом случае имеют вид

$$|H(\omega)| = \sin(\omega)/T; \quad \arg H(\omega) = \pi/2 - \omega.$$

Соответствующий график АЧХ представлен на рис. 2.19 кривой 3, которая показывает, что данный фильтр имеет примерно такую же область дифференцирования, что и формула первой разности отсчетов, но усиление шумов проявляется в гораздо меньшей степени. В связи с этим центральная первая разность часто используется для цифрового дифференцирования в тех случаях, когда нет жестких требований к математической точности данной операции.

Более точно цифровое дифференцирование может быть реализовано с использованием специально рассчитанных КИХ-фильтров. При этом число коэффициентов может составлять несколько десятков. На рис. 2.20 приведены амплитудно-частотная и импульсная характеристики цифрового дифференцирующего фильтра 50-го порядка. Отметим, что импульсная характеристика КИХ-фильтра совпадает с набором его коэффициентов. Данный фильтр имеет АЧХ, близкую к характеристике идеального дифференциатора, в диапазоне нормированной частоты от 0 до приблизительно 0,14, а далее кривая резко спадает, что исключает эффект усиления помех.

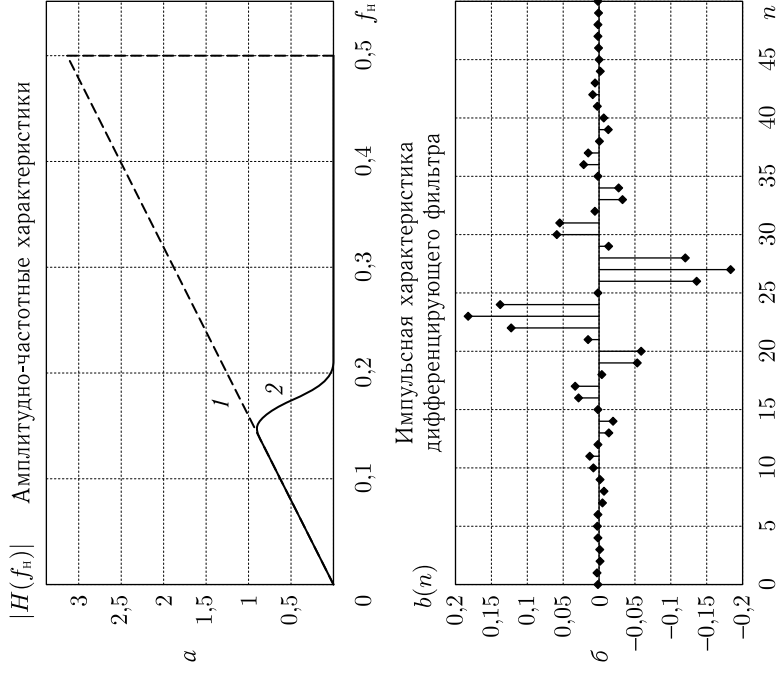


Рис. 2.20. Характеристики дифференцирующих фильтров: *a* — амплитудно-частотные характеристики идеального дифференциатора (1) и дифференцирующего фильтра 50-го порядка (2); *б* — импульсная характеристика (набор коэффициентов) дифференцирующего фильтра 50-го порядка

Одним из наиболее широко известных применений цифрового дифференцирования в обработке биомедицинских сигналов является решение задачи выделения R-зубца ЭКГ на фоне остальных компонентов кардиоцикла и помех. График *a* на рис. 2.21 демонстрирует пример фрагмента ЭКГ продолжительностью 2 с, содержащего типичные для ЭКГ помехи (сетевая наводка 50 Гц, широкополосная микрографическая помеха, дрейф изоэлектрической линии). Кроме того, присутствующие в ЭКГ T-зубцы также затрудняют выделение R-зубцов. Частота дискретизации сигнала в данном случае составляет 250 Гц — типичное значение для многих приборов и систем, предназначенных для контроля сердечного ритма. Справа от графика ЭКГ приведен график спектральной плотности мощности (график *б*), рассчитанной для данного фрагмента сигнала. Дрейф изолинии проявляется в спектре в виде высоких значений на частотах, близких к нулю, сетевая наводка дает

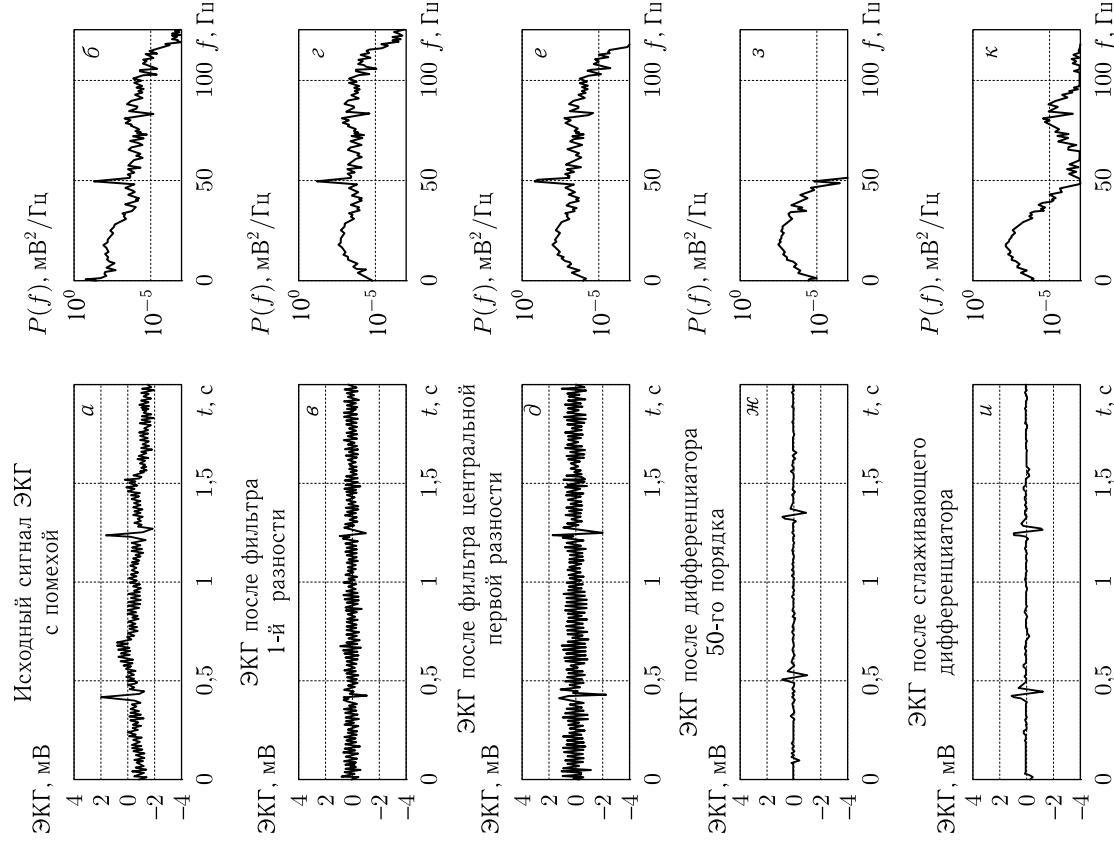


Рис. 2.21. Фрагмент ЭКГ, содержащий помехи (*a*) и график спектральной плотности мощности для этого фрагмента (*б*). Фрагмент ЭКГ и соответствующие графики спектральной плотности мощности после применения различных процедур цифровой фильтрации: формулы первой разности отсчетов (*в*, *г*); формулы центральной первой разности отсчетов (*в*, *г*); дифференциатора 50-го порядка (*ж*, *з*); сглаживающего дифференциатора (*и*, *к*)

в спектре пик на частоте 50 Гц, а миографический шум — примерно равномерный фоновый уровень во всем диапазоне частот. Собственно компоненты кардиоцикла дают основной вклад в спектр в диапазоне от 0,5 до 50 Гц. Составляющие спектра ЭКГ в области частот выше 7–10 Гц обусловлены в основном вкладом QRS-комплекса. Требуется подобрать такую процедуру фильтрации, которая позволила бы в максимальной степени подавить все компоненты сигнала, за исключением QRS-комплекса, что позволило бы обеспечить его надежное обнаружение на следующем этапе обработки сигнала. Так как QRS-комплекс является наиболее высокочастотным компонентом кардиоцикла, для его выделения часто используют дифференцирующие фильтры.

Графики *в* и *г* (рис. 2.21) демонстрируют результаты применения формулы первой разности отсчетов. При этом оказались подавленными низкочастотные помехи и Т-волны ЭКГ. Однако отчетливо проявился свойственный данному фильтру эффект усиления шумов. С точки зрения решения задачи выделения R-зубца ЭКГ данный фильтр не дает удовлетворительного результата. Использование центральной первой разности (графики *д* и *е*) лишь несущественно улучшает результат за счет меньшего эффекта усиления помех.

Применение описанного выше дифференциатора 50-го порядка (графики *ж* и *з*) дает наилучший результат: существенно подавлены все виды помех, а QRS-комплекс отчетливо выделен в сигнале. Отметим существенную задержку, вносимую этим фильтром в сигнал, что связано с большим числом коэффициентов фильтра. Применив несложные пороговые правила к полученному на выходе фильтра сигналу, можно получить вполне работоспособный обнаружитель R-зубца (подробнее данный вопрос рассмотрен далее в главе 3).

Недостаток последнего из описанных фильтров — его относительно высокая вычислительная сложность — может сделать невозможным его применение в портативных устройствах контроля ЭКГ в реальном масштабе времени. Практически такой же результат с точки зрения решения задачи обнаружения QRS-комплекса ЭКГ может быть получен с использованием значительно более простого в вычислительном отношении фильтра. Примером такого простого фильтра служит сглаживающий дифференциатор, определяемый разностным уравнением вида

$$y(n) = \frac{1}{8} [x(n-8) + x(n-7) + x(n-6) + x(n-5) - \\ - x(n-3) - x(n-2) - x(n-1) - x(n)].$$

На рис. 2.22 приведена амплитудно-частотная характеристика этого фильтра (кривая 2). Пунктиром на графике показана АЧХ идеального дифференциатора (кривая 1). Видно, что данный фильтр в диапазоне от 0 до 15 Гц обладает свойствами дифференциатора. Далее характеристика спадает, т. е. фильтр пропускает преимущественно часть спектральных составляющих ЭКГ, которые относятся к QRS-комплексу. Еще одно полезное свойство фильтра — нулевой коэффициент передачи

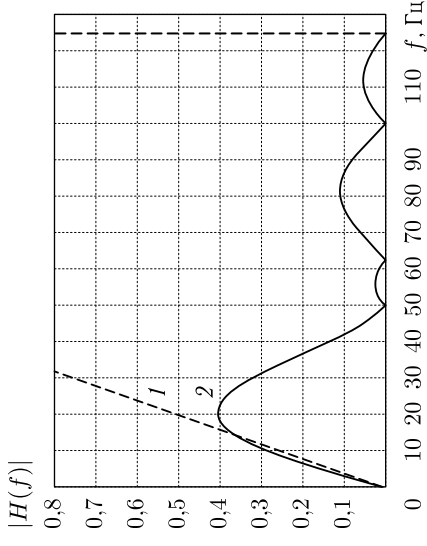


Рис. 2.22. Амплитудно-частотные характеристики идеального дифференциатора (1) и сглаживающего дифференциатора (2) для частоты дискретизации, равной 250 Гц

на частоте 50 Гц (это справедливо только для частоты дискретизации сигнала 250 Гц!), что означает одновременное решение как задачи выделения QRS-комплекса, так и задачи подавления сетевой наводки. Графики *и* и *к* рис. 2.21 демонстрируют результат применения данного фильтра. Как можно видеть, задача выделения QRS-комплекса ЭКГ на фоне остальных волн кардиоцикла и помех решается в данном случае достаточно успешно.

Взятие второй производной от гармонического сигнала дает

$$x''(t) = -\omega^2 e^{j\omega t}.$$

Тогда амплитудно-частотная и фазово-частотная характеристики соответственно определяются выражениями

$$|H(\omega)| = \omega^2; \quad \arg H(\omega) = \pi.$$

Кривая 1 на рис. 2.23 показывает вид АЧХ операции дифференцирования 2-го порядка для нормированной частоты.

Если приближенным цифровым выражением для первой производной считать первую разность дискретных отсчетов, то в качестве аналога второй производной можно рассматривать двукратное применение формулы первой разности:

$$z(n) = \frac{x(n) - x(n-1)}{T}; \\ y(n) = \frac{z(n) - z(n-1)}{T} = \frac{[x(n) - x(n-1)] - [x(n-1) - x(n-2)]}{T^2} = \\ = \frac{1}{T^2} x(n) - \frac{2}{T^2} x(n-1) + \frac{1}{T^2} x(n-2).$$

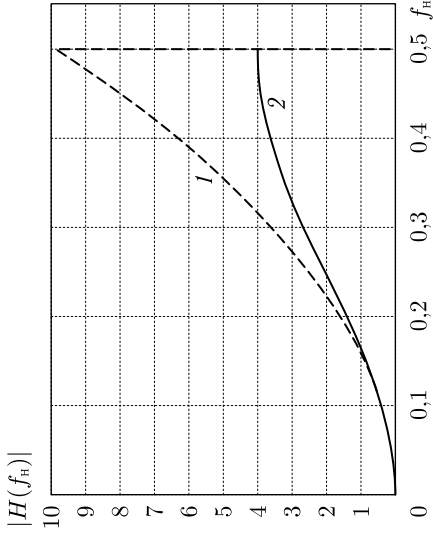


Рис. 2.23. Амплитудно-частотные характеристики идеальной дифференциатора 2-го порядка (1) и формулы второй разности отсчетов (2) для нормированной частоты

Используя z -преобразование, получим выражение для передаточной функции данного фильтра

$$\begin{aligned} H(z) &= (1 - 2z^{-1} + z^{-2})/T^2; \\ H(\omega) &= (1 - 2e^{-j\omega} + e^{-j2\omega})/T^2 = e^{-j\omega}(e^{j\omega} - 2 - e^{-j\omega})/T^2 = \\ &= 2e^{-j\omega}[\cos(\omega) - 1]/T^2. \end{aligned}$$

Соответственно выражения для АЧХ и ФЧХ будут иметь вид

$$|H(\omega)| = 2[\cos(\omega) - 1]/T^2; \quad \arg H(\omega) = -\omega.$$

Так как в случае использования нормированной частоты $T = 1$, то получим

$$|H(\omega)| = 2[\cos(\omega) - 1].$$

Кривая 2 на рис. 2.23 показывает вид АЧХ для формулы второй разности отсчетов. Как видно из графика, эта формула дает результат, близкий к дифференцированию 2-го порядка, в диапазоне нормированной частоты от 0 до примерно 0,15.

С другой стороны, такая кривая соответствует АЧХ фильтра верхних частот. Это последнее свойство можно, например, использовать для оценки уровня широкополосного (миографического) шума в ЭКГ. Как известно, основная доля мощности ЭКГ сосредоточена в области частот не выше 50 Гц [2,8]. Если выбрать частоту дискретизации равной 250 Гц, то формула второй разности практически не будет пропускать полезные составляющие электрокардиосигнала, а будет усиливать частоты, лежащие в более высокочастотном диапазоне. Так как спектр широкополосной помехи примерно равномерен во всем диапазоне частот, то по мощности высокочастотной части диапазона (заведомо

не имеющей отношения к полезной ЭКГ) можно судить о содержании широкополосных помех в сигнале. На рис. 2.24 приведен пример использования второй разности отсчетов для решения такой задачи. На графике *a* показан фрагмент ЭКГ, в котором на участке от 0,65

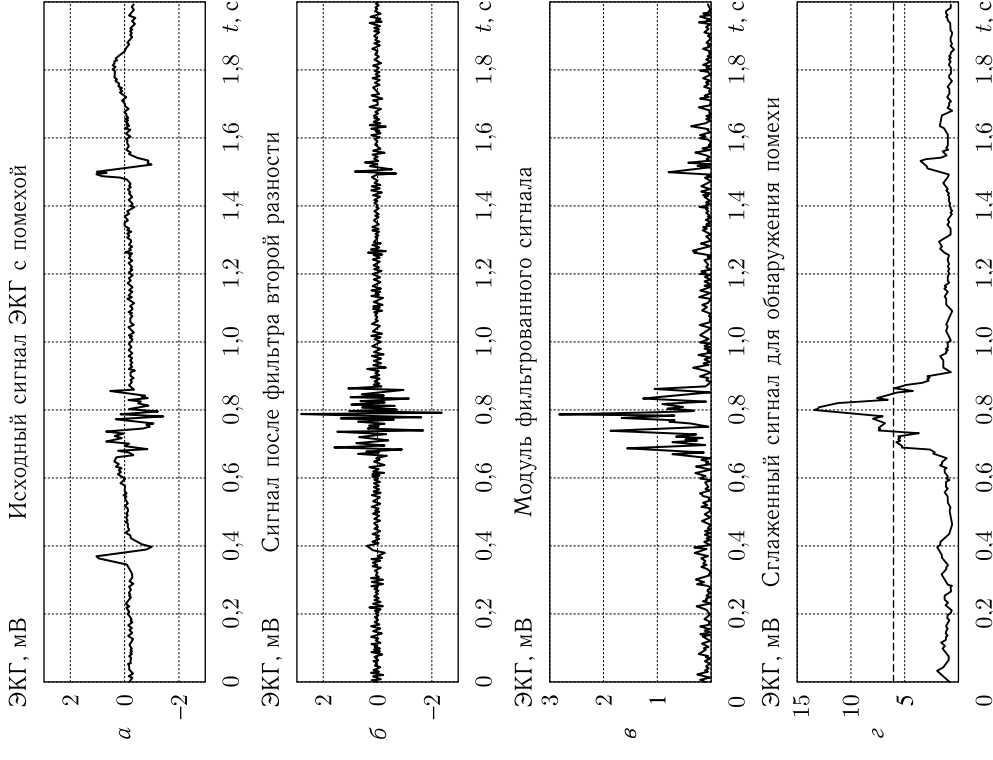


Рис. 2.24. Пример использования второй разности отсчетов для оценки интенсивности миографической помехи в ЭКГ: *a* — фрагмент ЭКГ, в котором наблюдается всплеск миографической помехи; *б*, *в* — результат применения формулы второй разности отсчетов и операции взятия модуля полученного сигнала; *г* — результат применения сглаживающего фильтра и возможный уровень порога (горизонтальная штриховая линия) для определения зашумленных участков сигнала

до 0,9 с наблюдается всплеск миографической помехи. Графики *б* и *в* демонстрируют соответственно результаты применения формулы второй разности отсчетов и операции взятия модуля полученного сигнала. Как можно видеть, формула второй разности позволила выделить ши-рокополосную помеху. Для большей устойчивости такого выделителя помех желательно далее применить фильтрацию нижних частот, для чего, например, можно использовать сглаживание по определенному числу последовательных точек:

$$y(n) = \frac{1}{N} \sum_{i=0}^{N-1} x(n-i).$$

На графике *г* рис. 2.24 показан результат применения этого фильтра для случая $N = 10$. Пунктирная горизонтальная линия на этом же графике показывает возможное положение порогового значения, по превышению которого можно определять участки сигнала, где уровень миографической помехи недопустимо высок.

Аналогично операции дифференцирования для рассмотрения операции интегрирования с позиций частотного подхода необходимо исследовать ее воздействие на гармонический сигнал [2.20].

Интеграл от гармонического сигнала определяется выражением

$$\int x(t) dt = \int e^{j\omega t} dt = \frac{1}{j\omega} e^{j\omega t}.$$

Тогда передаточную функцию операции интегрирования можно выразить как

$$H(\omega) = \frac{\int x(t) dt}{x(t)} = \frac{1}{j\omega}.$$

Модуль и аргумент данной функции (АЧХ и ФЧХ) определяются выражениями

$$|H(\omega)| = 1/\omega; \quad \arg H(\omega) = -\frac{\pi}{2}.$$

Это означает, что усиление дифференциатора по отношению к непрерывному гармоническому сигналу обратно пропорционально частоте, а вносимый фазовый сдвиг всегда равен $-\pi/2$. Кривая *1* на рис. 2.25 показывает амплитудно-частотную характеристику идеального (аналогового) интегратора от нормированной частоты.

Как известно, операция вычисления определенного интеграла соответствует нахождению площади под кривой, изображающей подынтегральную функцию. В вычислительной математике известен целый ряд методов, позволяющих численно оценить значение интеграла заданной функции. Рассмотрим наиболее распространенные из них.

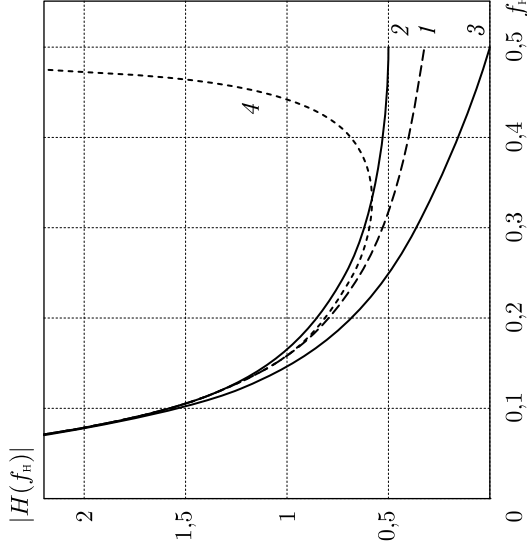


Рис. 2.25. Амплитудно-частотные характеристики идеального интегратора (1) и цифровых интегрирующих фильтров: формула прямоугольников (2); формула трапеций (3) и формула Симпсона (4)

Формула прямоугольников. Разностное уравнение, соответствующее формуле, имеет вид

$$y(n) = y(n-1) + Tx(n-1),$$

где $x(n)$ и $y(n)$ — соответственно отсчеты входной и выходной последовательностей, T — интервал дискретизации, а n — целочисленный порядковый номер отсчета.

Интеграл по данной формуле рассчитывается как сумма прямоугольников, образованных вертикальными прямыми, проходящими через точки взятия отсчетов и имеющих высоту, соответствующую значению отсчета $x(n-1)$, предшествующего текущему $x(n)$. Графики *а* и *г* на рис. 2.26 иллюстрируют принцип расчета интеграла по формуле прямоугольников. На верхнем из этих графиков штриховой кривой показана исходная непрерывная функция, на которой точками отмечены моменты взятия дискретных отсчетов. В данном случае интервал дискретизации равен $T = 0,05$ с, что соответствует частоте дискретизации $F_d = 20$ Гц. Заштрихован прямоугольник, площадь которого рассчитывается для отсчета $x(n)$, соответствующего здесь моменту времени $t = 0,2$ с. Как можно видеть, расхождение между площадью прямоугольника и площадью под соответствующим участком исходной кривой достаточно существенно. Этим объясняется заметная погрешность оценки интеграла, что иллюстрирует график *г* на рис. 2.26. Здесь сплошная линия показывает истинное значение интеграла, а точки,

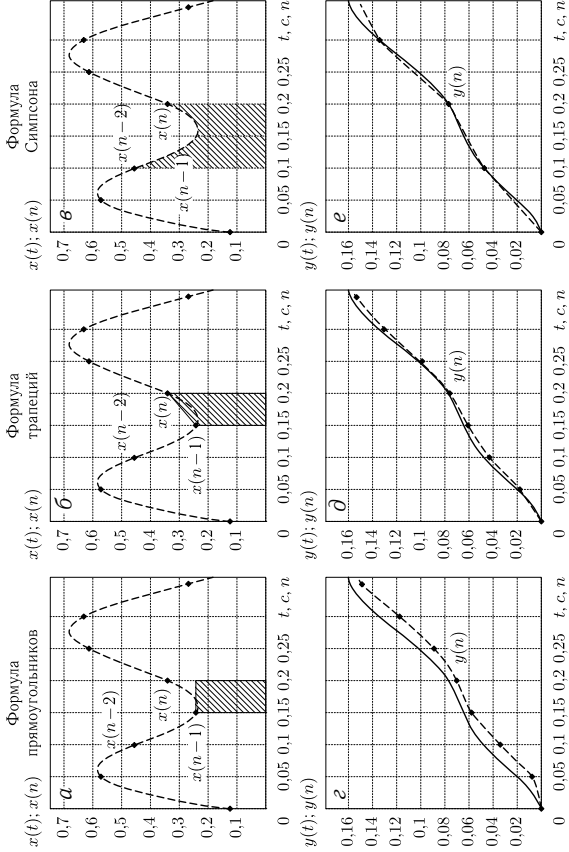


Рис. 2.26. Иллюстрация принципа вычисления интеграла по формулам: прямоугольников (*a* и *c*); трапеций (*b* и *d*) и Симпсона (*e* и *e*)

соединенные штриховой линией, соответствуют результату оценки интеграла с использованием формулы прямоугольников.

Так как формула прямоугольников представляет собой рекурсивный цифровой фильтр, то с использованием z -преобразования можно считать его передаточную функцию:

$$H(z) = T \frac{z^{-1}}{1 - z^{-1}},$$

$$H(\omega) = \frac{T}{1 - e^{-j\omega T}} = \frac{T e^{-j\omega T/2}}{e^{j\omega T/2} - e^{-j\omega T/2}} = \frac{T e^{-j\omega T/2}}{2j \sin(\omega T/2)}.$$

Соответственно АЧХ и ФЧХ фильтра описываются выражениями

$$|H(\omega)| = \frac{T}{2 \sin(\omega T/2)}; \quad \arg H(\omega) = -\frac{\pi}{2} - \frac{\omega T}{2}.$$

Кривая 2 на рис. 2.25 показывает вид АЧХ формулы прямоугольников для нормированной частоты. Из графика видно, что область совпадения данной кривой с характеристикой идеального интегратора (кривая 1) охватывает диапазон нормированной частоты от 0 до примерно 0,1. Это означает, что удовлетворительные результаты интегрирования могут быть получены только для сигналов, частотный состав которых не выходит за пределы 10 % от частоты дискретизации сигнала.

Более качественный результат может быть получен с помощью формулы трапеций, определяемой выражением

$$y(n) = y(n-1) + T[x(n-1) + x(n-1)].$$

Принцип вычисления интеграла по этой формуле и получаемый результат иллюстрируются на графиках *b* и *d* рис. 2.26. Как хорошо видно из графиков, погрешность вычисления в этом случае значительно ниже, чем для формулы прямоугольников. Найдем передаточную функцию формулы трапеций:

$$H(z) = \frac{T}{2} \frac{1 + z^{-1}}{1 - z^{-1}},$$

$$H(\omega) = \frac{T}{2} \frac{1 + e^{-j\omega T}}{1 - e^{-j\omega T}} = \frac{T}{2} \frac{e^{j\omega T/2} + e^{-j\omega T/2}}{e^{j\omega T/2} - e^{-j\omega T/2}} = \frac{T}{2} \frac{2 \cos(\omega T/2)}{2j \sin(\omega T/2)} = \frac{T}{2j} \operatorname{ctg}(\omega T/2).$$

Соответственно АЧХ и ФЧХ фильтра описываются выражениями

$$|H(\omega)| = \frac{T}{2} \operatorname{ctg}(\omega T/2); \quad \arg H(\omega) = -\frac{\pi}{2}.$$

Кривая 3 на рис. 2.25 показывает вид АЧХ формулы трапеций для нормированной частоты.

Одной из наиболее популярных формул численного интегрирования является формула Симпсона

$$y(n) = y(n-2) + \frac{T}{3} [x(n-1) + 4x(n-1) + x(n-2)].$$

Формула основана на интерполяции сигнала полиномом второй степени по трем последовательным точкам. Принцип вычисления интеграла по этой формуле и получаемый результат иллюстрируются на графиках «*e*» и «*e*» рис. 2.26. В данном случае видно, что достигается очень хорошее приближение к истинному результату. Однако результат получается с частотой отсчетов вдвое ниже исходной, так как для вычисления каждого следующего значения используются три подряд идущих отсчета. Найдем передаточную функцию формулы Симпсона:

$$H(z) = \frac{T}{3} \frac{1 + 4z^{-1} + z^{-2}}{1 - z^{-2}},$$

$$H(\omega) = \frac{T}{3} \frac{1 + 4e^{-j\omega T} + e^{-2j\omega T}}{1 - e^{-2j\omega T}} = \frac{T}{2} \frac{e^{j\omega T} + 4 + e^{-j\omega T}}{e^{j\omega T} - e^{-j\omega T}} =$$

$$= \frac{T}{3} \frac{4 + 2 \cos(\omega T)}{2j \sin(\omega T)} = \frac{T}{3} \frac{2 + \cos(\omega T)}{j \sin(\omega T)}.$$

Соответственно АЧХ и ФЧХ для формулы Симпсона описываются выражениями

$$|H(\omega)| = \frac{T}{3} \frac{2 + \cos(\omega T)}{\sin(\omega T)}; \quad \arg H(\omega) = -\frac{\pi}{2}.$$

Кривая 4 на рис. 2.25 показывает вид АЧХ данного фильтра для нормированной частоты. Как видно из графика, АЧХ формулы Симпсона демонстрирует очень хорошее совпадение с характеристикой идеального интегратора для значений нормированной частоты от 0 до 0,2, но начиная с частоты примерно 0,3 она резко уходит вверх. На практике это означает, что формулу Симпсона нельзя использовать для расчета интеграла от сигналов, содержащих компоненты с частотой выше 25 % от частоты дискретизации.

Операция цифровой интеграции используется во многих задачах компьютерной обработки биомедицинских сигналов и данных. В частности, в монографии [2.20] описан метод оценки сердечного выброса (объема крови, выталкиваемого сердцем за одно сокращение) с использованием контрастного вещества, вводимого в вену пациента через катетер. Концентрация этого вещества регистрируется датчиком, введенным в аорту. Интегрирование получаемой кривой концентрации контрастного вещества позволяет получить величину, прямо пропорциональную сердечному выбросу, по которой и рассчитывается данный показатель.

Также в разд. 4.1 описана методика анализа вариабельности сердечного ритма по последовательности интервалов времени между сокращениями сердца. Расчет важнейших спектральных показателей вариабельности (значений суммарной мощности в трех определенных стандартом диапазонах частот) осуществляется путем интегрирования функции спектральной плотности мощности в указанных частотных диапазонах. Отметим, что в данном случае наиболее корректный подход — использование метода интегрирования по формуле прямоугольников, что определяется взаимосвязью между спектральной плотностью мощности и суммарной мощностью сигнала (см. разд. 2.3).

Применение цифровой фильтрации в медицинских приборах и системах

Цифровая фильтрация широко применяется при автоматической обработке биомедицинских сигналов. С использованием ЦФ решаются следующие основные задачи:

- устранение помех;
- выделение информативных составляющих;
- анализ сигналов по частотным диапазонам.

При выборе процедур цифровой фильтрации необходимо исходить как из характеристик обрабатываемого сигнала и помех, так и из конечных задач анализа сигналов.

Рассмотрим пример использования ЦФ при анализе электроэнцефалограммы, т.е. записи электрических потенциалов мозга, снимаемых с поверхности головы человека. Выделяют четыре основных ритма ЭЭГ [2.5]:

- δ (дельта) с частотой 0,5–3 Гц и амплитудой 40–300 мкВ;
- θ (тета) с частотой 4–6 Гц и амплитудой 40–300 мкВ;
- α (альфа) с частотой 8–13 Гц и амплитудой до 100 мкВ;
- β (бета) с частотой 14–40 Гц и амплитудой до 15 мкВ.

Использование полосовой цифровой фильтрации позволяет выделить каждый из ритмов отдельно, после чего по полученным сигналам вычисляются некоторые численные показатели, характеризующие данный ритм. На рис. 2.27 слева показаны исходные сигналы ЭЭГ по 19 каналам, а справа — эти же сигналы, пропущенные через полосовой ЦФ, рассчитанный на выделение полосы частот, соответствующей

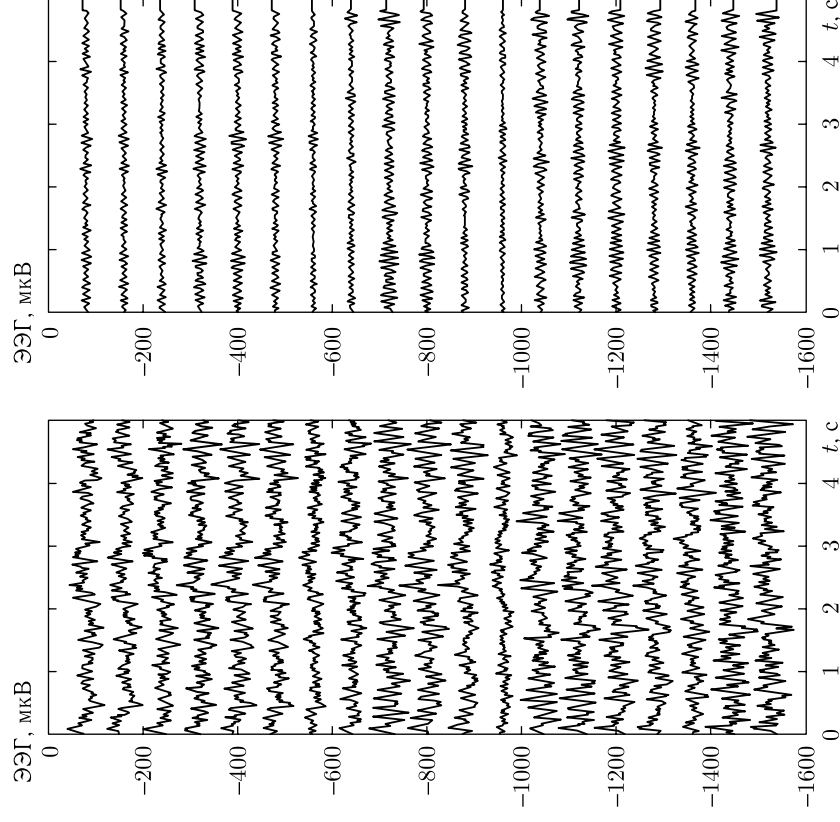


Рис. 2.27. Слева — исходные сигналы ЭЭГ по 19 каналам; справа — эти же сигналы, пропущенные через фильтр, выделяющий α -ритм

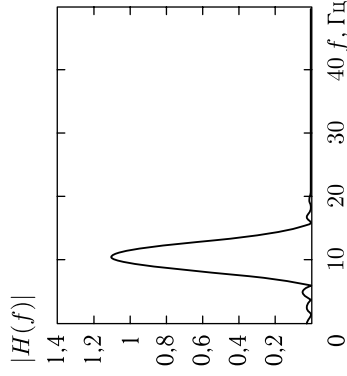


Рис. 2.28. АЧХ полосового ЦФ, рассчитанного на выделение полосы частот, соответствующей α -ритму ЭЭГ

α -ритму ЭЭГ. На рис. 2.28 показана АЧХ данного фильтра. Можно видеть, что в результате фильтрации в сигнале отчетливо проявились участки, где присутствует α -ритм.

В гл. 3 подробно рассмотрены также примеры применения ЦФ для автоматической обработки и анализа ЭКГ в системах непрерывного контроля сердечного ритма.

2.3. Цифровые методы спектрального оценивания сигналов

Основные понятия и определения

При исследовании сигналов часто используются методы *спектрального анализа*, позволяющие получить численные оценки частотного состава сигнала. Спектральный анализ основан на *частотном подходе* к анализу сигналов (см. разд. 2.1).

Наиболее распространены методики спектрального анализа, основанные на дискретном преобразовании Фурье (ДПФ)

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)nk},$$

где $x(n)$ — отсчеты дискретного сигнала; N — число отсчетов; $k = 0, 1, \dots, N-1$ — номера частотных составляющих разложения. Для ускорения расчета применяют алгоритмы быстрого преобразования Фурье (БПФ) [2.1, 2.19, 2.20].

В результате выполнения процедуры ДПФ получается ровно N значений комплексных чисел, описывающих гармонические колебания (синусоиды) с частотами $0, f_0, 2f_0, 3f_0, \dots, (N-1)f_0$, где $f_0 = f_N/N$ — основная частота преобразования (f_N — частота дискретизации сигнала). Амплитуда и фазовый сдвиг каждой из синусоид определяются

как модуль и аргумент соответствующего комплексного элемента разложения.

По полученному разложению исходный сигнал может быть полностью восстановлен (в точках взятия отсчетов) с использованием процедуры обратного дискретного преобразования Фурье.

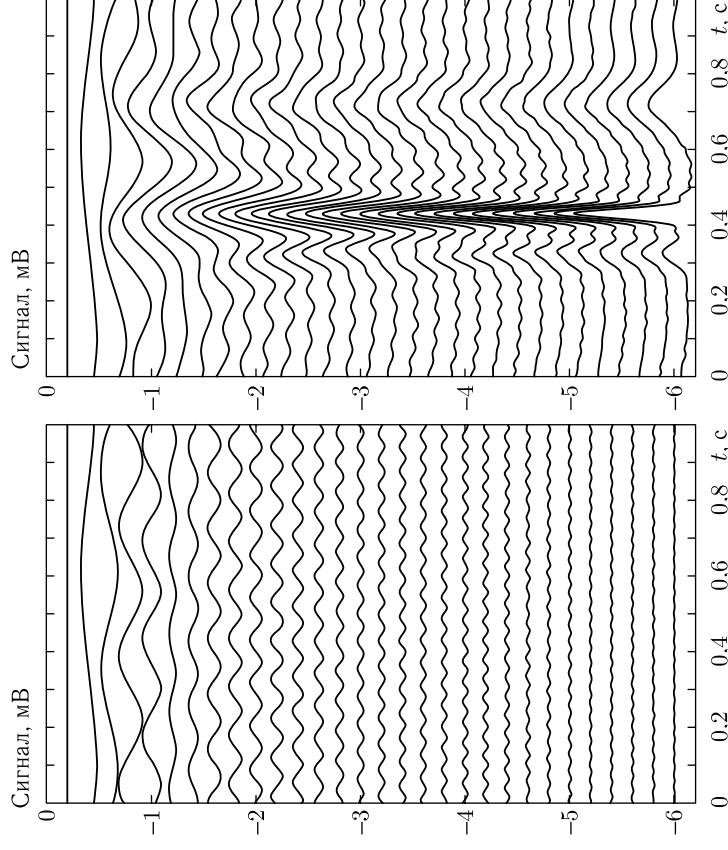


Рис. 2.29. Иллюстрация процесса реконструкции фрагмента ЭКГ при помощи синусоид, определяемых значениями элементов ДПФ

На рис. 2.29 иллюстрируется процесс реконструкции фрагмента ЭКГ при помощи синусоид, определяемых значениями элементов ДПФ. Левые графики показывают синусоиды со значениями частот $0, f_0, 2f_0, 3f_0, \dots, (N-1)f_0$, а правые — этапы реконструкции сигнала с помощью этих синусоид. Рисунок позволяет проследить, как суммирование синусоидальных составляющих дает возможность получить описание кривой практически произвольной формы.

Спектральное оценивание на основе дискретного преобразования Фурье

В случае использования алгоритма БПФ для получения спектральных оценок возможны два основных варианта: расчет БПФ для всего

фрагмента сигнала целиком (периодограммный метод) или использование усреднения спектров, рассчитанных по перекрывающимся фрагментам сигнала (метод Уэлча).

При использовании периодограммного метода выполняется следующая последовательность процедур [2.19, 2.25].

1. Удаление из сигнала среднего значения (или линейного тренда) с целью устранения постоянной составляющей.
2. Умножение сигнала на сглаживающее окно (см. далее) для уменьшения спектральной утечки, вызываемой разрывами на краях анализируемого фрагмента, возникающими при его периодическом продолжении.
3. Дополнение фрагмента сигнала нулями до размера, соответствующего какой-либо целой степени числа 2, что позволяет далее использовать алгоритм БПФ.
4. Расчет дискретного преобразования Фурье с помощью алгоритма БПФ.
5. Расчет спектральной плотности мощности (СПМ):

$$p(f_k) = \frac{1}{N f_a} \sum_{f_k=0}^{0.5} \{ [\operatorname{Re} x(f_k)]^2 + [\operatorname{Im} X(f_k)]^2 \},$$

где f_a — частота дискретизации, $f_k = k\Delta f = k f_a / N$, $k = 0, 1, 2, \dots, N/2$, а $X(f_k)$ и $p(f_k)$ — соответственно значения компонент разложения по Фурье и СПМ для частот f_k .

6. Умножение СПМ на корректирующий коэффициент, учитывающий потери мощности при использовании сглаживающего окна:

$$k_w = \frac{1}{\frac{1}{N_w} \sum_{n=1}^{N_w} w^2(n)},$$

где N_w — размер окна в отсчетах, а $w(n)$ — отсчеты оконной функции.

7. Умножение СПМ на корректирующий коэффициент, учитывающий потери мощности при дополнении нулями:

$$k_0 = N/N_0,$$

где N_0 — число отсчетов фрагмента сигнала до дополнения нулями.

При использовании метода Уэлча [2.19, 2.25] исходный сигнал разбивается на сегменты, для каждого из которых спектральная оценка рассчитывается при помощи описанной выше процедуры. Далее полученные оценки усредняются по соответствующим частотам. Оценки спектральной плотности усредняются по соответствующим частотам. Оценки Уэлча, отличаются большей статистической устойчивостью по сравнению с простым периодограммным методом.

Из-за того что при использовании ДПФ неявно предполагается бесконечное периодическое продолжение анализируемого сигнала,

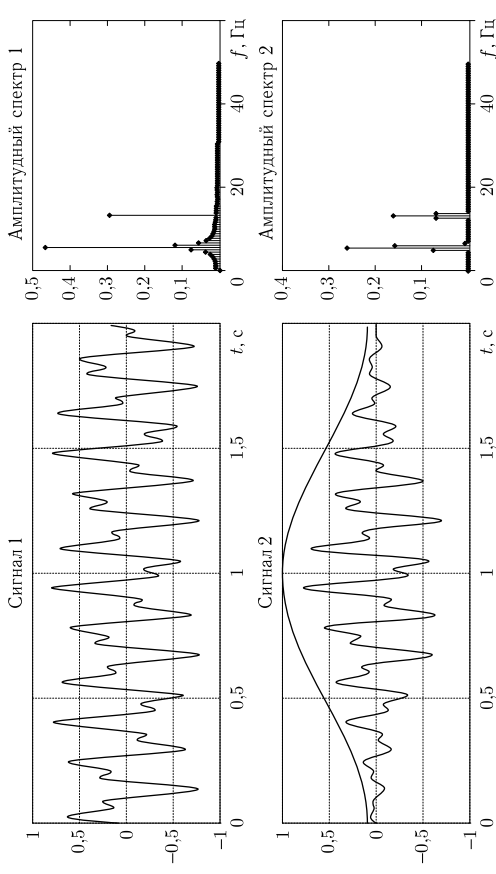


Рис. 2.30. Снижение спектральной утечки при использовании сглаживающего окна

в получаемом спектре может наблюдаться эффект, называемый *спектральной утечкой*, который проявляется в «растекании» мощности спектральных составляющих по диапазону частот. Этот эффект может быть снижен за счет применения так называемых *сглаживающих окон*. Под последними понимаются функции, имеющие на краях близкие к нулю значения, которые плавно возрастают до единицы в средней части функции. На рис. 2.30 демонстрируется эффект спектральной утечки и результат ее снижения с использованием сглаживающего окна. На левом верхнем графике показан фрагмент сигнала, представляющего собой смесь двух синусоид (с частотами 7 и 13 Гц). Справа сверху показан амплитудный спектр этого сигнала, где для составляющей с частотой 7 Гц наблюдается спектральная утечка. На нижнем левом графике показаны сглаживающая функция и сигнал, полученный в результате умножения исходного сигнала на эту функцию. Как можно видеть, значения последней функции приближаются к нулевым на концах фрагмента. Правый нижний график содержит амплитудный спектр полученного сигнала, в котором эффект спектральной утечки существенно снижен. Однако следует иметь в виду, что применение сглаживающих окон, хоть и снижает спектральную утечку, но приводит в то же время к ухудшению спектрального разрешения, т.е. способности спектра к различению близких по частоте составляющих.

Параметрические методы спектрального оценивания

Альтернативный подход к расчету спектральных оценок реализуется с использованием так называемых *параметрических* методов.

Наиболее распространенные из них основаны на модели авторегрессии (АР), описываемой уравнением

$$x(n) = u(n) - \sum_{k=1}^p a_k x(n-k),$$

где $x(n)$ — отсчеты процесса авторегрессии, $u(n)$ — отсчеты возбуждающего белого шума, a_k — коэффициенты модели, а p — порядок модели.

Если по последовательности отсчетов некоторого анализируемого сигнала удастся найти набор коэффициентов авторегрессии a_k и получить оценку дисперсии возбуждающего белого шума ρ_w , то оценка спектральной плотности мощности (СПМ) данного сигнала будет определяться выражением [2.19, 2.25]

$$P(f) = T\rho_w \left[\frac{1}{1 + \sum_{k=1}^p a_k e^{-j2\pi f k T}} \right],$$

где $T = 1/f_d$ — период дискретизации сигнала, а f — частота соответствующего значения СПМ.

Идея авторегрессионного метода иллюстрируется на рис. 2.31. Пусть требуется оценить спектральную плотность мощности некоторого наблюдаемого дискретного сигнала $y(n)$. Подадим на вход рекурсивного цифрового фильтра, имеющего набор коэффициентов $a_1, a_2, \dots, a_p$ и передаточную функцию $H(f)$, воздействие $u(n)$ в виде отсчетов белого шума (абсолютно случайного сигнала). Тогда в случае если удастся так подобрать коэффициенты фильтра, что средняя энергия ошибки $e(n)$ окажется минимальной, можно считать, что передаточная функция фильтра $H(f)$ (с точностью до множителя ρ_w , представляющего собой оценку дисперсии возбуждающего белого шума $u(n)$) является оценкой спектральной плотности мощности сигнала $y(n)$ [2.19].

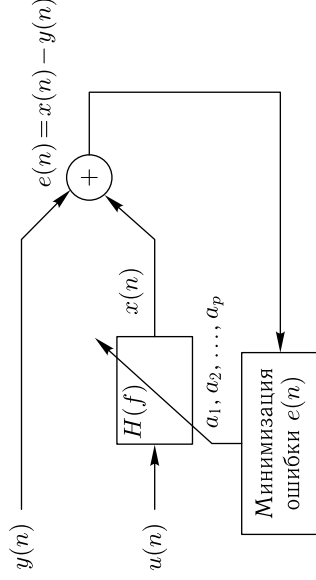


Рис. 2.31. Иллюстрация авторегрессионного метода получения спектральных оценок

Разработан целый ряд алгоритмов оценки параметров авторегрессии (коэффициентов $a_1, a_2, \dots, a_p$ и параметра ρ_w), из которых наиболее известны следующие четыре [2.19]:

- автокорреляционный метод Юла–Уокера;
- метод максимальной энтропии Берга;
- ковариационный метод;
- модифицированный ковариационный метод.

Из перечисленных алгоритмов наилучшие результаты дают последние два, основанные на методах линейного предсказания. В то же время эти методы наиболее сложны в вычислительном отношении.

Процедура получения спектральных оценок с помощью авторегрессионных методов может быть описана следующей последовательностью этапов.

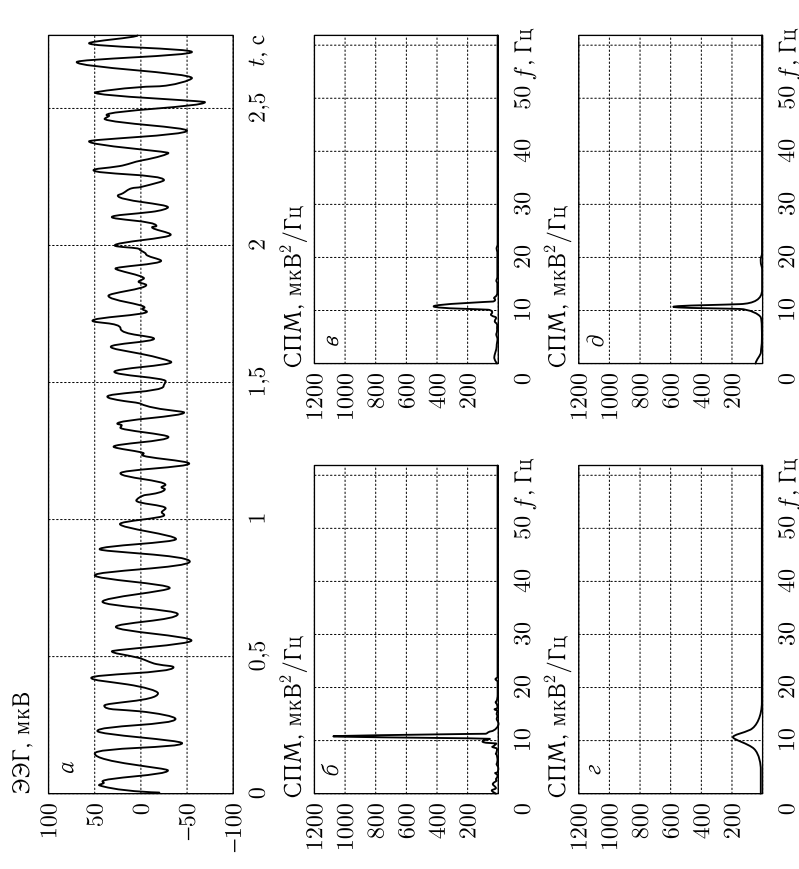


Рис. 2.32. Пример расчета СПМ альтернативными методами спектрального анализа: а — фрагмент ЭЭГ, содержащий выраженный α -ритм; б — СПМ, рассчитанная периодограммным методом; в — СПМ, рассчитанная методом Уэлча; г — СПМ, рассчитанная АР-методом с порядком, равным 8; д — СПМ, рассчитанная АР-методом с порядком, равным 15

Таблица 2.2

Сравнение альтернативных методов спектрального анализа

	Методы спектрального анализа	
	Классические (на основе ДПФ)	Параметрические (авторегрессионные)
Достоинства	Математическая простота. Обратимость ДПФ.	Более реалистические, чем в случае классических методов, предположения о сигнале за пределами анализируемого фрагмента. Возможность анализа спектра по относительно коротким фрагментам сигнала.
Недостатки	Появление спектральной утечки вследствие предполагаемой периодичности сигнала (т. е. нереалистического предположения о сигнале за пределами анализируемого фрагмента). Необходимостью наличия относительно продолжитель- ного фрагмента сигнала.	Математическая сложность. Необратимость спектра. Необходимость субъективного выбора порядка АР модели.

1. Удаление из сигнала среднего значения (или линейного тренда) с целью устранения постоянной составляющей.
2. Выбор порядка модели p .
3. Оценивание параметров АР модели $(a_1, a_2, \dots, a_p \text{ и } \rho_w)$.
4. Вычисление оценки СПМ $P(f)$.
5. Анализ полученных результатов и при необходимости выбор нового значения порядка p (т. е. возвращение к шагу 2).

Последний из этапов выполняется до тех пор, пока не будет достигнут приемлемый компромисс между частотным разрешением СПМ и дисперсией получаемых спектральных оценок [2.19]. Чем больше выбранный величина порядка p , тем выше разрешение получаемой оценки СПМ (т. е. возможность различения близких по частоте составляющих), но при этом большим оказывается и разброс спектральных оценок, т. е. статистическая устойчивость результатов анализа снижается. Как видим, применение данных методов предполагает необходимость контроля результатов человеком-исследователем, и это один из главных недостатков параметрических методов.

На рис. 2.32 показан фрагмент сигнала ЭЭГ и примеры оценок СПМ, рассчитанных альтернативными методами. Как можно видеть, полученные оценки СПМ существенно отличаются друг от друга. В то же время на каждом из графиков отчетливо виден пик на частоте чуть выше 10 Гц, соответствующий α -ритму. Необходимо отметить, что хотя высота пиков на графиках СПМ различна, суммарная мощность, равная площади под кривой СПМ, во всех случаях одна и та же.

В табл. 2.2 дана сравнительная характеристика методов на основе ДПФ и параметрических методов. В практических задачах необходимо выбирать конкретный метод расчета спектральных оценок, учитывая как частотные свойства сигнала, так и конечные цели и условия анализа сигнала.

Взаимная спектральная плотность мощности и функция когерентности

Взаимосвязь колебательных процессов, присутствующих в двух одновременно наблюдаемых сигналах, наиболее полно позволяет отразить их взаимный спектр, который содержит информацию как о мощности совместных колебаний, так и о фазовых сдвигах между колебаниями с одинаковыми частотами [2.4, 2.25]. Функция взаимной спектральной плотности мощности (ВСПМ) может быть получена как преобразование Фурье от взаимной корреляционной функции (ВКФ) двух сигналов [2.25]:

$$P_{xy}(f) = \text{FT}[C_{xy}(\tau)] = X^*(f)Y(f),$$

где $P_{xy}(f)$ — значение ВСПМ сигналов x и y для частоты f , $C_{xy}(\tau)$ — ВКФ для временного сдвига τ , а $X(f)$ и $Y(f)$ — преобразования Фурье для соответствующих сигналов.

На практике при цифровом анализе сигналов оценка ВСПМ может быть получена различными способами, из которых наиболее известны следующие [2.4, 2.19, 2.25].

1. Непосредственное вычисление ВСПМ по ДПФ двух синхронно снятых дискретных выборок сигналов $x(n)$ и $y(n)$, где $n = 0, 1, \dots, N - 1$.

Если ДПФ сигналов определены как

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk}, \quad k = 0, \dots, N - 1,$$

и

$$Y(k) = \sum_{n=0}^{N-1} y(n)e^{-j\frac{2\pi}{N}nk}, \quad k = 0, \dots, N - 1,$$

то оценка ВСПМ может быть рассчитана как

$$P_{xy}(k) = \frac{2}{N^2} X^*(k)Y(k), \quad k = 0, \dots, N - 1.$$

2. Получение оценки ВСПМ как ДПФ от оценки ВКФ.

В этом случае сначала вычисляется смещенная оценка взаимной корреляционной функции двух сигналов:

$$C_{xy}(m) = \begin{cases} \frac{1}{N\sigma_x\sigma_y} \sum_{n=0}^{N-m-1} x(n+m)y(n), & 0 \leq m \leq N-1, \\ \frac{1}{N\sigma_x\sigma_y} \sum_{n=0}^{N-|m|-1} x(n)y(n+m), & -(N-1) \leq m < 0, \end{cases}$$

где m — значение сдвига, а σ_x и σ_y — стандартные отклонения соответствующих сигналов. Тогда оценка ВСПМ может быть рассчитана по формуле

$$P_{xy} = \frac{1}{2} \sum_{m=-(N-1)}^{N-1} C_{xy}(m) e^{-j \frac{2\pi}{2N-1} mk}, \quad k = 0, \dots, 2(N-1).$$

3. Расчет оценки ВСПМ с использованием одного из распространенных методов практического спектрального анализа — периодограммного метода Уэлча, который включает следующие этапы:

- разбиение анализируемых фрагментов на перекрывающиеся сегменты одинаковой длительности;
- вычисление ВСПМ для каждого из сегментов;
- усреднение ВСПМ по всем сегментам.

Далее для любого из трех методов модуль ВСПМ и ее аргумент (взаимный фазовый спектр, ВФС) могут быть рассчитаны по формулам

$$|P_{xy}(k)| = \sqrt{[\operatorname{Im} P_{xy}(k)]^2 + [\operatorname{Re} P_{xy}(k)]^2}$$

и

$$\arg P_{xy}(k) = \operatorname{arctg} \left[\frac{\operatorname{Im} P_{xy}(k)}{\operatorname{Re} P_{xy}(k)} \right].$$

Полученное значение $|P_{xy}(k)|$ имеет размерность произведения разностей обих сигналов, а значение $\arg P_{xy}(k)$ выражается в радианах. Обычно бывает удобнее представлять фазовый сдвиг в градусах:

$$\arg P_{xy}^{\circ}(k) = \arg P_{xy}(k) \frac{180^{\circ}}{\pi}.$$

Временной сдвиг между сигналами для какой-либо определенной частоты $f_k = k f_N / N$ (где f_N — частота дискретизации, а N — число элементов ДПФ) может быть определен как

$$\Delta T(f_k) = \frac{\arg P_{xy}(f_k)}{2\pi f_k},$$

если ВФС выражен в радианах, или

$$\Delta T(f_k) = \frac{\arg P_{xy}^{\circ}(f_k)}{360^{\circ} f_k},$$

если ВФС выражен в градусах.

Важно отметить, что порядок вычисления взаимного спектра влияет на знак получаемого фазового сдвига. Положительным значениям фазы будет соответствовать запаздывание первого из сигналов относительно второго и наоборот.

Помимо ВСПМ часто рассматривается функция когерентности, определяемая выражением [2.4, 2.25]

$$\Gamma_{xy}(f) = \frac{P_{xy}(f)^2}{P_{xx}(f)P_{yy}(f)}.$$

Функция когерентности представляет собой нормированный вариант ВСПМ, и ее модуль может принимать значения не более единицы. Это может оказаться удобным для оценки относительной степени взаимосвязанности двух процессов на определенной частоте.

В разделе 4.1 представлен пример применения ВСПМ для решения задачи совместного анализа сигналов сердечного ритма и артериального давления.

Применение методов спектрального оценивания при анализе биомедицинских сигналов

Методы спектрального анализа находят широкое применение для исследования частотного состава таких биомедицинских сигналов, как электрокардиограмма, электроэнцефалограмма, электромиограмма, фонокардиограмма и ряда других [2.1, 2.5, 2.20, 2.25].

Рассмотрим пример использования данных методов для анализа ЭЭГ, которая представляет собой запись электрических потенциалов мозга, снимаемых с поверхности головы человека. Выделяют четыре основных ритма ЭЭГ [2.5]:

δ (дельта)	с частотой 0,5–3 Гц и амплитудой 40–300 мкВ;
θ (тета)	с частотой 4–6 Гц и амплитудой 40–300 мкВ;
α (альфа)	с частотой 8–13 Гц и амплитудой до 100 мкВ;
β (бета)	с частотой 14–40 Гц и амплитудой до 15 мкВ.

Рассчитав спектральную оценку сигнала ЭЭГ, можно получить количественные характеристики каждого из ритмов в отдельности. На рис. 2.33 показан пример получения спектральных оценок для двух синхронно снятых каналов ЭЭГ человека. Справа приведены графики фрагментов сигналов, а слева — графики рассчитанных по ним оценок СПМ. Верхние графики соответствуют сигналам с лобной части головы, а нижние — с затылочной. На верхнем правом графике (б) отчетливо виден пик, соответствующий выраженному альфа-ритму с частотой около 10 Гц. Этот ритм наблюдается у здоровых людей в состоянии покоя с закрытыми глазами и наиболее сильно проявляется в затылочных отведениях. В то же время в спектре ЭЭГ, снятой с электрода, размещенного в лобной части головы, вклад альфа-ритма на порядок меньше (график г).

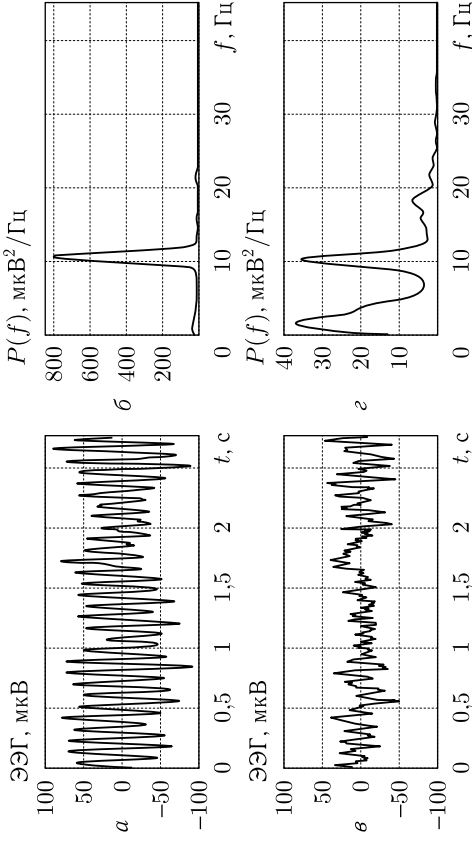


Рис. 2.33. Фрагменты сигнала ЭЭГ (а, в) и соответствующие им оценки СПМ (б, г)

Для получения количественных оценок частотных свойств ЭЭГ по рассчитанной спектральной плотности мощности для заданного частотного диапазона ЭЭГ вычисляются следующие основные численные показатели [2.5]:

A_{cp} — средняя амплитуда;

A_{max} — максимальная амплитуда;

F_{cp} — средневзвешенная частота;

F_{max} — частота максимальной амплитуды спектра.

Если общее число элементов спектрального разложения равно N , а частота дискретизации f_d , то шаг спектра по оси частот составит $\Delta f = f_d/N$. При этом частота, соответствующая элементу спектра с номером k , будет определяться выражением $f_k = k \cdot \Delta f$, а соответствующее значение СПМ может быть обозначено как $P(f_k)$. Если задать границы какого-либо частотного диапазона ЭЭГ номерами соответствующих составляющих спектра k_1 и k_2 , то первые три из перечисленных выше спектральных показателей могут быть рассчитаны по формулам

$$A_{cp} = \frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} P(f_k), \quad A_{max} = \max_{k=k_1}^{k_2} P(f_k),$$

$$F_{cp} = \frac{\sum_{k=k_1}^{k_2} f_k \cdot P(f_k)}{\sum_{k=k_1}^{k_2} P(f_k)}.$$

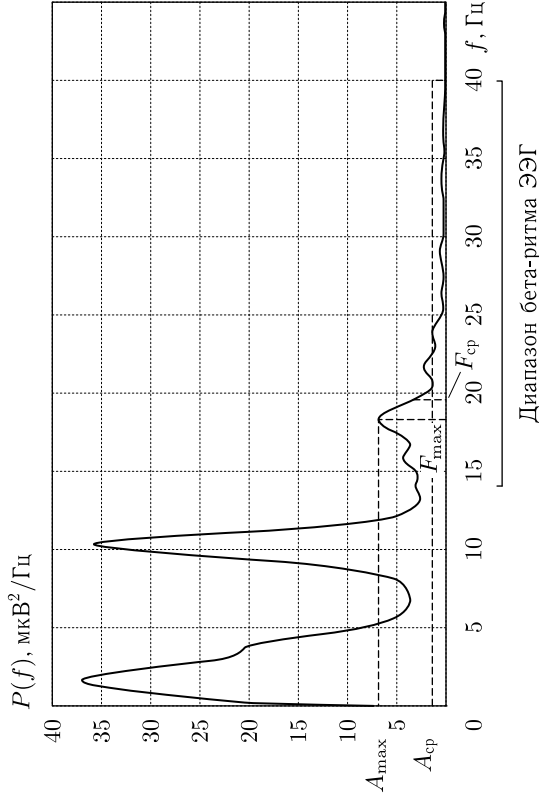


Рис. 2.34. Иллюстрация расчета спектральных показателей для частотного диапазона, соответствующего бета-ритму ЭЭГ

Показатель F_{max} определяется как частота, соответствующая значению A_{max} .

Смысл описанных спектральных показателей для частотного диапазона, соответствующего бета-ритму ЭЭГ (14–40 Гц), проиллюстрирован на рис. 2.34.

Еще одно применение частотного анализа ЭЭГ — автоматический контроль глубины анестезии в ходе хирургических операций. Используемый для решения этой задачи метод подробно описан в разд. 4.3.

При автоматическом анализе ЭКГ спектральные методы используются для решения целого ряда диагностических задач, среди них: обнаружение опасных аритмий по спектральному описанию электрокардиосигнала (см. 3.4), классификация форм желудочковых комплексов ЭКГ (см. 3.6), анализ вариабельности сердечного ритма (см. 4.1).

2.4. Получение цифровых оценок корреляционных функций

Цифровые оценки автокорреляционной и взаимной корреляционной функций

При анализе сигналов и процессов широкое применение находят *автокорреляционная функция* (АКФ) и *взаимная корреляционная функция* (ВКФ). Первая позволяет судить о наличии внутренних периодичностей в сигнале, а вторая — об общих периодичностях в двух

сигналах. Оценка АКФ для дискретного сигнала рассчитывается по формуле [2.1]

$$R_{xx}(m) = \frac{1}{N\sigma_x^2} \sum_{n=0}^{N-|m|-1} x(n+m)x(n), \quad m = -(N-1), \dots, N-1, \quad (2.2)$$

где $x(n)$ — отсчеты сигнала, N — число отсчетов, m — число сдвигов, σ_x — оценка среднеквадратичного отклонения сигнала $x(n)$.

В свою очередь оценка ВКФ для двух дискретных сигналов рассчитывается с использованием следующих выражений [2.1]:

$$R_{xy}(m) = \begin{cases} \frac{1}{N\sigma_x\sigma_y} \sum_{n=0}^{N-m-1} x(n)y(n+m), & m = 0, 1, \dots, N-1; \\ \frac{1}{N\sigma_x\sigma_y} \sum_{n=0}^{N+m-1} x(n-m)y(n), & m = -1, -2, \dots, -(N-1), \end{cases} \quad (2.3)$$

где $x(n)$ и $y(n)$ — отсчеты двух сигналов, N — число отсчетов для каждого из них, m — число сдвигов, σ_x и σ_y — оценки среднеквадратичных отклонений этих сигналов.

Величины, рассчитанные по формулам (2.1) и (2.2), называются нормированными смещенными оценками АКФ и ВКФ. Их значения лежат в пределах от -1 до $+1$. АКФ показывает наличие внутренних периодичностей в сигнале. Она всегда равна единице при нулевом сдвиге (т.е. при $m = 0$), симметрична относительно оси ординат и затухает с увеличением m тем медленнее, чем сильнее выражены периодичности в сигнале. ВКФ позволяет оценить наличие сходных по частоте колебаний в двух сигналах. Если при каких-то значениях сдвига модуль значения ВКФ приближается к единице, то это служит признаком наличия в сигналах однотипных изменений, происходящих с задержкой, равной сдвигу.

При анализе медико-биологических сигналов часто применяется коэффициент взаимной корреляции (КВК), который по сути представляет собой значение ВКФ при нулевом сдвиге. Оценка КВК для двух дискретных сигналов $x(n)$ и $y(n)$ рассчитывается следующим образом:

$$R_{xy} = \frac{1}{N\sigma_x\sigma_y} \sum_{n=0}^{N-1} x(n)y(n).$$

Значение КВК, рассчитанное по данной формуле, лежит в пределах от -1 до $+1$. Если модуль КВК приближается к единице, то это свидетельствует о существовании какого-то вида линейной связи между сигналами. Если модуль КВК близок к нулю, то это признак отсутствия линейной зависимости между сигналами.

Применение корреляционных методов при анализе биомедицинских сигналов

АКФ и ВКФ широко используются при анализе электроэнцефалограмм (ЭЭГ). АКФ позволяет выявить присутствие ритмов в ЭЭГ, а ВКФ дает возможность оценить взаимосвязь электроэнцефалографических событий в сигналах ЭЭГ, снятых в разных зонах на поверхности головы.

На рис. 2.35 приведен пример расчета АКФ (графики g, δ, e) и ВКФ (графики $ж, з, и$) для трех синхронно зарегистрированных фрагментов ЭЭГ человека (графики $a, б, в$). ЭЭГ на графиках a и $б$ сняты с затылочной области головы и содержат выраженный альфа-ритм, проявляющийся в виде колебаний почти синусоидальной формы с частотой примерно 10 Гц. Нижняя ЭЭГ ($в$) получена с лобной области головы и носит характер хаотического процесса. Из графиков АКФ видно, что присущая первым двум сигналам периодичность приводит к тому, что их автокорреляционная функция затухает относительно медленно. В то же время АКФ для третьего сигнала затухает очень быстро, что демонстрирует его случайный характер. ВКФ сигналов 1 и 2 затухает медленно, что показывает наличие в этих сигналах сходных периодичностей. ВКФ, рассчитанная для двух других пар сигналов, принимает

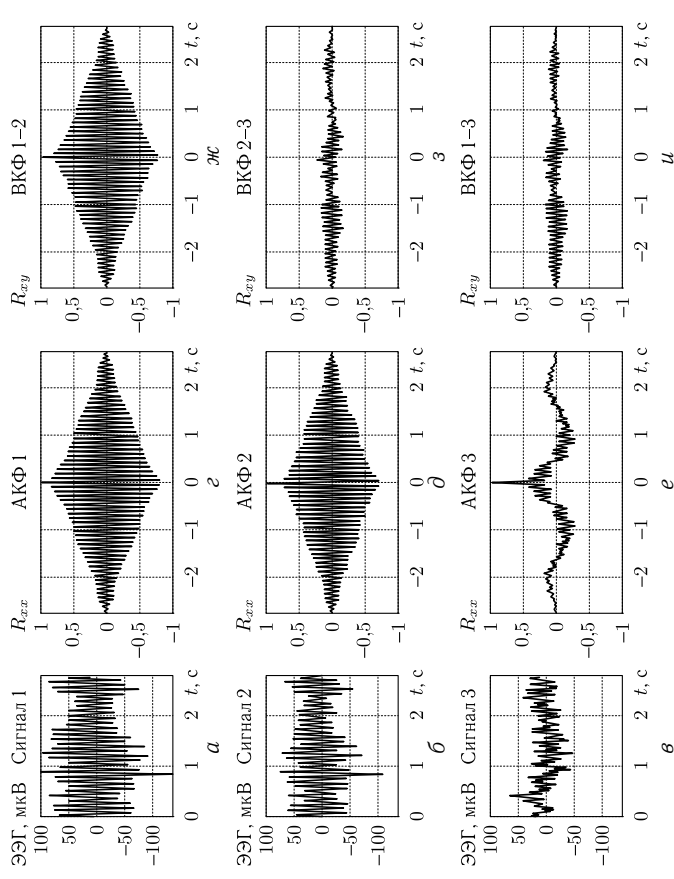


Рис. 2.35. Пример расчета АКФ и ВКФ для трех фрагментов ЭЭГ

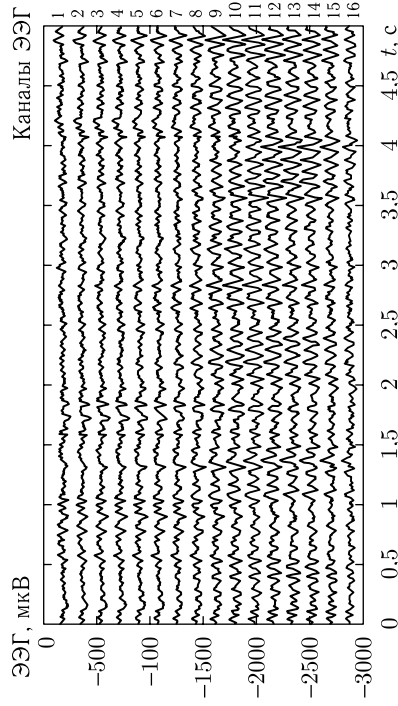


Рис. 2.36. Графики синхронно снятых 16 каналов ЭЭГ

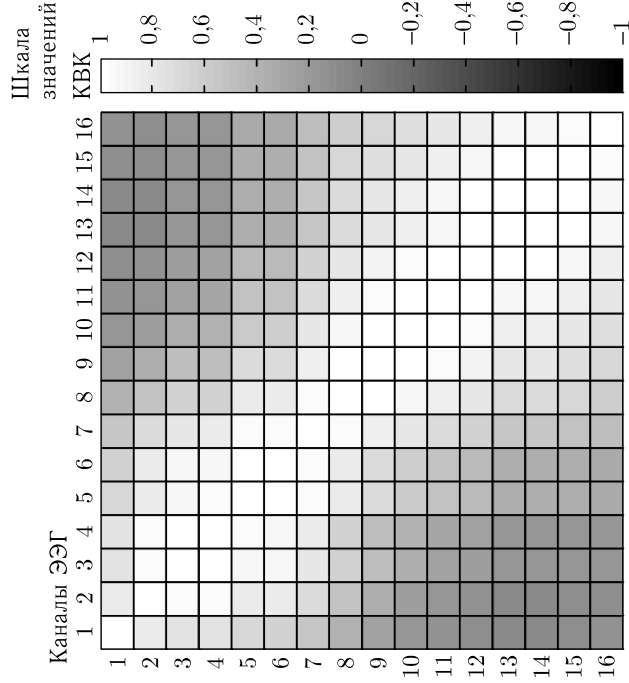


Рис. 2.37. Графическое изображение матрицы коэффициентов взаимной корреляции

относительно небольшие значения, что свидетельствует об отсутствии существенной связи между соответствующими сигналами.

Коэффициент взаимной корреляции часто используется для оценки степени связанности сигналов ЭЭГ, снимаемых с электродов, расположенных в различных точках головы человека. Это дает возможность

судить о взаимосвязи процессов, происходящих в различных областях головного мозга. На рис. 2.36 показаны графики синхронно снятых 16 каналов ЭЭГ. Верхним графикам соответствуют сигналы, полученные с лобной части головы, а нижним — с затылочной. Графическое изображение значений матрицы коэффициентов взаимной корреляции приведено на рис. 2.37. Здесь значениям КВК в диапазоне от -1 до 1 соответствуют оттенки серого цвета. Видно, что значения КВК симметричны относительно диагонали матрицы. Кроме того, можно отметить убывание значений по мере удаления от диагонали, что демонстрирует более слабую связь между сигналами, полученными от далеко расположенных электродов.

Корреляционные функции используются также при автоматическом анализе ЭКГ. В частности, в 3.6 рассмотрено применение ВКФ для решения задачи классификации форм желудочковых комплексов ЭКГ.

2.5. Идентификация сложных биосигналов с использованием оценок условной энтропии

Ряд задач анализа биомедицинских сигналов заключается в распознавании фрагментов конечной длины по степени выраженности детерминированных и случайных компонент, в обнаружении нерегулярных дискретных последовательностей. Это может быть анализ вариабельности сердечного ритма в различных состояниях живого организма, распознавание приступов мерцательной аритмии и т.п. Одним из методов их решения является подход, основанный на теории информации Э. Шеннона [2.27].

Под регулярностью временного ряда, преобразованного посредством квантования и символьного кодирования в последовательность состояний, понимается статистическая устойчивость структуры переходов, наблюдаемая во времени. В качестве индикатора появления новых последовательностей, не согласованных с предшествующими, предлагается использовать условную энтропию [2.16, 2.31, 2.32]. Анализ изменения этого показателя во времени, т.е. при удлинении цепочек символов, может служить эффективной мерой нерегулярности временного ряда.

Вычисление и выбор параметров условной энтропии

Вычисление условной энтропии предполагает выполнение определенной последовательности шагов. Вначале для заданной выборки $x(i)$, $i = 1, 2, \dots, N$, формируется ряд L -мерных последовательностей $\{X_L(i), i = 1, 2, \dots, (N - L + 1)\}$, $X_L(i) = (x(i), x(i + 1), \dots, x(i + L - 1))$. Затем посредством операции квантования и последующего преобразования из числовой в символьную форму это множество задается в виде множества последовательностей состояний $\{Z_L(i), i = 1, 2, \dots, (N - L + 1)\}$. При этом каждый элемент

последовательности $Z_L(i) = (z(i), z(i+1), \dots, z(i+L-1))$ кодируются символами из произвольного алфавита $A = \{a_p\}$, $p = 1, \dots, \xi$. В результате такого преобразования получаем упорядоченные по i последовательности событий, связанных с интервальной оценкой каждого отсчета $x(i)$ в диапазоне $(x_{\min}, x_{\max})$. Они рассматриваются далее как цепочки символов длины L . Изменяя параметр L , можно сформировать последовательности событий разной длины и перейти к вычислению условной энтропии.

Известно [2.27], что условная энтропия определяется в виде

$$E(L | L-1) = - \sum_{L-1} P_{L-1} \sum_{L|L-1} P_{L|L-1} \log P_{L|L-1}, \quad (2.4)$$

где $P_{L|L-1}$ — вероятность появления конкретного символа в последней, L -й ячейке цепочки длины L при условии выделения всех одинаковых цепочек укороченной длины $L-1$; P_{L-1} — вероятность появления конкретной цепочки событий длины $L-1$.

Величина $E(L | L-1)$ определяет дополнительную информацию, необходимую для предсказания L -го события в любой последовательности $\{Z_L(i)\}$, $i = 1, 2, \dots, (N-L+1)$, если известна вся последовательность предыдущих событий. Она может быть вычислена еще и как приращение, получаемое безусловной энтропией $E(L)$ при переходе от последовательности событий длины $(L-1)$ к длине L :

$$E(L | L-1) = E(L) - E(L-1) = - \sum_L P_L \log P_L + \sum_{L-1} P_{L-1} \log P_{L-1}. \quad (2.5)$$

Для выборки большой длины $(N \rightarrow \infty)$ значение $E(L | L-1)$ при увеличении L стремится к нулю для регулярных процессов и к постоянному положительному значению для процессов с чисто случайными свойствами. Однако при анализе конечной выборки отсчетов замена вероятностей в (2.4) и (2.5) значениями частот попадания $x(i)$ в соответствующие интервалы приводит к значительному смещению оценки $E(L | L-1)$. Известно, что для последовательности событий длиной L значение $E(L)$ ограничено сверху значением $\log(N-L+1)$ и поэтому $E(L | L-1)$ с ростом L будет стремиться к нулю независимо от вида анализируемого процесса. Это приводит к тому, что даже для чисто случайного процесса может быть обнаружена ложная регулярность.

В процессе обработки значения $x(i)$ квантуются по ξ интервалам и для $\{Z_{L-1}(i)\}$ пространство состояний оказывается разделенным на ξ^{L-1} областей. Каждая из них — возможное местоположение любой цепочки указанной длины. С увеличением L количество ячеек экспоненциально возрастает, вследствие чего увеличивается и число областей, содержащих лишь одну последовательность событий. При добавлении L -го элемента к таким цепочкам длины $(L-1)$ получаем

нулевой вклад в вычисляемое приращение энтропии и в результате оценки $E(L | L-1)$ редкие события идентифицируются как детерминированные. На самом деле такие цепочки являются «неизвестными» в смысле предсказуемости динамики поведения на последующих интервалах $(L, L+1, L+2, \dots)$. Для исключения ложной регулярности, вносимой в оценку условной энтропии одиночными цепочками, могут применяться следующие виды коррекции:

$$\tilde{E}_1(L | L-1) = E(L | L-1) \cdot \left(1 + \frac{N_{L-1}^{(1)}}{N_L - N_{L-1}^{(1)}}\right), \quad (2.6)$$

$$\tilde{E}_2(L | L-1) = E(L | L-1) + E(1) \cdot \frac{N_{L-1}^{(1)}}{N_L}, \quad (2.7)$$

где N_L — число анализируемых цепочек символов длиной L ; $N_{L-1}^{(1)}$ — число одиночных цепочек символов длиной $(L-1)$.

В выражении (2.6) одиночным цепочкам придается среднее приращение энтропии, оцениваемое по $(N_L - N_{L-1}^{(1)})$ — числу последовательностей событий. В формуле (2.7) таким событиям присваиваются веса $E(1)$, соответствующие энтропии одиночных событий. Благодаря этому «неизвестные» цепочки рассматриваются как случайные, а не как регулярные последовательности. Это позволяет избежать ложного представления о неподтвержденных регулярных изменениях в исследуемых последовательностях. Введение весов, задаваемых величиной $E(1)$, было обосновано в работе [2.32], однако придавались они всем одиночным событиям, обнаруженным на L -м шаге, что приводило к значительному смещению энтропии. В работах [2.16, 2.31] в разряд совершенно случайных событий переводятся лишь те цепочки, которые были представлены лишь одним объектом в ячейках $(L-1)$ -мерного пространства состояний.

На рис. 2.38 представлены зависимости оценок условной энтропии от длины цепочек L для трех видов модельных сигналов: a — гармонического сигнала, b — шума, θ — аддитивной смеси сигнала и шума. Из анализа кривых следует, что для гармонического сигнала их изменения практически совпадают, а при наличии шума — существенно различаются.

Эти различия связаны с коррекцией условной энтропии лишь по числу одиночных цепочек, равному $N_{L-1}^{(1)}$. При этом оценка $\tilde{E}_1(L | L-1)$ незначительно уменьшает крутизну изменения условной энтропии, а оценка $\tilde{E}_2(L | L-1)$ приближается к постоянному уровню, задаваемому величиной $E(1)$, что можно интерпретировать как искусственное «зашумление» анализируемого сигнала. На рис. 2.38, g приведены зависимости величины $N_L^{(1)}$ от длины L для гармонического сигнала (кривая a), шума (b) и смеси гармонического сигнала и шума (θ). Из рисунка видно, что для выборки ограниченной длины N с увеличением L быстро растет число «неизвестных» цепочек

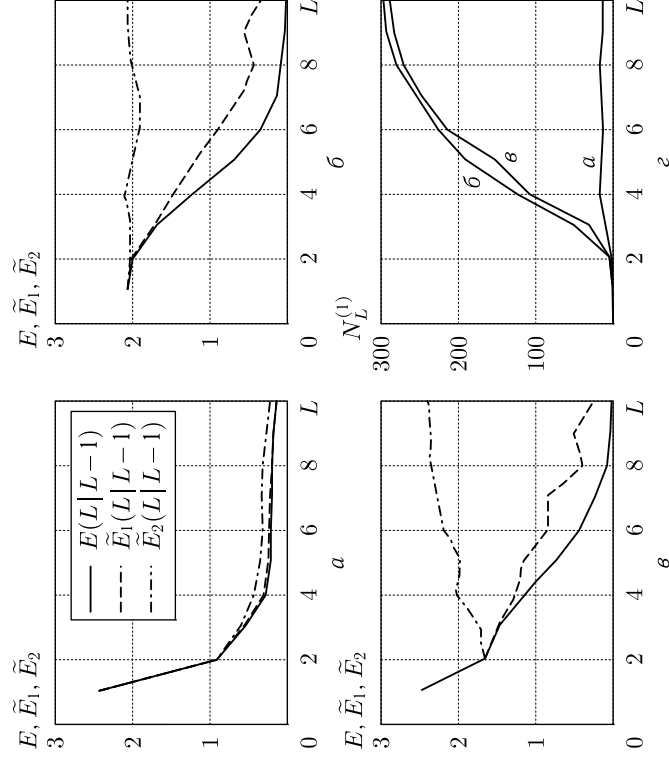


Рис. 2.28. Зависимости оценок условной энтропии $E(L|L-1)$, $\tilde{E}_1(L|L-1)$, $\tilde{E}_2(L|L-1)$ для разных видов модельных сигналов

при наличии в сигнале случайной составляющей, что и требует введения коррекции функции $E(L|L-1)$.

На основе анализа полученных зависимостей выбраны следующие параметры для идентификации сигналов. Во-первых, это могут быть точечные оценки скорректированной кривой $\tilde{E}_1(L|L-1)$ или $\tilde{E}_2(L|L-1)$ при значениях $L = 2, 3, 4$, т.е. в той области, где вклад одиночных цепочек незначителен. Это позволяет в дальнейшем использовать их в качестве дискриминантных признаков при решении задач классификации. Во-вторых, в качестве меры нерегулярности процесса может быть использована нижняя граница кривой $\tilde{E}_2(L|L-1)$ [2.32], которая принимает малое значение для регулярных процессов и большое в случае наличия выраженных случайных изменений. В работах [2.16, 2.31] с этой целью вычислялся относительный минимум ME , задаваемый в виде $ME = E(1) - \min_{L=2, \dots, 10} \{E_2(L|L-1)\}$. Чем выше регулярность процесса, тем больше значение показателя ME . Эксперименты на реальных данных позволили установить, что функция $\tilde{E}_2(L|L-1)$ и показатель ME могут использоваться для оценки степени регулярности изменений, наблюдаемых во временных рядах

конечной длины, а также могут найти применение в задаче распознавания нарушений ритма со сложной динамикой распределения кардиоинтервалов.

Распознавание фрагментов мерцательной аритмии

Оценка ME была использована при анализе нарушений ритма сердца, в частности, в задаче распознавания мерцательной аритмии по последовательности кардиоциклов длиной $N = 300$ [2.16, 2.18]. В процессе исследования было обработано более 150 записей ЭКГ, для которых были построены ритмограммы и вычислен относительный минимум функции $\tilde{E}_2(L|L-1)$. Анализ результатов эксперимента показал, что параметр ME имеет устойчивые интервальные оценки при уровне значимости $\alpha = 0,05$ для разных видов рассматриваемых ритмограмм, причем значение этого параметра при мерцательной аритмии значительно меньше. Так, для нормального ритма с волновой структурой ритмограммы значение ME равно $0,85 \pm 0,09$. При частой желудочковой экстрасистолии, для которой характерны кратковременные регулярные изменения в последовательности RR -интервалов, оно несколько меньше: $0,74 \pm 0,10$. И только при мерцательной аритмии, отличающейся нерегулярным распределением длительностей кардиоциклов, показатель ME значительно меньше: $0,17 \pm 0,11$.

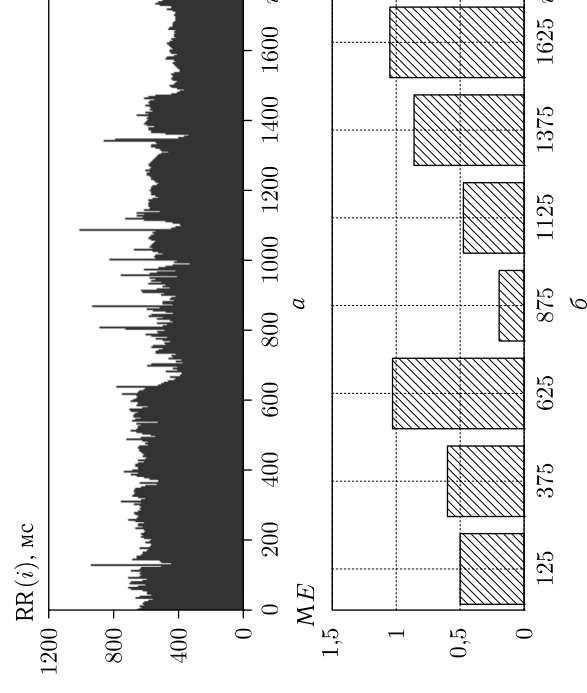


Рис. 2.39. Последовательность RR -интервалов, содержащая фрагмент выраженной мерцательной аритмии (а), и результат анализа параметра ME (б)

На рис. 2.39 в качестве примера показан результат динамического анализа параметра ME для сложной последовательности RR-интервалов, содержащей кратковременный приступ мерцательной аритмии. При попадании текущего окна обработки длиной $N = 250$ на область данных, содержащих хаотические изменения ритмограммы, значение ME резко уменьшается, что является индикатором приступа аритмии. В ходе многочисленных модельных и реальных экспериментов установлено, что показатель ME , который представляет собой интегральный параметр скорректированной условной энтропии, может эффективно использоваться для идентификации нерегулярных последовательностей, в частности для распознавания фрагментов мерцательной аритмии в ходе автоматического слежения за ЭКГ больного.

2.6. Распознавание биосигналов с нелинейными свойствами по параметрам аппроксимированной энтропии

Методы нелинейной динамики все шире используются в практических задачах анализа сложных биомедицинских сигналов (ЭКГ, ЭКГ, ритм сердечных сокращений и др.). Такие сигналы при определенных условиях можно рассматривать как процессы, генерируемые режимом динамического хаоса. Это представление значительно расширяет спектр количественных критериев для диагностики состояний, поскольку основано на анализе характеристик, применяемых для исследования детерминированного хаоса. Так, например, в работах [2.33, 2.34] в качестве такого критерия предложено использовать *аппроксимированную энтропию* (approximate entropy, ArEn), которая количественно определяет степень сложности сигнала.

Аппроксимированная энтропия тесно связана с энтропией Колмогорова (К-энтропией), которая считается важнейшей характеристикой хаотического движения в фазовом пространстве произвольной размерности [2.28]. К-энтропия определяется как средняя скорость потери информации о состоянии динамической системы во времени. При этом для описания состояния системы вводится понятие аттрактора, под которым понимается множество точек или подпространств в фазовом пространстве (образованном компонентами вектора состояний $\mathbf{X}(t)$), к которому приближаются хаотические траектории [2.21]. Предполагается, что d -мерное фазовое пространство разделено на ячейки размером r^d , а состояние системы $\mathbf{X}(t)$ определяется дискретно через интервалы времени τ . Поведение системы на странном аттракторе (аттракторе, имеющем фрактальную размерность) оценивается в виде дополнителной информации, необходимой для предсказания того, в какой ячейке i_{m+1}^* фазового пространства будет находиться система в момент времени t_{m+1} , если известно, что на предыдущих шагах она находилась на траектории $i_1^*, i_2^*, \dots, i_m^*$. К-энтропия равна нулю

для регулярного движения, бесконечна для случайных систем, полужительна и постоянна для систем с детерминированным хаосом [2.28]. Анализируя некоторый сигнал, кажущийся случайным, можно отличить нерегулярность, связанную с динамикой движения на аттракторе, от связанной с воздействием белого шума [2.33].

На практике значение К-энтропии сложно вычислить ввиду того, что анализируется одна реализация сложного процесса, хотя аттрактор, характеризующий свойства системы, вложен в d -мерное фазовое пространство. Однако в соответствии с теоремой Такенса [2.28] некоторые свойства аттрактора можно восстановить по одной составляющей процесса. При этом можно найти нижнюю границу К-энтропии

$$K_2 = \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \ln \frac{C_m(r)}{C_{m+1}(r)}, \quad (2.8)$$

где $C_m(r)$ — обобщенный корреляционный интеграл, m — длина последовательности точек, r — пороговое расстояние.

Корреляционный интеграл вычисляется в виде

$$C_m(r) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_i^N \sum_j^N \Theta(r - |X(i) - X(j)|), \quad i \neq j, \quad (2.9)$$

где $X(i) = (x(i), x(i+1), \dots, x(i+m-1))$ — последовательность, состоящая из m отсчетов сигнала; $\Theta(\cdot)$ — функция Хевисайда ($\Theta(s) = 1$, если $s \geq 0$, и $\Theta(s) = 0$, если $s < 0$); N — объем выборки данных.

Если показатель $K_2 > 0$, то это достаточное условие существования хаоса. Однако требование $m \rightarrow \infty$, $r \rightarrow 0$, $N \rightarrow \infty$ в (2.8), (2.9) делает невозможным применение критерия K_2 в практических задачах анализа сигнала на конечных выборках. Это приводит к необходимости нахождения приближенной оценки энтропии Колмогорова в виде аппроксимированной энтропии.

Рассмотрим алгоритм вычисления функции аппроксимированной энтропии.

Алгоритм вычисления функции аппроксимированной энтропии

Пусть имеется выборка исходных данных $x(1), x(2), \dots, x(N)$, где N — длина выборки. Вначале зададим значения двух параметров: m — длина анализируемых цепочек (последовательностей отсчетов); r — порог, задающий размеры ячеек фазового пространства.

1. Формируем последовательности $X(1), \dots, X(N-m+1)$, определяемые выражением $X(i) = (x(i), x(i+1), \dots, x(i+m-1))$, $i = 1, \dots, (N-m+1)$.

2. Определяем расстояние между $X(i)$ и $X(j)$ как

$$d(X(i), X(j)) = \max_{k=0, \dots, (m-1)} (|x(i+k) - x(j+k)|).$$

3. Вычисляем $C_r^m(i) = N^m(i)/(N - m + 1)$, где $N^m(i)$ — количество значений $d(X(i), X(j))$, удовлетворяющих условию $d(X(i), X(j)) \leq r$, $j = 1, \dots, (N - m + 1)$.
4. Находим натуральный логарифм от каждого $C_r^m(i)$ и усредняем его значение по индексу i : $\theta^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N - m + 1} \ln C_r^m(i)$.
5. Увеличиваем значение m до $m + 1$. Повторяем шаги 1–4 и находим значения $C_r^{m+1}(i)$, $\theta^{m+1}(r)$.
6. Находим оценку К-энтропии

$$\text{ArEn}(m, r) = \lim_{N \rightarrow \infty} (\theta^m(r) - \theta^{m+1}(r)),$$

которая для выборки данных ограниченной длины принимает вид

$$\text{ArEn}(m, r, N) = (\theta^m(r) - \theta^{m+1}(r)).$$

В [2.33] был предложен быстросействующий алгоритм вычисления этой оценки.

Значение аппроксимированной энтропии зависит от параметров m и r . Автор работы [2.34] предложил принять значение $m = 2$ и r равным $(0,1, \dots, 0,25)\sigma_x$, где σ_x — стандартное отклонение исходной выборки данных. Все последующие расчеты аппроксимированной энтропии проведены для $N = 300$, $m = 1, \dots, 6$ и $r = 0,15\sigma_x$.

Сложность анализа аппроксимированной энтропии как функции длины анализируемых цепочек m при конечной длине выборки N связана с тем, что при возрастании m значение $\text{ArEn}(m)$ стремится к нулю независимо от степени регулярности исследуемого процесса. Это происходит из-за увеличения вероятности того, что цепочки длиной m будут представлять собой лишь однажды произошедшие события. По этой причине аппроксимированная энтропия, вычисленная в соответствии с многошаговой процедурой, дает ошибочную оценку степени хаотичности сигнала. Избежать этого можно, вводя коррекцию оценки $\text{ArEn}(m)$ [2.14, 2.17, 2.18]:

$$\text{ArEn}_{\text{cor}}(m) = \text{ArEn}(m) + \text{ArEn}(0)N_m^{(1)}/N_{m+1}, \quad (2.10)$$

где N_{m+1} — число анализируемых цепочек длины $(m + 1)$; $N_m^{(1)}$ — число лишь однажды встретившихся цепочек длиной m ; $\text{ArEn}(0)$ — значение абсолютной энтропии, вычисленное для исходной последовательности отсчетов.

При этом нижнюю границу К-энтропии можно аппроксимировать нижней оценкой $\text{ArEn}_{\text{cor}}(m)$. Отметим, что вид коррекции (2.10) аналогичен корректирующей оценке (2.7), используемой при анализе условной энтропии Шеннона.

Основное отличие аппроксимированной энтропии от условной энтропии $E(L | L - 1)$ заключается в способе формирования элементарных областей фазового пространства. При расчете условной энтропии

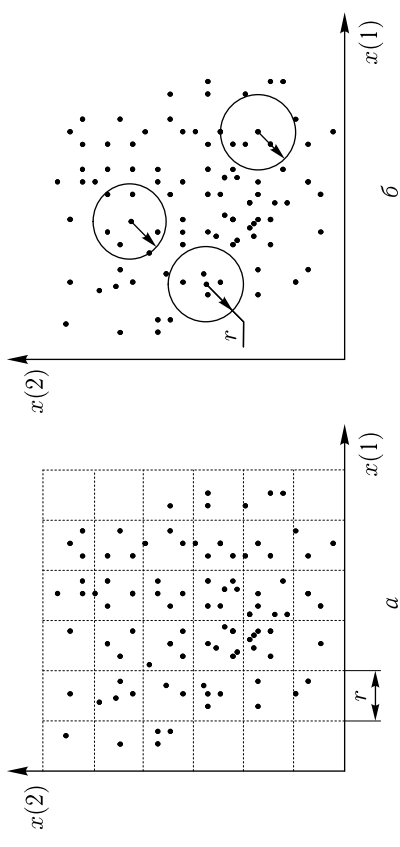


Рис. 2.40. Иллюстрация механизма формирования ячеек фазового пространства при расчете: a — условной энтропии, b — аппроксимированной энтропии

гиперкубы формируются жесткой сеткой, задаваемой числом уровней квантования ξ в пределах минимального и максимального значений $(x_{\min}, x_{\max})$ временного ряда. При расчете $\text{ArEn}(m)$ работает механизм адаптивного формирования элементарных областей путем построения гиперсфер радиусом r с центром, задаваемым каждой точкой, в которой находится объект фазового пространства. Схематическое изображение механизмов формирования элементарных ячеек для двумерного пространства приведено на рис. 2.40.

Анализ свойств и выбор параметров аппроксимированной энтропии

В работах [2.14, 2.17, 2.33] проведены модельные эксперименты, направленные на исследование свойств аппроксимированной энтропии. Среди основных свойств этой характеристики можно выделить следующие.

1. Аппроксимированная энтропия отражает степень сложности сигнала — чем выше его регулярность, тем меньше значение этой величины.
 2. Она позволяет получать надежные оценки хаотических свойств сигналов, используя короткие выборки данных.
- Результаты экспериментов показали, что как для длинных, так и для более коротких последовательностей получаются устойчивые оценки энтропии. В дальнейшем было решено анализировать фрагменты сигналов длиной $N = 300$ отсчетов.
3. Приближенная оценка К-энтропии может использоваться для обнаружения динамических нерегулярностей, связанных как с нелинейными, так и со стохастическими свойствами сигналов.

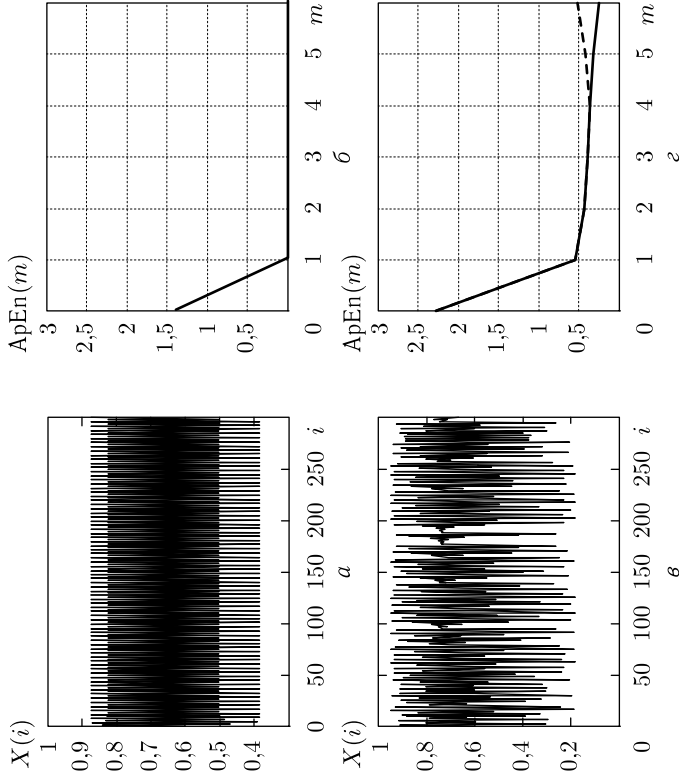


Рис. 2.41. Логистическое отображение при $a = 3,5$ и $a = 3,8$: a и $в$ — исходный сигнал $X(i)$; $б$ и $з$ — зависимости $ArEn(m)$ (сплошная линия) и $ArEn_{cor}(m)$ (штриховая линия) от m

В качестве примера на рис. 2.41 приведены результаты вычисления $ArEn(m)$ и $ArEn_{cor}(m)$ для логистического отображения [2.28], определяемого в виде $X(i+1) = a \cdot X(i)(1 - X(i))$, где a — внешний параметр. Данное отображение представляет собой простейшее нелинейное разностное уравнение, которое при $a = 3,5$ имеет периодический характер (рис. 2.41, a), а при $a = 3,8$ — хаотический (рис. 2.41, $в$). На представленных графиках наблюдаются отличия полученных кривых, а именно: при $a = 3,5$ оценка $ArEn(m)$ с ростом m быстро уменьшается до нуля (рис. 2.41, $б$), что характерно для периодических сигналов, а введение коррекции не вносит изменений в $ArEn_{cor}(m)$ (графики зависимостей совпадают). В то же время при переходе к режиму детерминированного хаоса ($a = 3,8$) зависимость $ArEn(m)$ с ростом m стремится не к нулю, а к некоторому постоянному уровню (рис. 2.41, $з$). При этом на шаге $m = 5$ включается корректирующая поправка, что вызвано возрастанием числа одиночных цепочек, обнаруженных при анализе элементарных гиперсфер.

При анализе стохастического процесса вид кривых $ArEn(m)$ и $ArEn_{cor}(m)$ резко изменяется. На рис. 2.42 представлен результат

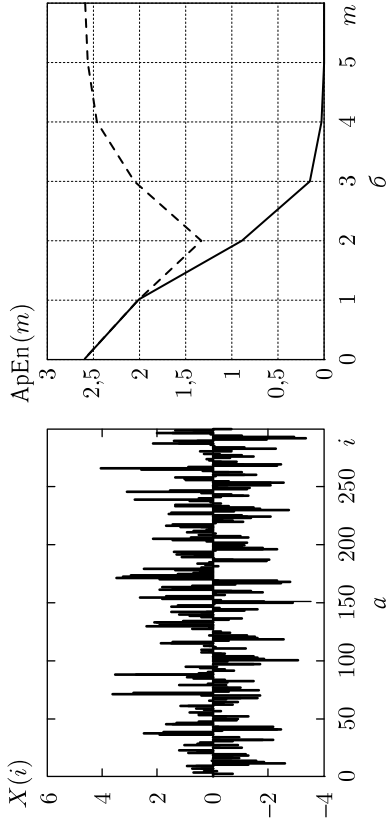


Рис. 2.42. Авторегрессионный процесс 2-го порядка: a — исходный сигнал $X(i)$; $б$ — зависимость $ArEn(m)$ (сплошная линия) и $ArEn_{cor}(m)$ (штриховая линия) от m

анализа процесса авторегрессии второго порядка, который задан следующим уравнением:

$$X(i) - \mu = a_1(X(i-1) - \mu) + a_2(X(i-2) - \mu) + Z(i),$$

где μ — математическое ожидание случайной величины $Z(i)$; $Z(i)$, $i = 1, 2, \dots$, — белый шум с нормальным законом распределения; $a_1 = 1,0$, $a_2 = -0,5$.

Как видно из рис. 2.42, $б$, на графиках наблюдается ярко выраженный минимум $ArEn_{cor}(m)$. Значения $ArEn(m)$ при малых m достаточно велики, а при возрастании m стремятся к нулю, в то время как функция $ArEn_{cor}(m)$ — к значению абсолютной энтропии $ArEn(0)$. Характер зависимостей $ArEn(m)$ и $ArEn_{cor}(m)$ аналогичен закономерностям, выявленным для случая белого шума.

4. Приближенная оценка К-энтропии применима для анализа как детерминированных хаотических процессов, так и стохастических сигналов. Более того, она позволяет оценить степень зашумленности детерминированного сигнала в смешанном процессе — свойство, полезное для анализа биосигналов, поскольку многие из них содержат как регулярную, так и стохастическую составляющие.

5. Нечувствительность аппроксимированной энтропии к воздействию белого шума определяется выбором подходящего значения порога r . Если величина r больше амплитуды шума, то его влияние будет эффективно устранено.

Проиллюстрируем это следующим примером [2.33]. Пусть наблюдается смешанный процесс $X(i) = s(i) + cn(i)$, где $s(i) = 3 \cos(\pi \cdot i/20)$, $i = 1, \dots, N$, — гармонический сигнал; $n(i)$ — шум, равномерно распределенный в пределах значений $(-1, +1)$; c — коэффициент, задающий амплитуду шумовой составляющей сигнала.

Как видно из рис. 2.43, значения $\text{ArEn}(m)$ при всех m , кроме $m = 1$, помехоустойчивы. При $c < r$ ($r = 0,32$) значения $\text{ArEn}(m = 2, \dots, 6)$ постоянны и увеличиваются с ростом m лишь в случае, когда амплитуда шума превысит значение заданного порога r .

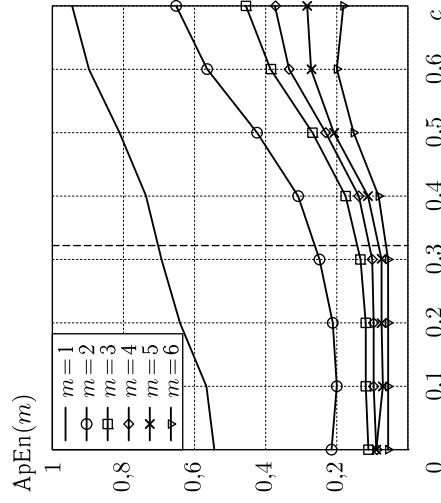


Рис. 2.43. Зависимость $\text{ArEn}(m)$ от амплитуды шума

6. Аппроксимированная энтропия нечувствительна к изменению частоты гармонического сигнала.

Исходя из анализа свойств аппроксимированной энтропии и результатов модельных экспериментов, можно сделать вывод о том, что для оценки степени регулярности изменений, наблюдаемых в дискретной последовательности отсчетов, могут быть использованы следующие параметры:

— значения $\text{ArEn}(m)$ и $\text{ArEn}_{\text{cor}}(m)$ при $m = 1, 2, 3$, где вклад одиночных цепочек незначителен;

— относительный минимум $\text{ArEn}_{\text{cor}}(m)$:

$$\text{ME}_K = \text{ArEn}(0) - \min_{m=1, \dots, 6} \{\text{ArEn}_{\text{cor}}(m)\},$$
 который аппроксимирует нижнюю границу K-энтропии.

Распознавание мерцательной аритмии

Эти параметры были использованы для распознавания мерцательной аритмии по последовательности кардиоинтервалов [2.13, 2.18]. Анализ корреляционной размерности ритмограмм при мерцательной аритмии (МА) показал [2.15], что сердечный ритм имеет хаотическую природу и, следовательно, в этой задаче применимы методы нелинейной динамики. Все исследования проведены на ритмограммах реальных записей электрокардиосигнала. Реализации ЭКГ пациентов были выбраны из базы архивных записей пациентов 4-го кардиологического отделения СПб ГУЗ «Городская больница № 31». Все реализации были

предварительно верифицированы и разделены на три класса: нормальный ритм (НР), включающий выраженную дыхательную аритмию; мерцательная аритмия (МА); частая экстрасистолия (ЧЭ). Затем в каждой из трех групп были отобраны наиболее характерные для данного класса реализации (по 50 реализаций длиной 300 отсч.) и сформированы обучающие выборки. Для этих фрагментов из обучающей выборки были рассчитаны значения $\text{ArEn}(m)$ и $\text{ArEn}_{\text{cor}}(m)$ для $m = 1, \dots, 6$, а также ME_K .

На рис. 2.44 приведены примеры ритмограмм и полученные зависимости $\text{ArEn}(m)$ от длины цепочек m для классов МА (а, б), НР (в, г) и ЧЭ (д, е) при $r_0 = 0,15$, $r = r_0 \sigma_x$. Как видно из рисунка, зависимости имеют различный характер, причем МА характеризуется наименьшим значением ME_K .

Построение решающих функций по оценкам аппроксимированной энтропии выполнено с использованием теории линейного дискриминантного анализа. Задача распознавания МА предполагает обнаружение аритмии на фоне других нарушений ритма, поэтому дискриминантный анализ был проведен для двухклассовой задачи: один класс составили ритмограммы МА (класс ω_1), в другой класс были объединены реализации НР и ЧЭ (класс ω_2). Классификация осуществлялась с использованием различного набора признаков, в качестве которых были выбраны параметры $\text{ArEn}(m)$ и $\text{ArEn}_{\text{cor}}(m)$.

Уравнение линейной разделяющей функции в пространстве заданных признаков определялось в виде $D(\mathbf{X}) = \mathbf{W}^T \cdot \mathbf{X} + w_0 = 0$, где $\mathbf{W} = [w_1 \ w_2 \ \dots \ w_d]^T$ — весовой вектор единичной длины, $\mathbf{X} = [x_1 \ x_2 \ \dots \ x_d]^T$ — вектор наблюдений, d — число признаков, w_0 — пороговая величина.

Значение порога w_0 определялось исходя из критерия оптимальности разбиения на классы, т.е. минимизации суммы ошибок первого рода α (ложная тревога) и второго рода β (пропуск цели): $(\alpha + \beta) \rightarrow \min$, поскольку в этом случае одинаково важно как не пропустить МА, так и не выдать сигнал ложной тревоги.

Для нахождения ошибок α и β были построены гистограммы распределения проекций объектов на единичный вектор $\mathbf{W}$, а также соответствующие этим гистограммам функции плотности распределения, имеющие нормальный закон распределения. Величина порога фильтрации r была задана равной $r = r_0 \cdot \sigma_x$, $r_0 = 0,2$, что увеличило число оцениваемых параметров $\text{ArEn}(m)$ до $m = 6$. Выбор информативных признаков осуществлялся по кривой зависимости критерия Фишера J от числа параметров, включенных в решающее правило. Это позволило ограничить описание объектов, используемое для распознавания классов ω_1 и ω_2 , следующим набором признаков: ME_K ; $\text{ArEn}(m)$ при $m = 1, 2, 3$. Эксперименты по оценке эффективности решающих правил показали, что в этом случае теоретические ошибки классификации минимальны и составляют менее 0,1 %. Эффективность решающего

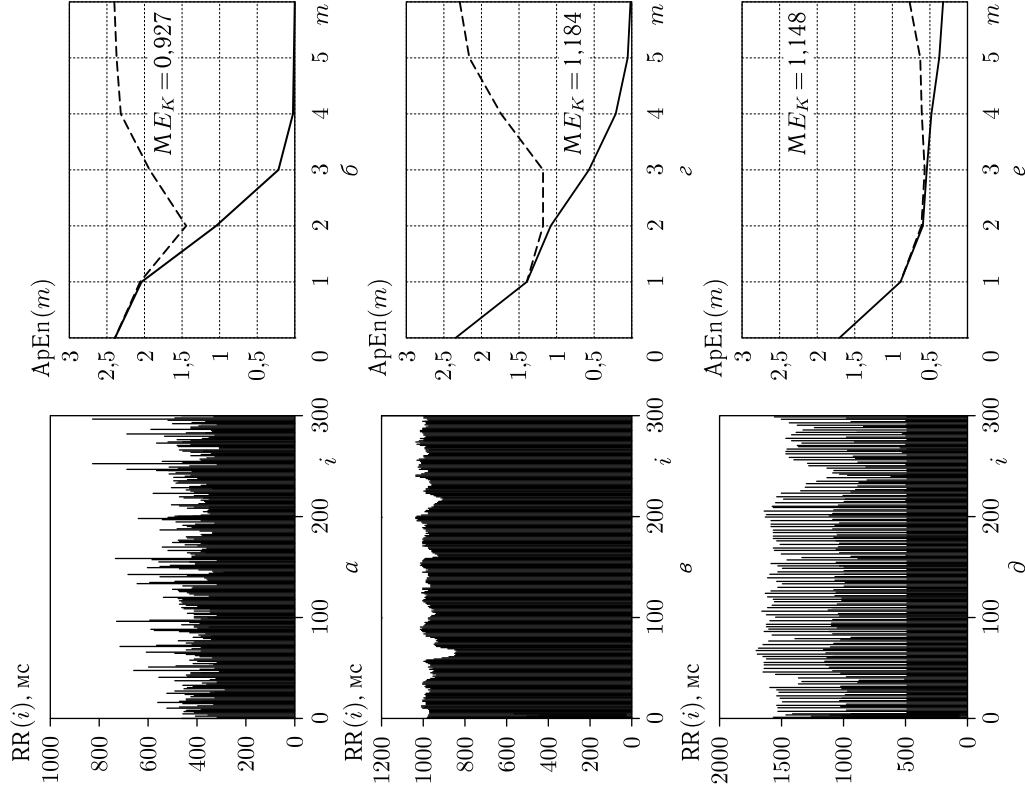


Рис. 2.44. Примеры оценок $ApEn(m)$ (сплошная линия) и $ApEn_{cor}(m)$ (штриховая линия) для реальных сигналов: а, б — мерцательной аритмии; в, г — нормального ритма; д, е — частой экстрасистолии

правила проверена на контрольной выборке, включающей по 100 реализаций ЭКГ для каждого из трех видов ритма (МА — класс ω_1 , НР и ЧЭ — класс ω_2). Как показали результаты расчетов, данный метод обладает высоким уровнем чувствительности ($S_e = 0.98$) и специфичности ($S_p = 0.985$) при распознавании МА. Полученные данные (метод $ApEn$) сравнивались с результатами обнаружения МА, использующими альтернативные методы распознавания [2.10, 2.11, 2.13]:

- проверка гипотезы о гауссовом законе распределения (критерий χ^2);
 - оценка среднеквадратичного отклонения ритмограммы (Dev);
 - предсказание авторегрессионной моделью (AR-модель);
 - оценка минимума уточненной условной энтропии (ME).
- В этих методах используется лишь один информативный признак — индикатор приступов МА. Результаты анализа эффективности этих методов приведены в табл. 2.3.

Таблица 2.3

Метод	Условия обнаружения МА — на фоне	Чувствительность метода, S_e , %	Специфичность метода, S_p , %
Критерий χ^2	НР	74,5	97,4
Dev	ЧЭ	98,8	64,7
AR-модель	НР и ЧЭ	93,3	91,4
ME	НР и ЧЭ	98,5	89,3
ApEn	НР и ЧЭ	98,0	98,5

Из данных таблицы следует, что метод распознавания МА на основе аппроксимированной энтропии наиболее надежен, что позволяет использовать его в практических задачах.

Оценка глубины наркоза по ЭЭГ

Рассмотренные ранее энтропийные характеристики применялись также для анализа свойств ЭЭГ-сигнала в разных стадиях анестезии [2.12, 2.22] и распознавания стадии глубокого наркоза [2.7, 2.30]. В работе [2.23] для оценки глубины наркоза по ЭЭГ предложено использовать показатель аппроксимированной энтропии $ApEn(2)$.

Известно, что изменение уровня функциональной активности головного мозга вызывает характерные изменения, наблюдаемые в ЭЭГ-сигнале [2.6]. Высокий уровень активности мозга сопровождается сложным ЭЭГ-сигналом, имеющим хаотический вид. Снижение степени функциональной активности отражается на ЭЭГ появлением регулярных низкочастотных колебаний достаточно большой амплитуды. В разных стадиях анестезии выраженность регулярных и хаотических изменений ЭЭГ-сигнала различна.

На рис. 2.45 приведены четыре фрагмента ЭЭГ-сигнала, зарегистрированного на одном пациенте в разных стадиях проводимой анестезии: до начала анестезии (а), в начальной стадии (б) и при выходе из анестезии (в), в стадии окончательного пробуждения (г). Частота дискретизации ЭЭГ-сигнала составляет 500 Гц, длительность фрагментов 10 с. Для этих сигналов рассчитаны рассмотренные энтропийные

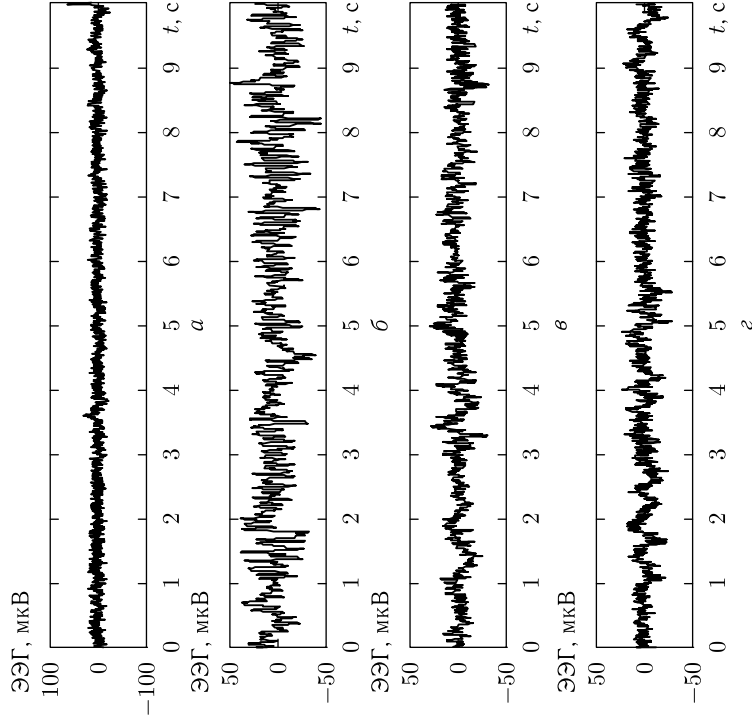


Рис. 2.45. Фрагменты ЭЭГ-сигнала при разных стадиях анестезии: а — до начала анестезии, б — глубокий наркоз, в — выход из анестезии, г — окончательное пробуждение

характеристики, причем длительность анализируемых фрагментов выбрана равной 5 с.

Рис. 2.46 иллюстрирует результат вычисления аппроксимированной энтропии для двух последовательных 5-секундных фрагментов ЭЭГ при значении $r = 0.2\sigma_x$. Как видно из рисунков, в начальной области изменения параметра m кривые имеют явные отличия при разных функциональных состояниях головного мозга.

Выбор параметра $\text{ArEn}(2)$ для оценки уровня анестезии обусловлен тем, что при увеличении m возрастает число одиночных последовательностей и из-за их нулевого вклада в оцениваемые показатели энтропии возникает эффект ложной регулярности сигнала. При выбранном значении $m = 2$ надежность показателя энтропии значительно выше. Как следует из численного анализа $\text{ArEn}(2)$, оцениваемые параметры более чем в 2 раза уменьшают свои значения при переходе от активного состояния к стадии глубокого наркоза. Кроме того, выход из наркоза

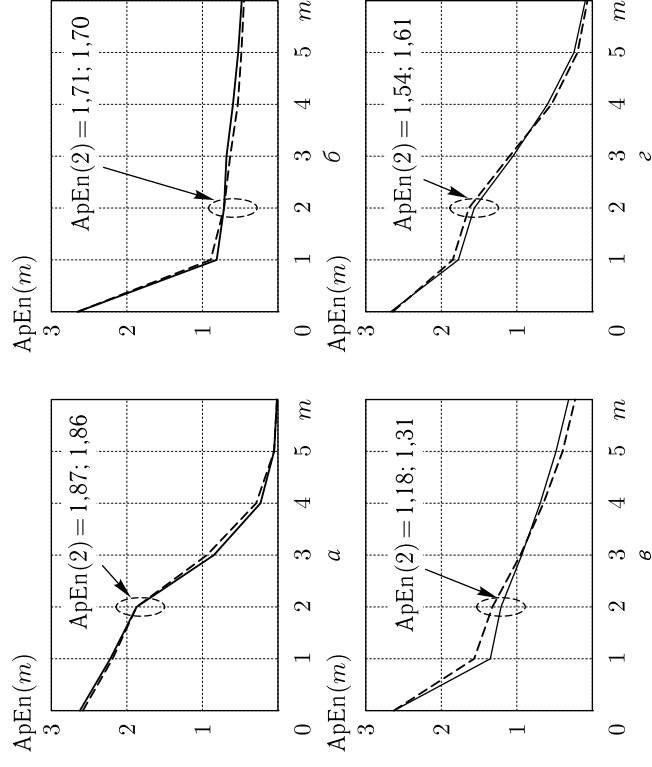


Рис. 2.46. Аппроксимированная энтропия $\text{ArEn}(m)$ для сигналов ЭЭГ, соответствующих разным стадиям анестезии: а — до начала анестезии; б — глубокий наркоз; в — выход из наркоза; г — пробуждение

и переход к окончательному пробуждению сопровождается постепенным возрастанием значений $\text{ArEn}(2)$.

Для приведения показателя энтропии к принятой в анестезиологии шкале (0...100) использована нормировка по значению абсолютной энтропии $\text{ArEn}(0)$. В этом случае следует ожидать значений энтропии, близких к 100, для сигнала ЭЭГ, зарегистрированного до начала анестезии и в стадии полного пробуждения. При глубоком наркозе, очевидно, показатель $\text{ArEn}(2)$ будет иметь довольно малые значения.

На рис. 2.47 приведены результаты вычисления нормированного показателя $\text{ArEn}(2)/\text{ArEn}(0)$ для выборки ЭЭГ-сигналов (64 реализации длиной 5 с каждая), зарегистрированных у восьми пациентов в клинических условиях во время проведения операции. Они представлены в графическом виде как разброс точек на вертикальных линиях с координатами: 1 — до начала анестезии, 2 — глубокий наркоз, 3 — выход из наркоза, 4 — окончательное пробуждение. Каждая точка — это результат вычисления энтропийного показателя для одной 5-секундной реализации ЭЭГ. Как видно из рисунка, на линиях 1 и 2 образованы пересекающиеся множества объектов. Очевидно, что при пороговом значении аппроксимированной энтропии на уровне 0,45 можно для исследуемой выборки данных решить задачу распознавания стадии

Аппроксимированная энтропия
 $\text{ArEn}(2)/\text{ArEn}(0)$

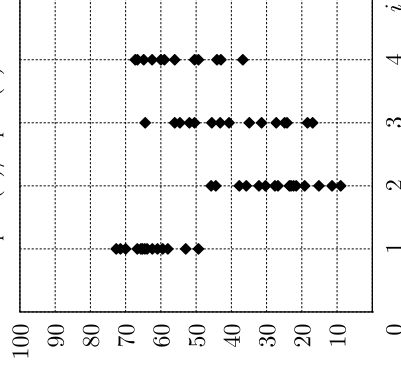


Рис. 2.47. Отображение энтропийных параметров на направления i , соответствующие четырем стадиям анестезии

глубокого наркоза. Стадии 3 и 4 образуют группы точек, занимающие промежуточное положение между состояниями 1 и 2, что соответствует существующему представлению о процессе выхода из наркоза.

В качестве показателя эффективности применения $\text{ArEn}(2)/\text{ArEn}(0)$ для распознавания стадий наркоза выбран критерий Фишера J . Он вычислялся по множествам значений параметра энтропии для двух следующих друг за другом состояний ($1 \rightarrow 2$; $2 \rightarrow 3$; $3 \rightarrow 4$). Всего в ходе эксперимента проанализированы электроэнцефалограммы 31 пациента. В результате обработки 124 реализаций ЭЭГ длительностью 10 с каждая получены следующие значения: $J_{1,2} = 3,07$; $J_{2,3} = 0,78$; $J_{3,4} = 0,47$. Оценка энтропийных параметров проводилась по коротким 5-секундным фрагментам ЭЭГ-сигнала.

Как следует из анализа значений критерия J , а также данных, приведенных на рис. 2.47, метод оценки параметра $\text{ArEn}(2)$ позволяет решить задачу распознавания двух состояний: бодрствование и глубокий наркоз. Это дает основу для дальнейшего исследования энтропийного подхода при распознавании промежуточных стадий анестезии.

2.7. Анализ фазовых портретов биомедицинских сигналов

Алгоритмы нелинейной динамики находят применение как при моделировании работы сердечно-сосудистой системы, головного мозга, так и для диагностики состояний различных физиологических систем человека [2.2, 2.33].

Обычно при оценке вариабельности и степени сложности сердечного ритма применяется параметрическая статистика и спектральный

анализ. В то же время установлено, что основная доля в спектре мощности ВСР приходится на хаотические компоненты, имеющие фрактальную природу. Поэтому все чаще исследуются характеристики фрактальности сердечного ритма. Применяя методы нелинейной динамики к анализу длительных записей ЭКГ, можно получить более полную картину сложных изменений в регуляции работы сердца.

Современные представления об особенностях ЭЭГ-сигнала, характерных для разных уровней активности головного мозга, также указывают на важность исследования его нелинейных компонент с целью распознавания различных состояний (бодрствование, сон, интеллектуальная нагрузка, эмоциональное напряжение) и некоторых патологий (эпилепсия, травмы головного мозга и др.).

Считается, что в норме ЭЭГ пациента имеет высокую степень хаотичности, а при некоторых патологиях более выражены его квазипериодические компоненты. В то же время при ряде заболеваний сердечно-сосудистой системы, появлении аритмий наблюдается обратная картина — возрастает степень нерегулярности сердечного ритма [2.2]. В целом результаты исследований показывают, что изменение степени хаотичности анализируемых сигналов может стать важным признаком развития патологии [2.33].

Меру хаотичности биомедицинских сигналов можно оценить, реконструируя фазовое пространство и анализируя особенности полученного фазового портрета.

Введем некоторые определения из теории нелинейной динамики [2.3, 2.9, 2.28].

Основные понятия и определения

Если система описывается конечным набором n параметров, то ее можно рассматривать в абстрактном пространстве, оси которого представлены переменными $x_1, \dots, x_n$. Такое n -мерное пространство получило название *фазовое пространство*. Каждому состоянию динамической системы соответствует точка в этом пространстве и каждой точке из этого пространства соответствует единственное состояние системы. Изменение состояния системы можно интерпретировать как движение точки в фазовом пространстве. Траектория такой точки, то есть последовательное изменение ее положения в фазовом пространстве, называется *фазовой траекторией*. Множество траекторий в фазовом пространстве дает общую картину поведения исследуемой системы, т.е. ее *фазовый портрет*. В общем случае фазовые траектории с течением времени стягиваются к некоторой области фазового пространства, называемой *аттрактором*. Аттракторы бывают статические (фиксированная точка), периодические (предельный цикл), квазипериодические (тор) и хаотические. Последний тип аттрактора называют *странным аттрактором* — он характеризуется дробной фрактальной размерностью.

В классическом представлении *хаос* означает состояние беспорядка и нерегулярности. В современной трактовке хаос определяют как «неожиданное поведение детерминированных систем», т. е. кажущееся случайным поведение, которое полностью описывается детерминированными правилами или моделями.

В хаотических системах фазовые траектории с течением времени стягиваются к странному аттрактору. Но на самом аттракторе движение неустойчиво: любые две траектории системы расходятся экспоненциально, оставаясь на странном аттракторе. При этом поведение их очень похоже на случайное, хотя динамика детерминирована и воспроизводима при условии точного повторения начальных условий. Классическими примерами таких систем являются отображение Хенона и аттрактор Лоренца [2.28].

Динамические системы, имеющие несколько аттракторов, при незначительных внешних воздействиях могут «перескакивать» с одного аттрактора на другой. Такой скачок называется *бифуркацией*, а момент времени, в который он произошел, — точкой бифуркации. При бифуркациях происходит значительная качественная перестройка динамики системы. Анализ бифуркаций в нелинейной динамике отводит важное место, поскольку несколько последовательных бифуркаций могут приводить к потере устойчивости и возникновению в системе странного аттрактора.

Наиболее широко распространенная величина, используемая при описании хаотической динамики, — *размерность аттрактора*. В общем случае она равна минимальной размерности пространства, содержащего все фазовые траектории системы. Другими словами, это число степеней свободы динамики системы. Существует большое количество численных методов нахождения размерности аттрактора. Самый используемый алгоритм предложен Грассбергером и Прокачича [2.3, 2.28, 2.33]. Он позволяет вычислить так называемую *корреляционную размерность* D_c аттрактора, характеризующую степень сложности динамики исследуемой системы. Важность этой характеристики состоит еще в том, что она может быть определена непосредственно по экспериментальным данным. При этом путем обработки только одного временного ряда можно оценить минимальное число переменных, которое должна включать математическая модель исследуемой динамической системы. Кроме того, в результате специальной обработки зарегистрированного сигнала становится возможной оценка размерности фазового пространства и построение фазового портрета, по которому проводится анализ состояния физиологической системы.

Реконструкция фазового пространства

Рассмотрим метод реконструкции фазового пространства для системы, состояние которой оценивается только по одной реализации. Аттрактор для такой временной последовательности может быть построен с помощью метода задержек, основанного на теореме

Такенса [2.3]. Согласно этой теореме, можно восстановить некоторые свойства аттрактора в фазовом пространстве по временной последовательности только одной составляющей, т. е. наблюдая за изменением лишь одного параметра этой системы.

Пусть состояние некоторой динамической системы оценивается только по одной из фазовых координат, которая представлена временным рядом

$$x(1), x(2), \dots, x(N), \quad (2.11)$$

где N — количество измерений.

Восстановление аттрактора системы сводится к построению псевдоаттрактора, в котором компонентами вектора $\mathbf{z}(t)$, соответствующего точке $x(t)$, являются элементы анализируемой последовательности, но полученные с некоторой временной задержкой:

$$\mathbf{z}(t) = (x(t), x(t + \tau), \dots, x(t + (m - 1)\tau)), \quad (2.12)$$

где τ — величина задержки (дискретная величина), m — размерность пространства вложения, $t = 1, 2, \dots, (N - (m - 1)\tau)$ — дискретное время.

В векторном пространстве этих координат $(x(t), x(t + \tau), \dots, x(t + (m - 1)\tau))$, изменяя последовательно параметр t , можно построить аттрактор таким образом, что он будет сохранять важнейшие свойства и динамику исходного аттрактора. При этом момент времени t в (2.12) можно соотносить как с первым, так и с последним элементом последовательности. Размерность m определяется из условия

$$m \geq 2[d] + 1, \quad (2.13)$$

где d — фрактальная размерность аттрактора, $[d]$ — обозначает целую часть фрактальной размерности [2.3].

Основной задачей анализа временного ряда (2.11) для последующей псевдофазовой реконструкции является определение величин временной задержки τ и размерности m .

В работах [2.3, 2.9] подробно рассмотрено несколько способов выбора временной задержки: метод автокорреляционной функции, метод взаимной информации, метод среднего отклонения. Поскольку компоненты вектора $\mathbf{z}(t)$, характеризующего систему, должны быть независимы, величина τ может быть получена как точка первого пересечения с нулевой линией автокорреляционной функции временного ряда. Данный метод достаточно прост и часто применяется в практических задачах.

Для оценки величины d используют алгоритм Грассбергера–Прокачича, позволяющий вычислить корреляционную размерность D_c . Процедура расчета этого показателя включает вычисление корреляционного интеграла $C(r)$ по формуле (2.9), которую в общем случае применяют для векторов $\mathbf{X}(i)$ и $\mathbf{X}(j)$, описывающих поведение системы в фазовом пространстве. Величина $C(r)$ равна нормированному числу

пар точек на аттракторе, расстояние между которыми не превышает величины r . На специально построенном графике, показывающем логарифмическую зависимость $\ln C(r) = f \ln(r)$, выделяют линейный участок (область скейлинга) и методом наименьших квадратов строят линейную зависимость. Корреляционную размерность, определяемую соотношением $D_c = \lim_{r \rightarrow 0} \frac{\ln C(r)}{\ln r}$, вычисляют как тангенс угла наклона этой прямой. Полученная величина и будет оценкой размерности d . Наряду с натуральным логарифмом при расчете можно использовать масштаб двойного или десятичного логарифма.

При анализе временного ряда (2.11) данная процедура применяется для множества последовательностей

$$X(i) = (x(i), x(i - \tau), \dots, x(i - (m - 1)\tau)), i = ((m - 1)\tau + 1), \dots, N,$$

длины m , которые можно сформировать из исходного временного ряда. В результате для заданного m и полученной выборки данных может быть найдена корреляционная размерность $D_c(m)$.

Рис. 2.48 иллюстрирует построение линейных зависимостей корреляционного интеграла для расчета $D_c(m)$ на примере отображения Хенона. Вычисленные значения корреляционной размерности для разных m сведены в табл. 2.4.

Для определения размерности вложения необходимо последовательно вычислять корреляционную размерность для каждого $m = 1, 2, \dots$

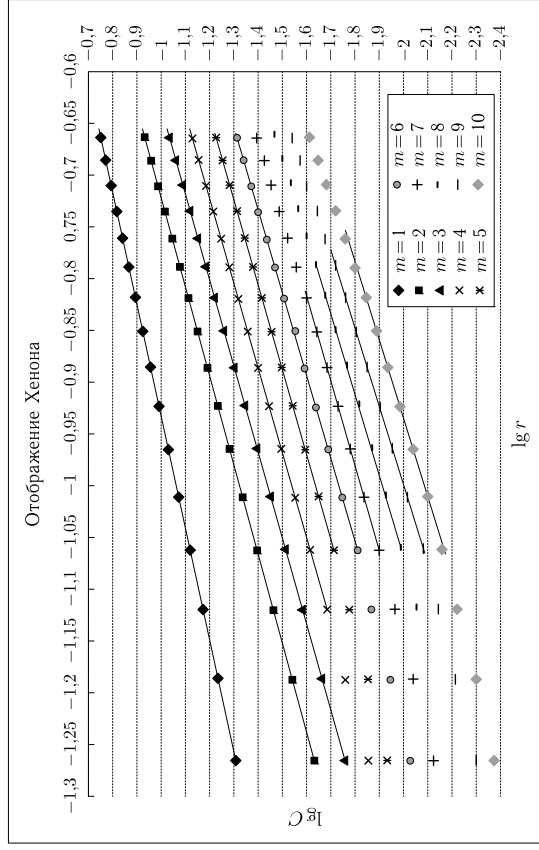


Рис. 2.48. Вид линейных зависимостей корреляционного интеграла для отображения Хенона

Таблица 2.4

m	1	2	3	4	5	6	7	8	9	10
D_c	0,936	1,153	1,186	1,203	1,233	1,23	1,234	1,231	1,232	1,239

и затем следует построить графическую зависимость $D_c = f(m)$. В начальной области графика при добавлении новых компонент псевдоаттрактора и соответственно увеличении m корреляционная размерность растет. Это означает, что m компонент недостаточно для сохранения основных свойств аттрактора. Начиная с некоторой размерности m кривая перестает изменяться, т.е. наблюдается насыщение величины $D_c(m)$ на некотором уровне d . Эта величина d определяет корреляционную (фрактальную) размерность аттрактора, т.е. минимальную размерность фазового пространства. Величина m , с которой начинается насыщение $D_c(m)$, задает минимальную размерность пространства вложения. Для сильно зашумленного сигнала насыщение $D_c(m)$ не наблюдается.

Далее в соответствии с (2.13) определяется размерность модельной динамической системы, которая необходима для реконструкции фазового пространства и построения фазового портрета динамической системы по наблюдаемому временному ряду. Эта размерность соответствует числу независимых переменных анализируемой дискретной последовательности, с помощью которых можно описать исходную систему.

Важно отметить следующие особенности практического применения рассмотренного алгоритма. Координаты временного ряда должны быть выбраны с шагом, равным величине τ . Причем величина шага между элементами ряда (2.11) и величина задержки τ могут и не совпадать. Кроме того, в практических задачах значение m , определяемое формулой (2.13), часто оказывается завышенным и можно ограничиться пространством меньшей размерности, а именно воспользоваться условием $m = [d + 1]$ [2.3] и даже $m < d$ [2.2].

На рис. 2.49 в качестве примера приведены графики зависимости корреляционной размерности D_c от размерности вложения m для модельных сигналов, а именно для гармонического сигнала (рис. 2.49, а), отображения Хенона (рис. 2.49, б) и белого шума (рис. 2.49, в). Все кривые построены по наблюдаемым последовательностям.

Модельные сигналы заданы в виде:

$$x(i) = 88 \cdot \sin(\pi \cdot i / 100) + 600 \text{ — гармонический сигнал;}$$

$x(i) = 220 \cdot \text{randn}(N, 1) + 600$ — белый шум с математическим ожиданием — 600 и дисперсией — 220, где randn — функция, генерирующая случайную величину с нормальным распределением; N — количество отсчетов дискретного сигнала.

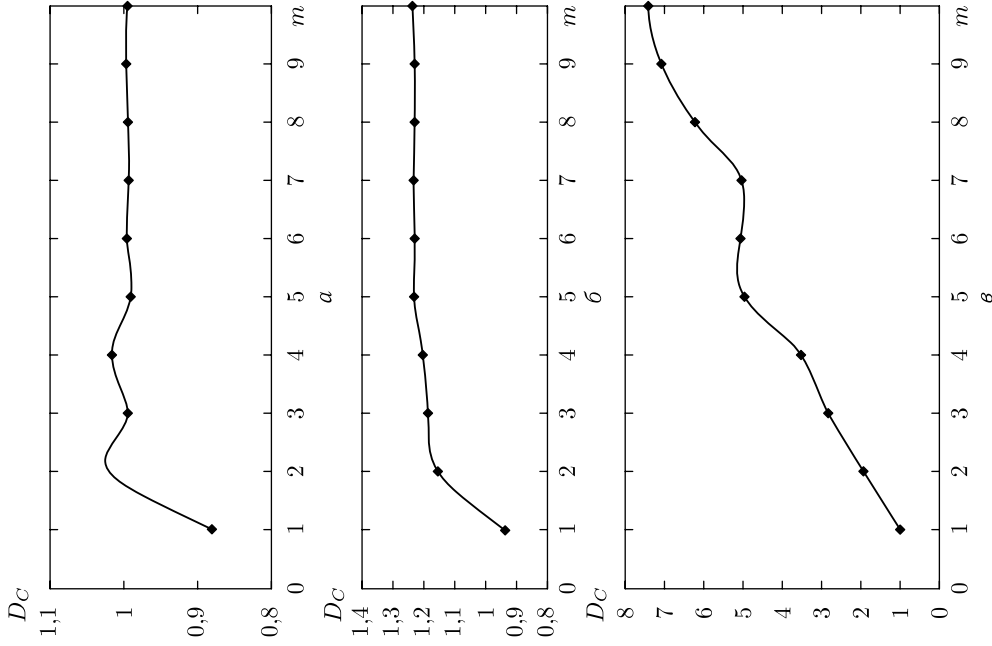


Рис. 2.49. Зависимость корреляционной размерности D_C от размерности вложения m для гармонического сигнала (a), отображения Хенона (b), белого шума (σ)

Отображение Хенона — двумерный аналог логистического отображения, имеет следующий вид:

$$x(i+1) = 1 - a \cdot x(i)^2 + b \cdot x(i-1),$$

где a и b — внешние параметры, равные $a = 1,4$ и $b = 0,3$.

Как видно из рисунка, кривая зависимости корреляционной размерности D_C от размерности вложения для гармонического сигнала с увеличением m возрастает и выходит на плато на уровне 1. Для отобра-

ражения Хенона кривая D_C переходит в насыщение на уровне дробной величины 1,23, что свидетельствует о хаотическом характере генерируемого процесса. Для случайного сигнала (белого шума) увеличение размерности вложения m на единицу приводит к примерно такому же увеличению показателя D_C , поэтому теоретически для его описания требуется бесконечное число переменных.

Рассчитанные значения корреляционной размерности для заданных модельных сигналов, а также значения, найденные в литературных источниках, практически совпадают (табл. 2.5). Небольшие различия в полученных значениях можно объяснить малой длиной анализируемой выборки, которая составляла 1000 отсчетов.

Таблица 2.5

Модельный сигнал	Значение корреляционной размерности d	
	Результаты экспериментов	Данные из литературных источников [2.28]
Гармонический сигнал	1,0	$\sim 1,0$
Отображение Хенона	1,23	1,21
Белый шум	∞	∞

На рис. 2.50 приведен вид временной последовательности $X(i)$, $i = 1, \dots, 300$, для отображения Хенона (рис. 2.50, a), а также его реконструированный фазовый портрет в пространстве размерности $m = 2$ (рис. 2.50, b).

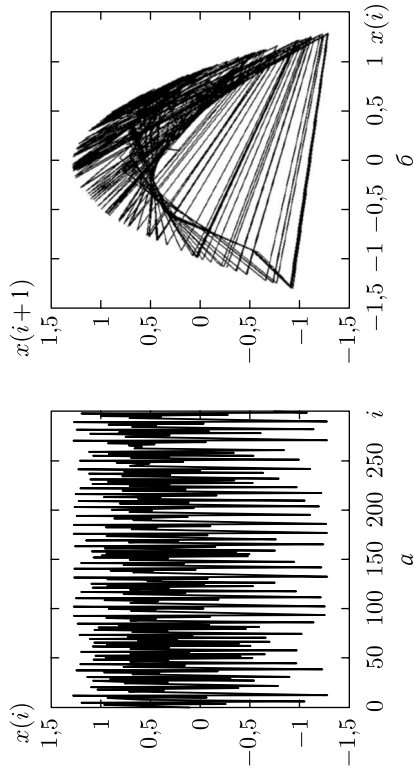


Рис. 2.50. Представление отображения Хенона в виде a — последовательности отсчетов, b — фазовой траектории на плоскости

По уравнению движения известно, что аттрактор Хенона является двумерным [2.9, 2.28], но здесь восстановленное фазовое пространство размерности $m = 2$ и фазовый портрет получены расчетным путем.

Распознавание аритмий по ритмограмме

Рассмотренный метод реконструкции фазового пространства применялся для анализа биомедицинских сигналов (ЭКГ, ЭЭГ). Проведен анализ корреляционной размерности реальных данных — сигнала сердечного ритма при некоторых видах аритмий [2.15]. В первую очередь, это мерцательная аритмия (МА, atrial fibrillation, фибрилляция предсердий), надежное обнаружение которой — одна из главных задач кардиомониторного наблюдения. При этом нарушении выражена нерегулярность в появлении желудочковых сокращений, наблюдается хаотический разброс длительностей кардициклов (рис. 2.51, а). Анализ ритмограмм показал, что зависимость корреляционной размерности от размерности вложения, как и в случае модельных экспериментов (отображение Хенона), достигает насыщения на уровне дробной величины $d = 5,5$ (рис. 2.51, б). Это свидетельствует о хаотической природе сердечного ритма при МА.

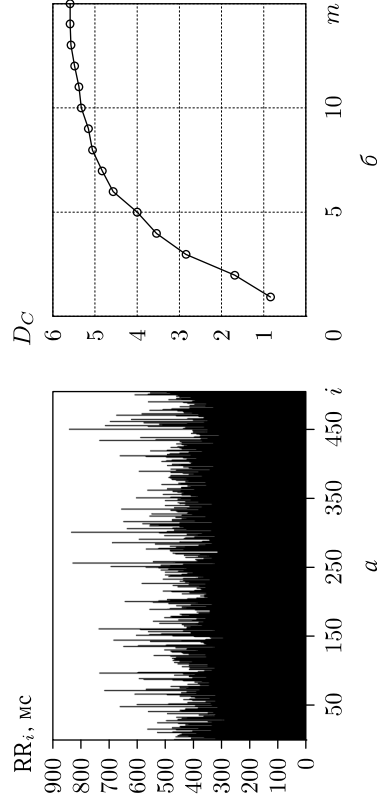


Рис. 2.51. Оценка корреляционной размерности последовательности кардиоинтервалов при мерцательной аритмии: а — вид ритмограммы; б — зависимость D_C от размерности вложения m

Аналогичные зависимости были получены для нормального синусового ритма (НР) и частой экстрасистолии (ЧЭ). Оказалось, что для НР характерно значение $d = 3,5$, а в случае ЧЭ $d = 1$. Это два наиболее распространенных вида ритма сердца, на фоне которых трудно распознать МА, используя традиционные статистические и спектральные методы. Кривые были построены согласно предположению о независимости смежных отсчетов ритмограммы, что, конечно, позволяет получить лишь приближенную оценку фрактальной размерности d .

В практических задачах выбор размерности m на основе расчетных значений d зависит от типа решаемых задач. Для задач моделирования, прогнозирования временных рядов, исследования свойств аттрактора сердечного ритма размерность пространства вложения необходимо выбирать в соответствии с приведенными выше рекомендациями, ориентируясь на полученные значения d и условие (2.13). Классификация сигналов возможна на основе более простых описаний, например представления псевдофазового портрета в пространстве меньшей размерности. При решении таких задач важно выявить отличительные особенности фазового портрета и сформировать набор признаков, обеспечивающий надежное распознавание сигналов заданных классов.

Учитывая большие различия в полученных оценках минимальной размерности фазового пространства, для решения задачи обнаружения МА на фоне НР и ЧЭ размерность псевдофазового пространства

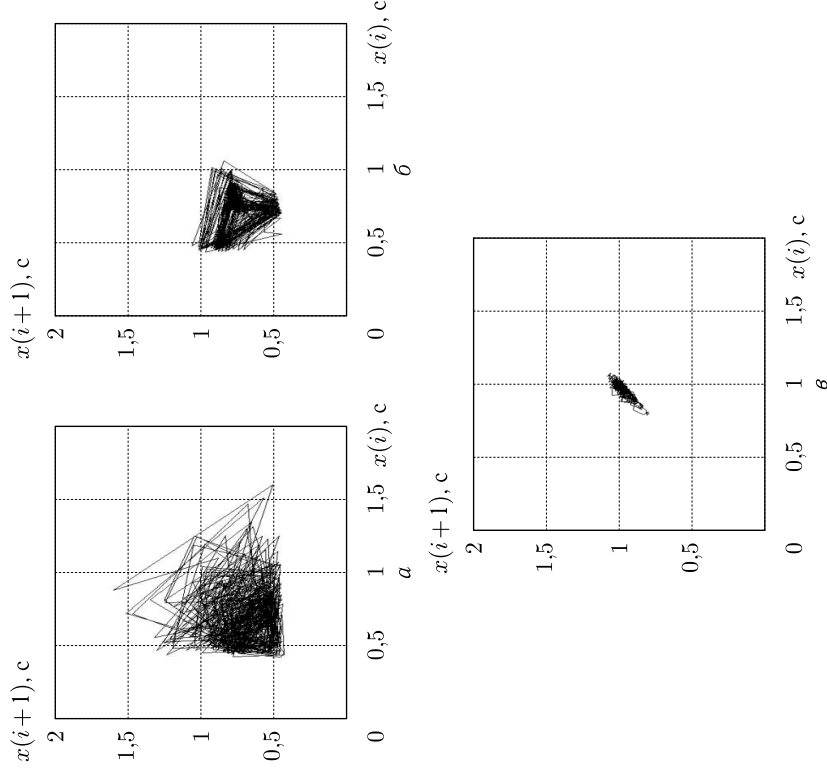


Рис. 2.52. Двумерное отображение псевдофазового портрета для мерцательной аритмии (а), частой экстрасистолии (б) и нормального синусового ритма (в)

выбрана равной 2. На рис. 2.52 представлены проекции псевдофазового портрета на плоскость для динамической модели сердечного ритма при МА, НР, ЧЭ.

Анализ особенностей полученных изображений позволил сформировать следующий набор морфологических признаков:

- длина фазовой траектории;
- площадь портрета, ограниченная контуром;
- длина контура;
- угловой показатель.

Первые три признака довольно просто вычисляются и традиционно используются при обработке изображений [2.35]. Четвертый признак определяет относительно число переходов от короткого интервала к длинному и наоборот на всей выборке данных. Именно такие переходы наиболее характерны для сигналов, имеющих слабую регулярность, в частности для последовательности кардиоинтервалов при МА. Фактически вычисляется число линий фазовой траектории, расположенных под углом $100\text{--}170^\circ$ к горизонтальной оси, задаваемой координатой $x(i)$.

Методом линейного дискриминантного анализа, основанном на использовании критерия Фишера, получено решающее правило, которое реализовано в алгоритме автоматического обнаружения МА. Результаты экспериментов, проведенных на выборке реальных записей ЭКГ (описание выборки реальных данных приведено в разд. 2.6) показали, что ошибка классификации составляет 2 % при длине последовательности, равной всего 50 кардиоинтервалам.

Важно отметить, что данный метод позволяет достаточно надежно обнаруживать МА на фоне НР и ЧЭ по коротким фрагментам ритмограммы, что отличает его от ранее рассмотренного метода, использующего оценки аппроксимированной энтропии.

АНАЛИЗ ЭЛЕКТРОКАРДИОГРАФИЧЕСКИХ СИГНАЛОВ В МОНИТОРНЫХ И ДИАГНОСТИЧЕСКИХ СИСТЕМАХ

3.1. Основные задачи и этапы непрерывного анализа ЭКГ

При оценке состояния сердца и сердечно-сосудистой системы ключевую роль играет анализ ЭКГ, представляющей собой запись наблюдаемых на поверхности тела проекций объемных электрических процессов, происходящих в сердце. Электрокардиограмма несет информацию как о текущем состоянии сердечно-сосудистой системы, так и о патологических изменениях в самом сердце [3.2, 3.4, 3.10].

К настоящему времени проблема автоматической обработки и анализа электрокардосигнала сложилась в отдельное направление, разветвленное на множество более частных задач, связанных с различными применениями и аспектами исследования сердечно-сосудистой системы. Одна из таких частных задач — длительный непрерывный контроль ЭКС, представляющий собой одну из важнейших функций аппаратуры для кардиологического наблюдения, к которой относятся приборы и системы для прикроватного и амбулаторного мониторинга наблюдения ЭКГ, телеметрического наблюдения за состоянием организма, автономного контроля сердечной деятельности, функциональных исследований сердечно-сосудистой системы и решения ряда других задач.

Общим свойством систем кардиологического наблюдения является необходимость автоматической обработки продолжительных непрерывных реализаций электрофизиологических сигналов; к наиболее важным среди них относится электрокардосигнал. При этом обработка должна выполняться либо в режиме реального времени, либо в ускоренном режиме (анализ накопленных данных в холтеровских системах). В системах кардиологического наблюдения решаются следующие основные задачи:

- текущий контроль частоты сердечных сокращений (ЧСС), служащий одним из основных гемодинамических показателей сердечно-сосудистой системы;

— распознавание нарушений сердечного ритма (аритмий), которые возникают вследствие нарушений в автоматизме и проводящей системе сердца;

— анализ ишемических изменений ЭКГ (в частности, изменений ST-сегмента кардиоцикла), служащий индикатором нарушений коронарного кровообращения;

— анализ вариабельности сердечного ритма, позволяющий получать оценки регуляторных функций вегетативной нервной системы.

Автоматический анализ ЭКГ в системах кардиологического наблюдения характеризуется рядом специфических особенностей, к которым, в первую очередь, относятся следующие:

— сложность и разнообразие сигнала, что проявляется в наличии его индивидуальных особенностей у каждого пациента, связанных как с типом конституции человека, так и с присутствием тех или иных патологий в проводящей системе сердца;

— изменчивость сигнала во времени, которая может быть обусловлена как изменениями положения тела, так и физиологическими феноменами — изменением общего состояния пациента, развитием патологий в сердечно-сосудистой системе и появлением нарушений ритма сердца;

— наличие различных видов помех, неизбежно возникающих в процессе длительного наблюдения сигнала;

— необходимость получения результатов анализа в реальном масштабе времени (или в ускоренном режиме, как в случае анализа холтеровских записей), что накладывает ограничения на вычислительную сложность используемых алгоритмов с точки зрения загрузки процессора и используемого объема оперативной памяти;

— использование ограниченного числа отведений ЭКГ (в некоторых применениях — только одного отведения), что затрудняет получение точных оценок диагностических показателей, ввиду недостаточно полной информации, содержащейся во входном сигнале.

Как правило, процесс автоматического анализа ЭКС в системах кардиологического наблюдения включает в себя следующие основные этапы.

1. Предварительная обработка сигнала:

— устранение или ослабление той части помех, которая поддается подавлению;

— оценка уровня помех в сигнале с целью исключения фрагментов, дальнейший анализ которых невозможен вследствие их высокой зашумленности;

— выделение информативных составляющих сигнала с целью обеспечения наилучших условий для последующих этапов обработки.

2. Обнаружение QRS-комплекса ЭКГ и измерение RR-интервала:

— обнаружение QRS-комплекса на фоне других элементов кардиоцикла ЭКГ;

— определение границ комплексов и условной опорной точки, служащей для измерения RR-интервала.

3. Распознавание нарушений ритма, непосредственно угрожающих жизни пациента.

4. Анализ морфологии QRS-комплекса:

— кластеризация QRS-комплексов по сходству их форм;

— идентификация сформированных кластеров с точки зрения их принадлежности к фоновому ритму или к какой-либо из категорий патологий.

5. Распознавание нарушений сердечного ритма:

— распознавание аритмических событий;

— формирование заключений о характере ритма сердца.

6. Анализ ишемических изменений ЭКГ:

— измерение смещения ST-сегмента относительно изоэлектрической линии;

— анализ типа формы ST-сегмента.

7. Анализ вариабельности сердечного ритма:

— формирование последовательности NN-интервалов (т. е. интервалов между смежными QRS-комплексами фонового ритма);

— расчет оценок временных и частотных показателей ВСР.

3.2. Характеристики электрокардиосигнала и помех

Электрокардиограмма представляет собой запись электрических потенциалов, снимаемых с поверхности тела человека и отражающих электрические процессы, происходящие в сердце на протяжении всех фаз сердечного цикла. На рис. 3.1 приведен вид типичного цикла ЭКГ здорового человека в норме. На электрокардиограмме различают

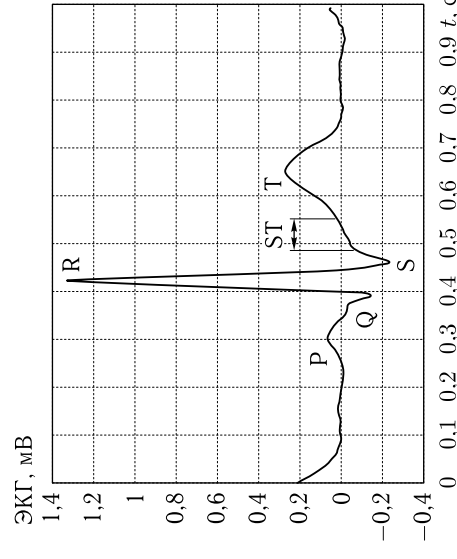


Рис. 3.1. Типичный цикл ЭКГ здорового человека в норме

следующие основные компоненты, соответствующие различным фазам сердечного цикла:

- Р-зубец (Р-волна), соответствует фазе деполяризации (сокращения) предсердий;
- QRS-комплекс (желудочковый комплекс), соответствует фазе деполяризации (сокращения) желудочков;
- ST-сегмент, соответствует периоду окончания деполяризации и началом реполяризации желудочков;
- Т-зубец (Т-волна), соответствует фазе реполяризации желудочков.

Измерение амплитудных и временных параметров перечисленных компонентов кардиоцикла ЭКГ (расшифровка ЭКГ) лежит в основе методов автоматической интерпретации ЭКГ, задача которой состоит в выявлении патологических изменений сердечной мышцы и проводящей системы сердца.

В то же время в приборах и системах кардиологического наблюдения наибольшее внимание уделяется анализу сердечного ритма и его нарушений или отслеживанию динамических изменений в электрокардиограмме в ходе ее продолжительного наблюдения. Это предполагает непрерывную регистрацию и анализ электрокардиосигнала на протяжении большого числа сердечных циклов. При этом основной автоматического анализа служит измерение интервалов времени между последовательными сокращениями желудочков сердца (RR-интервалов) и анализ характера формы (морфологии) основных волн кардиоцикла. Нарушения сердечного ритма (аритмии) вызываются патологическими изменениями автоматизма и проводимости сердца. В разд. 3.6 рассматриваются наиболее характерные виды нарушений сердечного ритма, а также описываются алгоритмы их автоматического распознавания.

Еще одна важная задача анализа электрокардиосигнала при кардиологическом наблюдении — выявление ишемических изменений ЭКГ, которые проявляются в трансформациях формы ST-сегмента (участка кардиоцикла между окончанием S-зубца и началом Т-зубца). Виды патологических изменений ST-сегмента ЭКГ и методы их автоматической количественной оценки подробно описываются в разд. 3.7.

Во многих системах и приборах кардиологического наблюдения обеспечивается возможность анализа вариабельности сердечного ритма (BCR). Анализ BCR заключается в оценке статистических и частотных характеристик сигнала, образуемого последовательностью величин смежных RR-интервалов между желудочковыми комплексами, относящимися к фоновому ритму (NN-интервалов). Предполагается, что вычисляемые при этом показатели несут информацию о регуляторных функциях вегетативной нервной системы. Описанию алгоритмов автоматического анализа BCR посвящается разд. 4.1.

В условиях продолжительного непрерывного съема электрокардиосигнала, характерных для систем кардиологического наблюдения, неизбежно появление в сигнале помех (артефактов), затрудняющих анализ сигнала и измерение его параметров. Различают следующие основные виды помех.

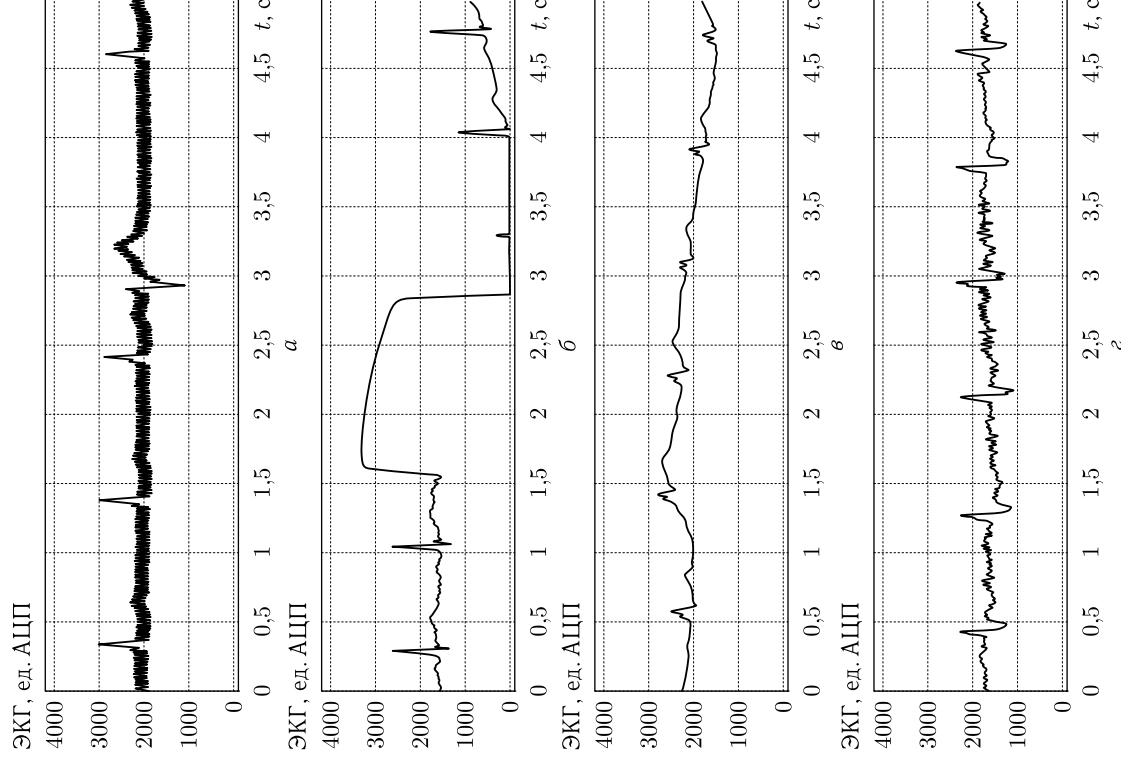


Рис. 3.2. Примеры некоторых видов помех в ЭКГ

1. *Сетевая наводка* (рис. 3.2, а) представляет собой синусоидальный сигнал с частотой 50 Гц (или 60 Гц в США и Канаде).
 2. *Контактный шум электродов* (рис. 3.2, б) — это преходящая помеха, вызываемая потерей контакта между электродом и кожей. Контактный шум электродов имеет вид случайным образом возникающего быстрого смещения (скачка) базовой линии.
 3. *Артефакты, связанные с движением* (рис. 3.2, в) — кратковременное изменение базовой линии, вызванное изменением сопротивления электрод–кожа, связанным с движением электрода.
 4. *Артефакты от мышечных сокращений (миографические шумы, рис. 3.2, г)*. Движения мышц пациента вызывают случайные широкополосные колебания, имеющие потенциал порядка одного милливольта.
 5. *Смещение базовой линии и модуляция амплитуды ЭКГ*, вызываемые дыханием, приблизительно могут быть представлены как добавление к ЭКГ синусоидальной составляющей с частотой дыхания.
 6. *Аппаратный шум*, генерируемый электронными устройствами съема и усиления сигналов. Вызывается насыщением входного усилителя. При этом информация об ЭКГ полностью теряется.
 7. *Электрохирургический шум* может быть представлен синусоидой большой амплитуды с частотами приблизительно между 100 кГц и 1 МГц. При этом информация об ЭКГ полностью теряется.
- Многообразие помех и их непредсказуемость существенно затрудняют автоматический анализ электрокардиосигнала. В связи с этим алгоритмы обработки ЭКС должны как включать процедуры подавления помех, так и предусматривать специальные меры, обеспечивающие устойчивость алгоритмов к шумам.

3.3. Предварительная цифровая фильтрация ЭКГ

На этапе предварительной обработки ЭКС решаются следующие основные задачи:

- устранение из сигнала той части помех, которая поддается подавлению;
- выявление фрагментов сигнала, где его автоматический анализ затруднен из-за чрезмерно высокого уровня помех;
- выделение информативных составляющих сигнала с целью обеспечения наилучших условий для последующих этапов обработки.

Наиболее часто для подавления помех используются методы цифровой фильтрации. Однако применение цифровых фильтров затрудняется тем, что помехи могут иметь спектральный состав, в значительной степени перекрывающийся со спектром полезного сигнала. Поскольку спектр случайной помехи обычно носит широкополосный характер, цифровая фильтрация способна эффективно удалять лишь те составляющие помехи, спектр которых лежит за пределами полосы частот, соответствующей электрокардиосигналу.

Обычно предварительная фильтрация ЭКС включает следующие основные этапы [3,4]:

- устранение сетевой наводки;
- анализ уровня помех;
- устранение высокочастотных помех;
- устранение низкочастотных помех.

Сетевая наводка 50 или 60 Гц носит характер детерминированного сигнала и наиболее легко поддается устранению. Для подавления сетевой наводки обычно применяют режекторные фильтры, частотная характеристика которых имеет провал на частоте помехи. При этом наиболее целесообразно использование адаптивной цифровой фильтрации, которая способна подавлять помеху с частотой 50 Гц, не затрагивая в то же время полезные составляющие сигнала на частотах, близких к 50 Гц.

Основная задача этапа анализа уровня помех — выявление участков сигнала, на которых дальнейший анализ невозможен из-за чрезмерно высокого уровня шумов. Распознавание помех выполняется как по простейшим критериям (выход сигнала за пределы динамического диапазона АЦП, высокоамплитудные импульсы или скачки сигнала), так и путем текущего контроля дисперсии сигнала или относительного содержания в нем высокочастотных составляющих.

Для устранения высокочастотных помех из электрокардиосигнала (в основном это высокочастотные составляющие миографической помехи) наиболее часто используются фильтры нижних частот, амплитудно-частотная характеристика которых имеет частоту среза 20–50 Гц. Недостаток таких фильтров — невозможность устранения тех составляющих помехи, спектр которых лежит в пределах полосы пропускания фильтров. Кроме цифровой фильтрации нижних частот предлагаются альтернативные методы, в значительной степени избавленные от указанного недостатка и основанные на следующих подходах:

- согласованная фильтрация;
- адаптивная фильтрация;
- использование вейвлет-преобразования;
- разложение сигнала на главные компоненты;
- нейронные сети.

Низкочастотная помеха (дрейф изоэлектрической линии) может быть устранена при помощи цифрового фильтра верхних частот с частотой среза 1–3 Гц. Однако это приводит к существенному искажению как формы QRS-комплекса, так и ST-сегмента ЭКГ. Поэтому в случаях когда одной из функций является анализ ишемических изменений ЭКГ, такой способ устранения помехи неприемлем. Для решения этой задачи наиболее часто прибегают к кусочно-полиномиальной аппроксимации с использованием сплайнов.

На рис. 3.3 и 3.4 приведены примеры фильтров соответственно нижних и верхних частот, используемых для предварительной обработки ЭКГ. Оба фильтра являются КИХ-фильтрами и рассчитаны на частоту

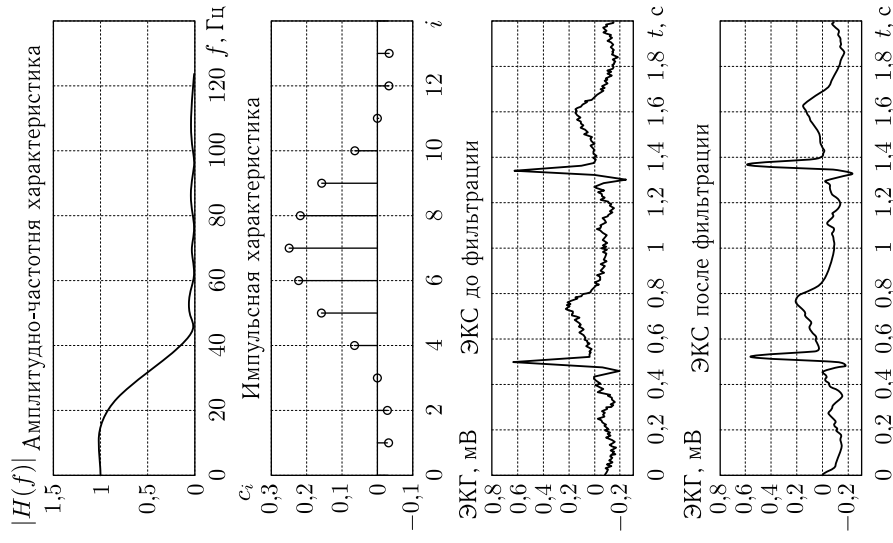


Рис. 3.3. Вверху: амплитудно-частотная и импульсная характеристики ФНЧ; внизу: пример фильтрации ЭКГ

дискретизации сигнала, равную 250 Гц. На этих же рисунках показаны примеры фильтрации фрагментов ЭКГ с помощью данных фильтров.

Как отмечалось выше, наиболее эффективное средство подавления сетевой наводки промышленной частоты 50 Гц — адаптивная цифровая фильтрация [3, 4, 12]. Это обусловлено тем, что адаптивные фильтры при сравнительно простоте их реализации обеспечивают высокую добротность подавления синусоидальной составляющей заданной частоты. Важным достоинством этих фильтров с точки зрения анализа ЭКГ является то, что они обладают способностью настраиваться на сетевую помеху и компенсировать ее, не влияя на близкие по частоте составляющие полезного сигнала. Рассмотрим принцип построения такого фильтра.

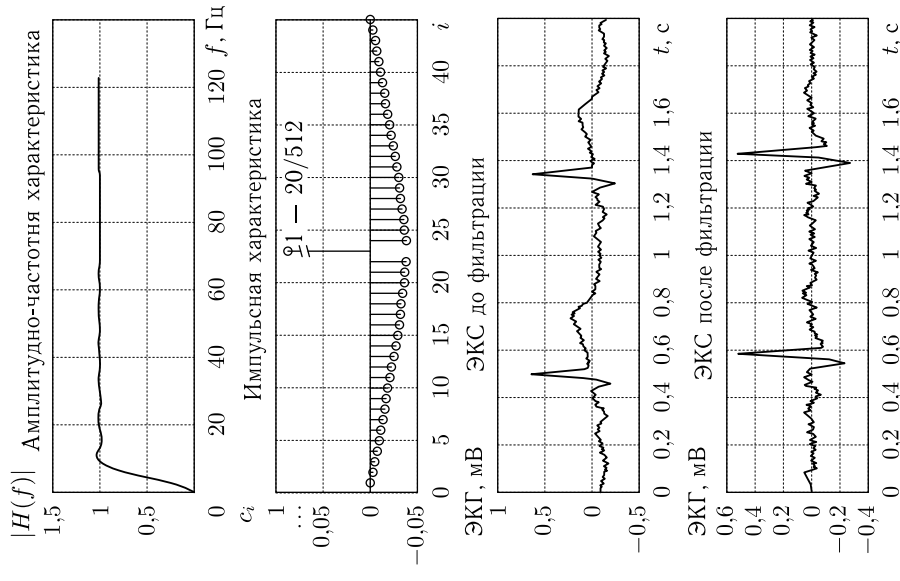


Рис. 3.4. Вверху: амплитудно-частотная и импульсная характеристики ФВЧ; внизу: пример фильтрации ЭКГ

Пусть последовательности отсчетов на входе и выходе фильтра (компенсатора помех) обозначены соответственно

$$x(1), x(2), \dots, x(i), \dots \quad \text{и} \quad y(1), y(2), \dots, y(i), \dots,$$

тогда любая i -я точка выходной последовательности может быть получена из соотношения

$$y(i) = x(i) - a(i), \quad i = 1, 2, 3, \dots, \quad (3.1)$$

где $a(1), a(2), \dots, a(i), \dots$ — рассчитанные фильтром отсчеты компенсирующего сигнала.

Известно тригонометрическое равенство

$$\sin(\theta + \delta) = 2 \sin \theta \cos \delta - \sin(\theta - \delta).$$

Умножим его на некоторую константу A :

$$A \sin(\theta + \delta) = 2 \cos \delta (A \sin \theta) - A \sin(\theta - \delta). \quad (3.2)$$

Теперь будем считать, что θ — текущее значение аргумента синусоидальной функции в точке взятия отсчета $a(i)$, δ — угловое расстояние между соседними отсчетами (по отношению к периоду компенсируемой помехи), а A — амплитуда этой синусоиды. Тогда (3.2) можно переписать в виде

$$a(i+1) = 2 \cos \delta \cdot a(i) - a(i-1), \quad (3.3)$$

где $a(i-1)$, $a(i)$, $a(i+1)$ — последовательные отсчеты синусоиды, отстоящие друг от друга на угол δ . Из (3.3) следует, что по двум известным соседним отсчетам синусоиды $a(i-1)$, $a(i)$ всегда можно предсказать значение следующего за ними отсчета $a(i+1)$. Это обстоятельство и положено в основу алгоритма подавления синусоидальной помехи.

Пусть для момента времени, соответствующего i -му отсчету входного сигнала $x(i)$, известны значение предыдущего отсчета $x(i-1)$, а также $a(i-1)$ и $a(i)$, равные ординатам соответствующих точек синусоидальной помехи. Тогда текущий и предшествующий отсчеты выходного сигнала могут быть найдены по формулам

$$\begin{aligned} y(i) &= x(i) - a(i), \\ y(i-1) &= x(i-1) - a(i-1). \end{aligned}$$

Рассмотрим разность

$$\Delta y = y(i) - y(i-1) = [x(i) - a(i)] - [x(i-1) - a(i-1)]. \quad (3.4)$$

Оказывается, что чем точнее определены отсчеты компенсирующего сигнала $a(i-1)$ и $a(i)$, тем ближе к нулю становится эта разность для каждого значения i , т.е. величина Δy используется в качестве критерия оценки полной выходной мощности фильтра. Поэтому подстройка фильтра под помеху заключается в том, что на каждом шаге фильтрации (для каждого следующего i) оценивается знак выражения (3.4) и в зависимости от того, положителен он или отрицателен, значение $a(i)$ соответственно наращивается или уменьшается на фиксированную величину Δa так, чтобы скорректировать Δy в сторону нуля. Полученное в результате значение $a(i)$ используется для вычисления

выходного отсчета по формуле (3.1). Далее с использованием (3.3) вычисляется предполагаемое значение отсчета компенсирующего сигнала $a(i+1)$, после чего вся процедура повторяется для следующего входного отсчета, и т.д.

Скорость сходимости, устойчивость, ширина полосы подавления и точность настройки фильтра определяются тремя параметрами: шагом адаптации Δa , постоянной $\cos \delta$ и точностью вычисления промежуточных величин. Выбор константы $\cos \delta$ выполняется из условия

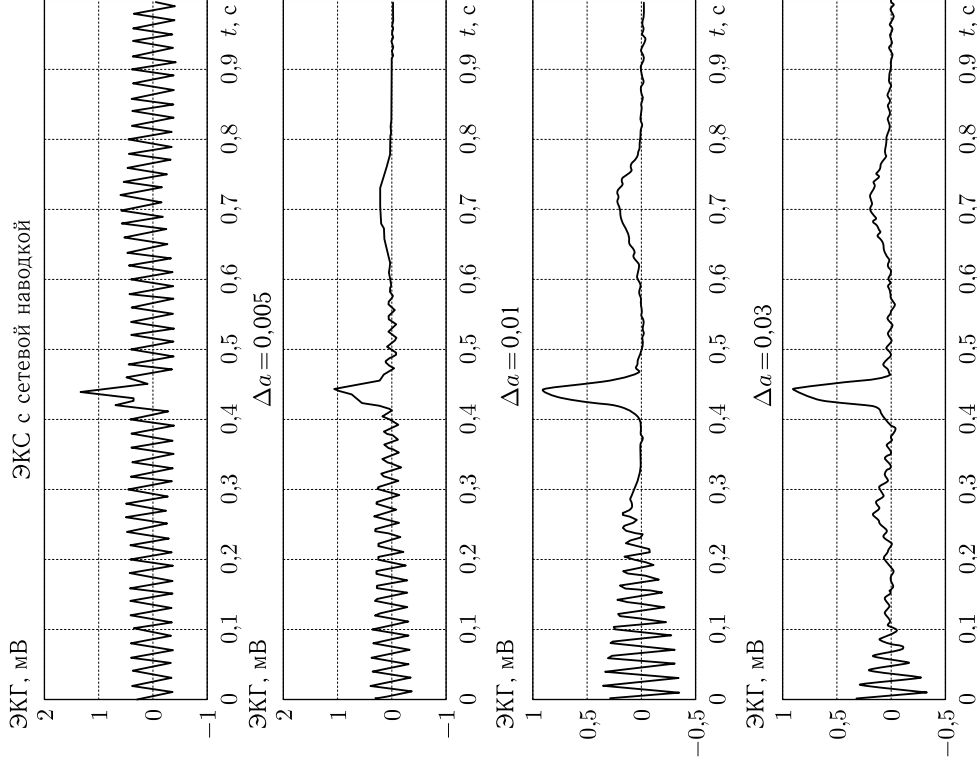


Рис. 3.5. Устранение сетевой наводки из электрокардосигнала при разных значениях шага адаптации Δa

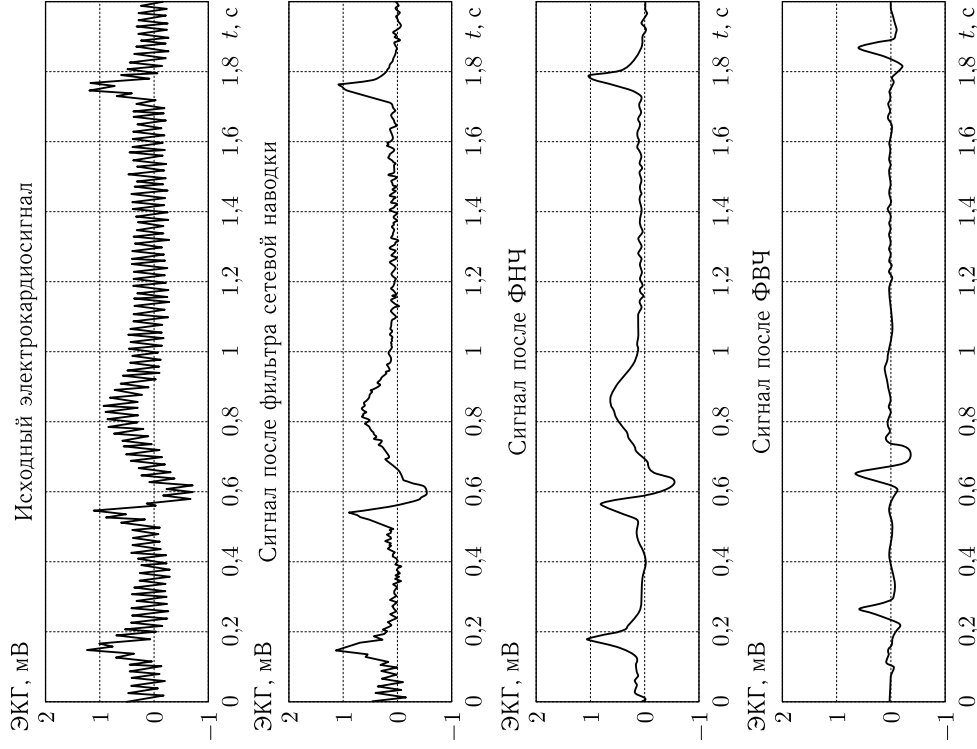


Рис. 3.6. Пример последовательной фильтрации фрагмента электрокардиосигнала с использованием цепочки цифровых фильтров

$\delta = 360^\circ f_{\text{сн}}/f_d$, где $f_{\text{сн}} = 50$ Гц — частота сетевой помехи, а f_d — частота дискретизации сигнала. Например, при $f_d = 250$ Гц получим

$$\delta = 360^\circ f_{\text{сн}}/f_d = 360^\circ \cdot 50/250 = 72^\circ, \\ \cos \delta = \cos 72^\circ = 0,309.$$

Наибольшее влияние на качество фильтрации оказывает выбор шага адаптации компенсирующего сигнала Δa . На рис. 3.5 показаны примеры работы фильтра при разных значениях этого параметра. Можно видеть, что при больших значениях Δa фильтр настраивается очень

быстро, но заметно искажает полезный сигнал. С уменьшением Δa искажения сигнала становятся меньше, но растет время настройки.

На рис. 3.6 показан пример последовательной фильтрации фрагмента электрокардиосигнала с использованием цепочки цифровых фильтров:

- фильтра нижних частот;
- фильтра верхних частот;
- адаптивного фильтра нижних частот.

3.4. Обнаружение опасных аритмий по спектральному описанию электрокардиосигнала

В задачах диагностики и наблюдения большое внимание уделяется анализу биомедицинских сигналов в частотной области. Особый интерес представляет спектральный анализ ЭКС при нарушениях ритма, приводящих к появлению трепетания и мерцания желудочков сердца. Сигнал при этом приобретает волнообразную форму и имеет довольно характерное спектральное описание. Автоматически эти волны довольно сложно идентифицировать, в то же время задача надежного обнаружения опасных аритмий остается первоочередной для систем кардиологического наблюдения.

Анализ сигналов в частотной области, как правило, предполагает оценку мощности основных гармонических компонент, а также выявление некоторых особенностей распределения спектральной плотности мощности (СПМ) [3.8]. В задачах распознавания биомедицинских сигналов это часто приводит к выбору содержательных признаков, имеющих явно различные для разных классов характеристик (например, наличие узкополосных компонент спектра, выраженность основного пика СПМ и др.) [3.1, 3.13, 3.18].

Исследуя распределения объектов в пространстве выбранных признаков, можно решить задачу классификации сигналов, реализуя линейные решающие функции или конструируя более сложные поверхности решений. Однако часто оптимизация вида решающих правил не приводит к требуемому качеству распознавания сигналов. Это происходит из-за того, что выбранные признаки, отобранные подчас интуитивно или опытным путем, не несут информации, существенной для задач классификации. Система признаков может оказаться неполной, а включение в решающие правила некоторых из них может даже снизить эффективность распознающей системы. Действительно, выбор множества признаков, обеспечивающих необходимое качество классификации, представляет собой трудную задачу [3.11]. В работах [3.5, 3.14, 3.15] решается ее предлагается путем формирования классификационного описания, рассматривающего полный ряд спектральных коэффициентов как упорядоченный набор параметров. Это приводит

к возможности построения в пространстве выбранных признаков эффективной разделяющей гиперплоскости [3.6, 3.7].

Пусть исходный сигнал $x(t)$ представляет собой последовательность N отсчетов $x_n, n = 0, \dots, (N-1)$, заданных с частотой дискретизации $f_d = 1/h$, Гц. Тогда на интервале времени $T = Nh$ может быть получена выборочная оценка СПМ [3.8]

$$G(f_k) = \frac{2h}{N} \left| \sum_{n=0}^{N-1} x_n \exp \left(j - \frac{2\pi kn}{N} \right) \right|^2,$$

где $f_k = \frac{k}{T} = \frac{k}{Nh}$, h — шаг дискретизации, $k = 0, 1, \dots, N-1$.

Далее следует анализ полученной оценки СПМ в области значений $k = 0, 1, 2, \dots, N/2$, что соответствует ограничению спектра до $f_c = f_d/2$ (частоты Найквиста).

Учитывая известный эффект просачивания энергии через боковые лепестки спектрального окна, который возникает в результате усечения исследуемого процесса прямоугольным окном шириной T , необходимо предварительно провести сглаживание концов исходной реализации. Для этого применяют специальные взвешивающие функции, например косинусоидальную, что позволяет уменьшить величину смещения выборочной оценки спектра [3.8].

Полученную оценку СПМ можно использовать для формирования спектральных признаков и построения решающих правил. Однако предварительно необходимо ограничить анализируемую область частот, что уменьшит размерность исходного спектрального описания и упростит процедуру распознавания сигналов. Для этого можно оценить расхождение выборочных описаний сигналов, сформировав обучающие выборки данных для заданных классов (ω_1 и ω_2) и вычислив для них оценки СПМ. В качестве меры близости между ω_1 и ω_2 можно использовать критерий, основанный на анализе среднего межгруппового расстояния:

$$\rho^{(k)}(\omega_1, \omega_2) = (1/n_1 n_2) \cdot \sum_{\mathbf{G}_q^{(k)} \in \omega_1} \sum_{\mathbf{G}_j^{(k)} \in \omega_2} \rho(\mathbf{G}_q^{(k)}, \mathbf{G}_j^{(k)}), \quad (3.5)$$

где n_1 и n_2 — число объектов, образующих классы ω_1 и ω_2 ; k — шаг итерации; $\mathbf{G}_q^{(k)}, \mathbf{G}_j^{(k)}$ — спектральное представление объектов q и j для классов ω_1 и ω_2 , используемое на k -м шаге (для k частотных составляющих).

Рассмотренный подход использовался при обработке ЭКС для решения задачи распознавания опасного нарушения ритма (фибрилляции желудочков сердца) (класс ω_1) на фоне других аритмий, не представляющих опасности для жизни больного (нормальный ритм с одиночными экстрасистолами, групповая экстрасистолия, другие нарушения работы предсердий и желудочков сердца) (класс ω_2) [3.5, 3.15].

В качестве примера на рис. 3.7 приведены реализации ЭКС и соответствующие выборочные оценки нормированного спектра для желудочковой фибрилляции (ЖФ) (а, б) и фонового ритма (ФР) (в, г). Фрагменты сигналов представлены записями ЭКГ длительностью 2 с, дискретизованными с частотой 250 Гц (50 ед. АЦП соответствуют 1 мВ).

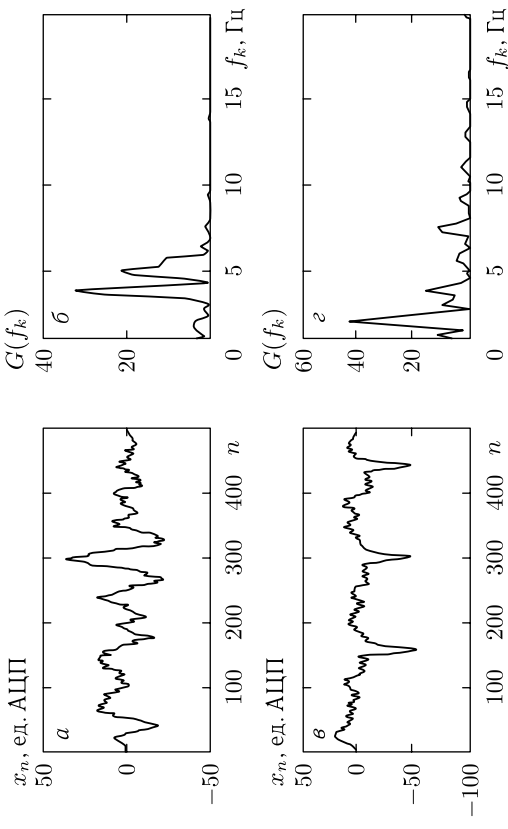


Рис. 3.7. Реализации ЭКС и соответствующие выборочные оценки нормированного спектра: а, б — для желудочковой фибрилляции; в, г — для фонового ритма

На рис. 3.8 приведены графики зависимости межгруппового расстояния от числа k выборок СПМ, включенных в спектральное описание сигналов. На рис. 3.8, а показана интегральная кривая $\rho^{(k)}(\omega_1, \omega_2)$, а на рис. 3.8, б — зависимость расстояния $\rho_1^{(k)}(\omega_1, \omega_2)$, вычисленного отдельно для каждого спектрального признака.

Зависимости получены по результатам обработки 60 записей ЭКС, отобранных из электрокардиографической базы MIT-BIH [3.17].

Анализ полученных зависимостей позволил выделить эффективную область анализа СПМ с верхней частотой, равной $f_{\max} = 14$ Гц ($k = 28$), поскольку более высокие частоты, как видно из рисунков, не вносят существенного вклада в оцениваемую меру расхождения классов.

Далее применяется процедура группировки спектральных коэффициентов сглаживанием их значений в границах заданных областей. Известно, что выборочный спектр, вычисляемый для независимых частот f_k , подчиняется закону χ^2 -распределения с двумя степенями свободы. Получаемая для каждой частоты величина не является

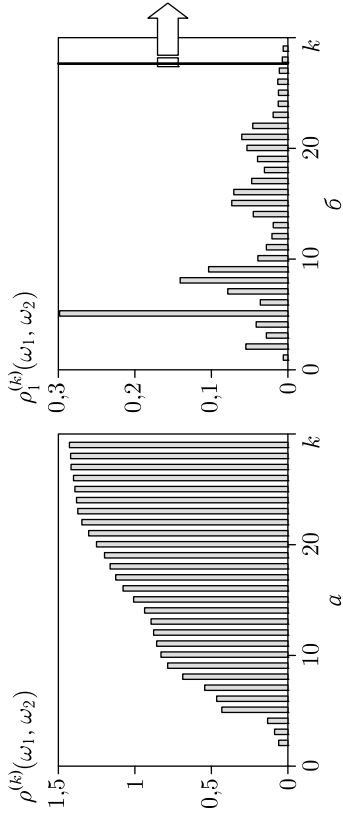


Рис. 3.8. Зависимость межгруппового расстояния от числа k выборов СПМ: a — интегральная кривая; b — значение $\rho_1^{(k)}(\omega_1, \omega_2)$ для каждого k -го признака

состоятельной оценкой СПМ, так как ее стандартное отклонение сравнимо по величине со средним значением $G(f_k)$ [3.8]. Используя сглаживание по l независимым частотам, перейдем к χ^2 -распределению с $2l$ степенями свободы, что приводит к уменьшению величины случайной ошибки в $\sqrt{l}$ раз. Таким образом, процедура объединения смежных частот способствует уменьшению дисперсии оценки СПМ, т. е. приводит к более устойчивым спектральным признакам.

Формирование набора спектральных признаков можно осуществить одним из предлагаемых способов [3.5, 3.15]. Первый способ основан на использовании периодограммной оценки Даньелла, формирующей новый ряд спектральных отсчетов путем усреднения $(2P + 1)$ соседних частот [3.8]:

$$G_D(f_i) = \frac{1}{2P + 1} \sum_{n=i-P}^{i+P} G(f_n).$$

При этом снижается размерность частотного описания сигналов, а также повышается устойчивость спектральных оценок. Учитывая возможное ухудшение разрешения соседних частот, можно использовать перекрывающиеся сегменты, что обеспечит более детальный анализ сигнала в частотной области. Но в этом случае корреляция соседних оценок станет отличной от нуля, а число оцениваемых признаков практически удвоится. Эффективность выбранного спектрального окна и способа формирования отсчетов СПМ может быть также оценена по критерию максимизации расстояния между заданными классами сигналов (3.5).

Второй способ предусматривает автоматическое формирование эффективных частотных областей путем нахождения оптимальных границ для группировки смежных спектральных коэффициентов. Вначале задается максимально допустимое число интервалов L , в пределах каждого из которых объединяются точечные оценки СПМ. Это значение определяет и число признаков, которые могут быть использованы

при построении дискриминантных функций. Затем оцениваются суммарные площади под кривой СПМ:

$$G_{Lp} = \frac{1}{Nh} \sum_{k=m_p}^{m_{p+1}} G(f_k), \quad p = 1, \dots, L.$$

Перебор вариантов расположения границ областей (m_p, m_{p+1}) в каждой из L частотных областей позволяет получить достаточно большой набор векторов $\{\mathbf{G}^{(L)}\}$, из которых наилучший может быть определен по максимальному значению критерия (3.5). Для сокращения времени анализа может быть использована процедура градиентного поиска экстремума целевой функции. Аналогичные действия выполняются и на последующих шагах обработки при последовательном уменьшении числа интервалов $(L, L-1, L-2, \dots)$. В результате формируется упорядоченный набор векторов $\{\mathbf{G}^{(L)}, \mathbf{G}^{(L-1)}, \mathbf{G}^{(L-2)}, \dots\}$, которые сравниваются друг с другом по значениям функции (3.5), и выбирается окончательный вариант спектрального описания $\mathbf{G} = (G_1, \dots, G_p, \dots, G_L)$.

На рис. 3.9 представлены зависимости межгруппового расстояния $\rho_i(\omega_1, \omega_2)$ от целочисленного аргумента i , связанного с местоположением границ частотных областей. График построен для двух областей частот ($L = 2$), и максимум функции $\rho_i(\omega_1, \omega_2)$, $i = 10$, приходится на последовательность границ 0,244; 6,832; 14,64 Гц. Все частоты крагны 0,244 Гц, что связано с дополнением нулями последовательности $x(n)$, $n = 0, \dots, N-1$ до $N = 1024$ отсчетов. Аналогичные вычисления для $L = 3, 4, 5$ привели к следующим результатам:

0,244; 2,44; 6,1; 14,64 Гц при $L = 3$;

0,244; 2,44; 5,124; 10,004; 14,64 Гц при $L = 4$;

0,244; 2,44; 5,124; 10,004; 12,444; 14,64 Гц при $L = 5$.

Видно, что при переходе к $L = 5$ четвертая область (10,004; 14,64 Гц) образовала два новых кластера, остальные области остались

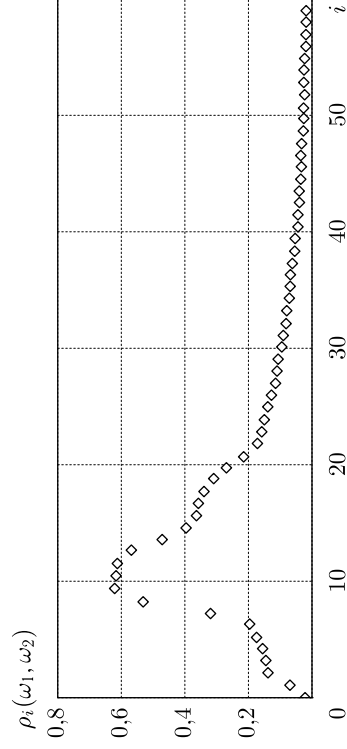


Рис. 3.9. Зависимость межгруппового расстояния $\rho_i(\omega_1, \omega_2)$ от местоположения границ частотных областей (i)

прежними. В результате получены наборы признаков, сформированные в виде суммарных значений СПМ по L частотным областям.

Последний этап исследований связан с распознаванием классов ЭКС ω_1 и ω_2 . Для построения решающих функций в пространстве спектральных признаков использовался метод нахождения линейного дискриминанта Фишера, рассмотренный в гл. 1. Сравнительный анализ полученных спектральных описаний проводился по критерию оптимизации J .

На рис. 3.10 в виде гистограмм представлены распределения проекций объектов на вектор $\mathbf{W}$ ($y = \mathbf{W}^T \mathbf{G}^{(L)}$) для сглаженных оценок СПМ в области частот 0–15 Гц. Слева (рис. 3.10, а) показан результат, полученный с использованием периодограммы Даньелла для $L = 15$ отсчетов частоты при ширине окна $\Delta f = 0,976$ Гц. Справа (рис. 3.10, б) приведены распределения, полученные при формировании $L = 4$ частотных областей спектра по критерию максимизации межгрупповых расстояний. Здесь класс ω_1 представлен реализациями ЖФ, а ω_2 — сигналами ФР.

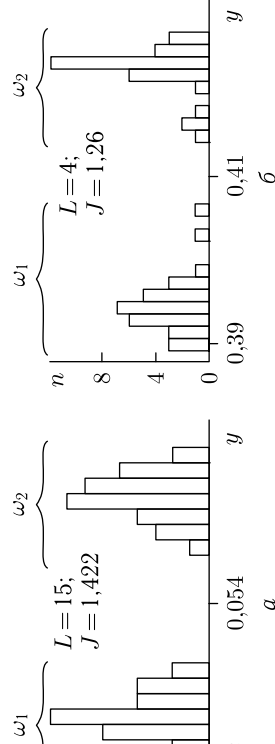


Рис. 3.10. Гистограммы распределений проекций объектов на вектор $\mathbf{W}$ для классов ω_1 и ω_2 : а — использовано окно Даньелла с $\Delta f = 0,976$ Гц; б — проведена оптимизация частотных областей

Анализ полученных данных показывает, что объекты в обоих случаях образуют достаточно компактные группы и классификацию сигналов можно эффективно осуществлять, используя один из предложенных способов построения пространства спектральных признаков. Важно отметить, что, применяя оптимизацию частотных областей, можно существенно сократить число используемых признаков. В результате задача обнаружения ЖФ была решена путем формирования четырех частотных областей СПМ. Аппроксимация выборочных распределений проекций объектов y на ось решений $\mathbf{W} = (0,433; 0,39; 0,397; 0,703)$ нормальным законом распределения позволила выбрать порог $\alpha = 0,415$ и оценить теоретические ошибки распознавания классов ω_1 и ω_2 . При этом ошибка классификации даже при условии минимизации риска, связанного с пропуском ЖФ, не превышает 0,5%. Результаты проверки эффективности обнаружения ЖФ на контрольной выборке данных, включающей более 200 коротких реализаций ЭКГ из базы

MIT-BIH, показали, что чувствительность метода $S_e = 0,98$ и специфичность $S_p = 0,96$ [3.15]. Эти результаты отличаются от теоретических оценок ввиду большого разнообразия представленных в контрольной выборке форм ЭКС, но уровень надежности классификации остается достаточно высоким.

В работе [3.16] рассмотрен способ коррекции положения разделяющей функции путем нахождения дополнительного вектора по критерию Фишера J и перехода в двумерное признаковое пространство векторов $\mathbf{W}_1, \mathbf{W}_2$ (см. разд. 1.3 настоящей монографии). В ходе экспериментов установлено, что введение дополнительного вектора $\mathbf{W}_2$ для спектрального описания ЭКС, использующего периодограммную оценку Даньелла ($L = 15, \Delta f = 0,976$ Гц), в 1,5 раза увеличило расстояние между ближайшими соседями рассматриваемых классов ω_1 (ЖФ), ω_2 (ФР). Такой переход к отображению объектов на плоскости улучшает разделимость классов и позволяет использовать более сложные разделяющие поверхности, аппроксимируя последние кусочно-линейными функциями.

3.5. Алгоритм обнаружения QRS-комплекса ЭКГ

QRS-комплекс представляет собой наиболее выраженный элемент кардиоцикла ЭКГ, соответствующий фазе деполяризации (сокращения) желудочков сердца. Обнаружение QRS-комплекса служит основой всех следующих этапов анализа ЭКГ. Решение задачи обнаружения QRS-комплекса в системах длительного непрерывного контроля ЭКГ осложняется разнообразием и изменчивостью форм комплексов во времени, а также неизбежными при длительном наблюдении помехами.

Алгоритмы обнаружения QRS-комплекса решают следующие основные задачи:

- обнаружение (локализация) QRS-комплекса в ЭКГ на фоне других элементов кардиоцикла;
- определение условной опорной точки комплекса, служащей для измерения RR-интервала;
- определение границ QRS-комплекса.

В литературе описываются многочисленные методы обнаружения QRS-комплекса [3.4, 3.9, 3.10], которые могут основываться на следующих подходах:

- вычисление производных сигнала;
- линейная цифровая фильтрация;
- адаптивная цифровая фильтрация;
- согласованная фильтрация;
- вейвлет-преобразование;
- нейронные сети;
- синтаксические методы.

Рассматриваемый далее алгоритм [3.4] относится к наиболее распространенным решениям. Он основан на выделении QRS-комплексов

при помощи цепочки процедур цифровой фильтрации и последующем анализе полученного сигнала с использованием пороговых правил.

Поскольку большинство временных интервалов кардиоцикла имеют тенденцию к укорочению с ростом частоты сердечных сокращений, то это необходимо учитывать при разработке алгоритма обнаружения желудочковых комплексов, так как частота ритма сердца даже у одного и того же больного может меняться в ходе наблюдения в 2–3 раза. На рис. 3.11 изображен типичный кардиоцикл и указаны возможные пределы изменения временных интервалов между его основными характеристичными точками. Для учета возможных вариаций временных и амплитудных характеристик алгоритм обнаружения содержит ряд параметров, адаптирующихся к текущим средним значениям как частоты сердечных сокращений, так и размаха QRS-комплекса.

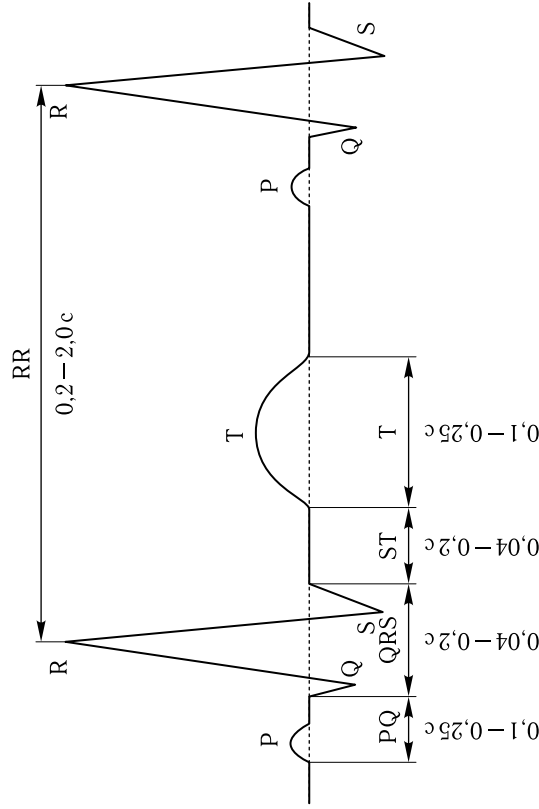


Рис. 3.11. Типичный кардиоцикл и возможные пределы изменения временных интервалов между его основными характеристичными точками

Алгоритму обнаружения QRS-комплекса ЭКГ, кроме описанной в 3.3 предварительной цифровой фильтрации, обычно предшествует этап выделения QRS-комплекса, включающий дифференцирующий ЦФ, процедуру взятия модуля сигнала и сглаживающий фильтр. Если задаться значением частоты дискретизации $f_d = 250$ Гц, то в качестве простого дифференцирующего ЦФ может использоваться, например, фильтр, определяемый выражением

$$y(n) = (1/4)[x(n) - x(n-1) - x(n-2) + x(n-4) + x(n-5) + x(n-6)].$$

АЧХ этого фильтра приведена на рис. 3.12, а. Как видим, фильтр обладает свойствами дифференциатора в полосе частот от 0 до примерно 30 Гц (см. 2.2). Дифференцирование позволяет выделить QRS-комплекс на фоне остальных волн ЭКГ, а также избавиться от низкочастотной помехи. Для удобства дальнейшего анализа сигнал берется по модулю и подвергается дополнительному сглаживанию, что повышает устойчивость обработки. Сглаживание может быть выполнено, например, с помощью такого фильтра:

$$y(n) = (1/8)[x(n) - 2x(n-1) - 2x(n-2) + 2x(n-3) + x(n-4)].$$

АЧХ этого фильтра приведена на рис. 3.12, б.

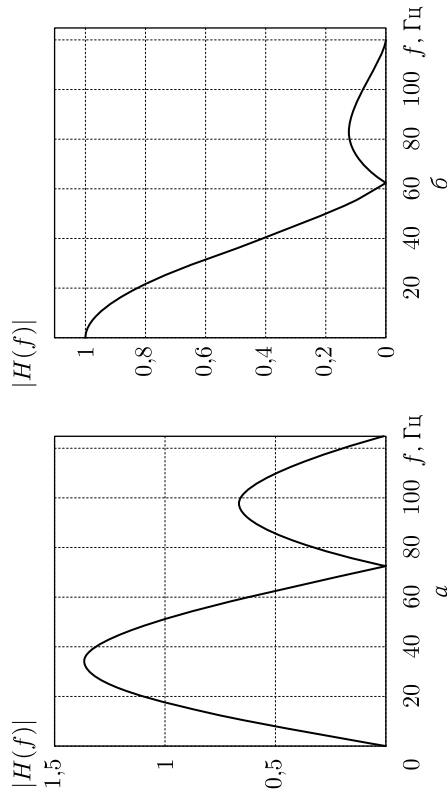


Рис. 3.12. Амплитудно-частотные характеристики цифровых фильтров, используемых для выделения желудочкового комплекса ЭКГ

На рис. 3.13 показаны все этапы выделения QRS-комплекса с использованием цифровых фильтров. На рис. 3.13, а приведен фрагмент исходной ЭКГ, содержащий три QRS-комплекса. В сигнале присутствуют сетевая помеха, дрейф изоэлектрической линии, а также незначительные высокочастотные шумы. На рис. 3.13, б показан этот же фрагмент после предварительной обработки фильтрами сетевой наводки и нижних частот, описанными в разд. 3.3. Используемый ФНЧ имеет частоту среза около 30 Гц. Поэтому оставшиеся в сигнале частотные составляющие полностью укладываются в область дифференцирования используемого далее фильтра-дифференциатора. Можно видеть, что сигнал с выхода дифференцирующего ЦФ (рис. 3.13, в) представляет собой производную от вышеуказанного сигнала: участкам возрастания и убывания соответствуют положительные и отрицательные значения, а точкам экстремума — нулевые. В полученном сигнале амплитуда колебаний в области QRS-комплексов стала существенно выше по отношению к другим элементам ЭКГ, а дрейф изолинии исчез, что и было

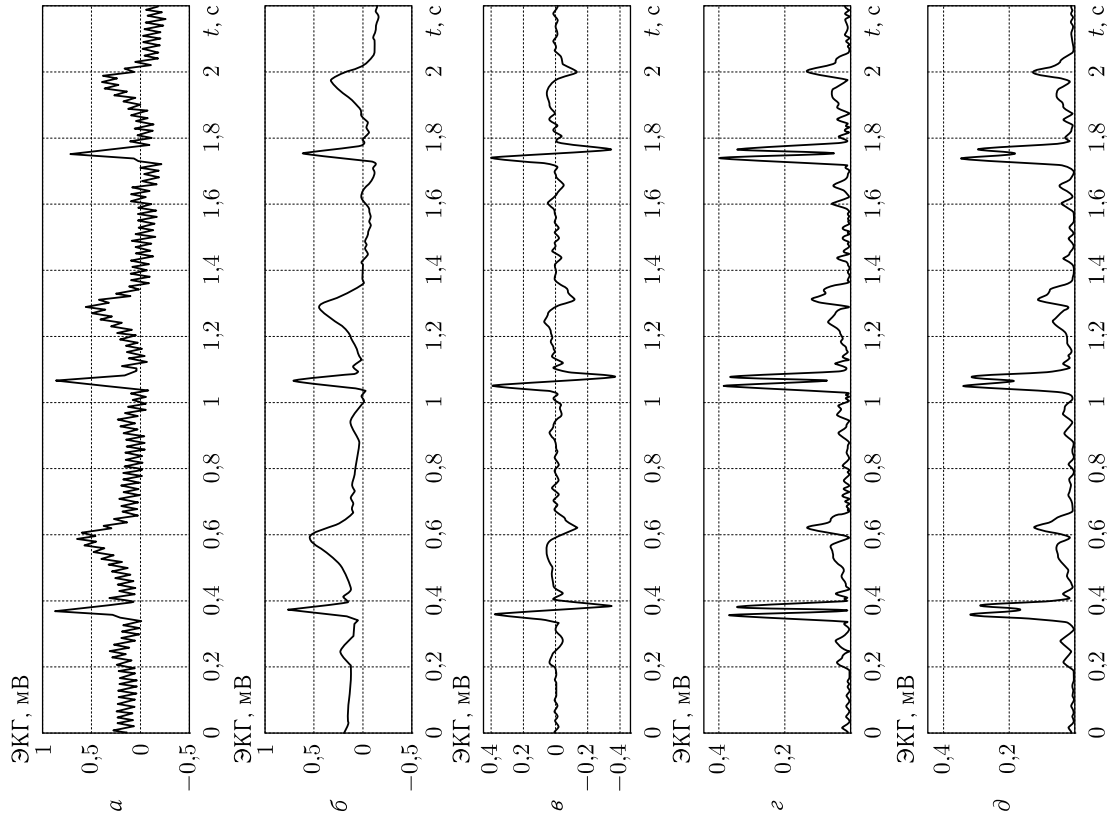


Рис. 3.13. Этапы выделения желудочкового комплекса ЭКГ

задачей дифференцирования ЭКГ. На рис. 3.13, *г* и *д* показаны соответственно сигналы после процедур взятия модуля и дополнительного сглаживания с целью повышения устойчивости обработки.

На вход алгоритма обнаружения желудочковых комплексов ЭКГ поступает сигнал после описанных выше процедур выделения. Принцип работы алгоритма иллюстрируется на рис. 3.14.

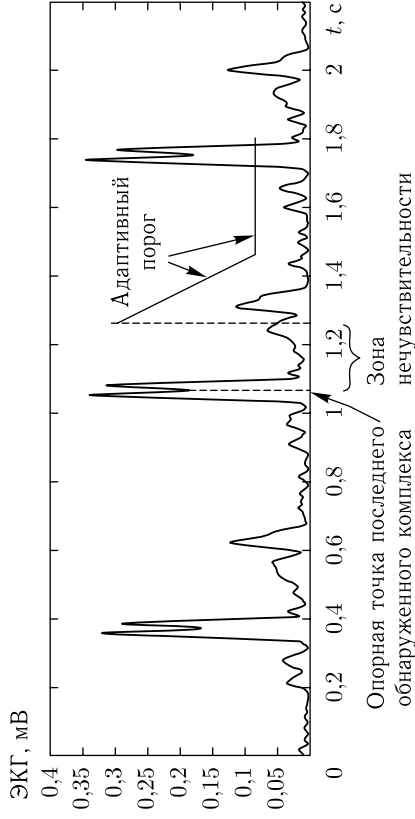


Рис. 3.14. Иллюстрация принципа работы алгоритма обнаружения желудочкового комплекса ЭКГ

Предполагается, что к моменту начала поиска очередного желудочкового комплекса известны:

- опорная точка последнего обнаруженного желудочкового комплекса (точка в пределах обнаруженного желудочкового комплекса относительно которой вычисляется оценка RR-интервала);
- текущее среднее значение размаха желудочковых комплексов;
- текущие значения RR-интервалов.

На первом этапе работы алгоритма вычисляются значения тех параметров, которые зависят от текущих средних значений размаха желудочковых комплексов и RR-интервалов. Это обеспечивает адаптивность алгоритма к текущим изменениям характеристик сигнала. Далее алгоритм пропускает промежуток времени от исходной точки до окончания зоны нечувствительности (адаптивно определяемой как функция от текущего среднего значения RR-интервала), так как предполагается, что на этом участке не может встретиться очередной комплекс. Начиная с отсчета, соответствующего моменту окончания зоны нечувствительности, выставляется линейно спадающий порог, начальное значение которого определяется с учетом среднего размаха желудочковых комплексов. Наклон и уровень окончания спада порога также являются адаптивно определяемыми параметрами. Такой подход позволяет снизить вероятность ложного срабатывания алгоритма из-за высокоамплитудных T-зубцов ЭКГ. Обнаружение очередного QRS-комплекса происходит при пересечении сигналом текущего уровня порога.

После того как обнаружен очередной QRS-комплекс, выполняется определение условной опорной точки, относительно которой измеряется значение RR-интервала. Во многих алгоритмах в качестве такой точки используется точка максимума R-зубца. Однако этот способ измерения может давать существенную погрешность в случаях, когда R-зубец

имеет расщепленную форму или когда форма QRS-комплекса подвергается трансформациям, связанным с дыханием. Более устойчивые результаты удается получить, используя некоторый интегральный метод (например, вычисление площади под кривой QRS-комплекса). Аналогичная процедура используется и для поиска границ обнаруженного QRS-комплекса [3.4].

3.6. Анализ форм ЭКГ и распознавание нарушений ритма

Одна из основных задач систем автоматического непрерывного контроля сердечной деятельности по ЭКГ — распознавание нарушений сердечного ритма (аритмий) [3.4, 3.10]. На рис. 3.15 показаны примеры ЭКГ при нормальном ритме сердца и при некоторых его нарушениях:

— нормальный синусовый ритм (рис. 3.15, а) — последовательность идентичных кардициклов нормальной формы, следующих друг за другом с приблизительно одинаковыми интервалами;

— суправентрикулярная (предсердная) экстрасистола (рис. 3.15, б) — появление на фоне нормального синусового ритма преждевременного сокращения сердца нормальной формы, вызванного импульсом возбуждения из предсердия; за экстрасистолой, как правило, следует удлиненный интервал, называемый компенсаторной паузой;

— выпадение желудочкового комплекса (рис. 3.15, в) — отсутствие QRS-комплекса после Р-зубца на фоне нормального синусового ритма;

— желудочковая экстрасистола (рис. 3.15, г) — появление на фоне нормального синусового ритма преждевременного сокращения сердца патологической формы, вызванного спонтанным импульсом возбуждения из одного из желудочков; за экстрасистолой, как правило, следует удлиненный интервал, называемый компенсаторной паузой;

— групповая желудочковая экстрасистола (рис. 3.15, д) — появление на фоне нормального синусового ритма двух и более последовательных преждевременных сокращений сердца патологической формы, вызванных спонтанными импульсами возбуждения из желудочков.

Как можно видеть из приведенных иллюстраций, нарушения ритма сердца проявляют себя на ЭКГ в виде изменения интервалов между смежными желудочковыми комплексами, а также морфологии элементов кардиоцикла, в частности QRS-комплекса. В связи с этим одной из задач анализа ЭКГ является идентификация принадлежащего каждому обнаруженного желудочкового комплекса либо к категории комплексов фонового ритма (обычно называемой «норма»), либо к какому-либо виду патологических форм. Решение этой задачи служит основой для ряда этапов последующего анализа сигнала, в частности следующих:

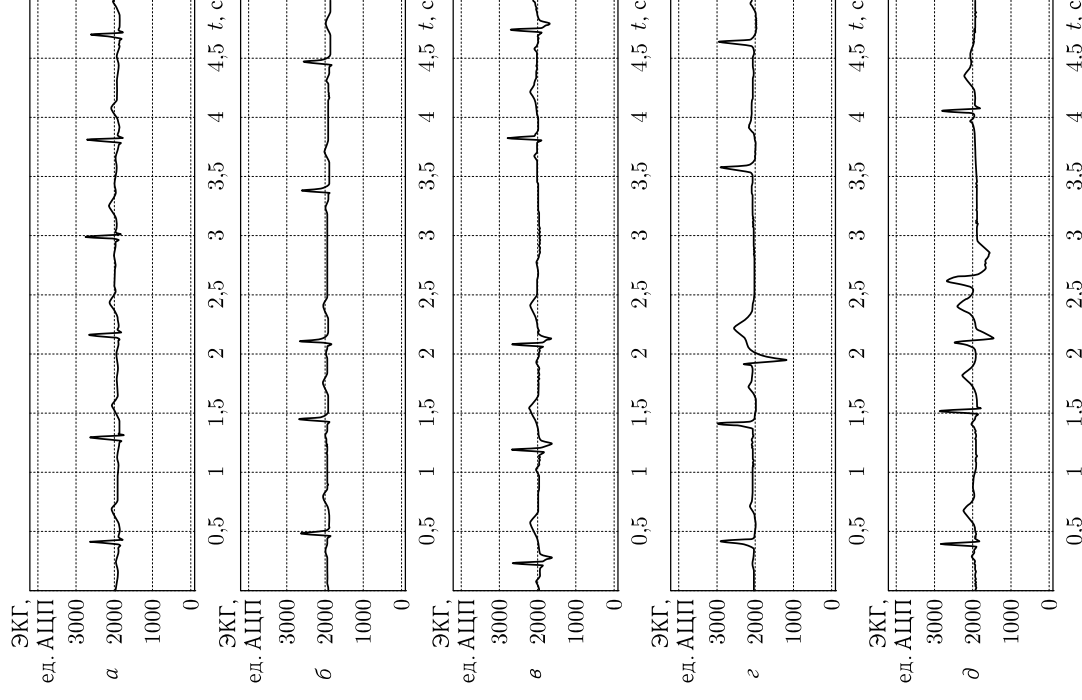


Рис. 3.15. Примеры ЭКГ при нормальном синусовом ритме (а) и при различных нарушениях ритма (б, в, г, д)

— распознавание аритмий (основано на одновременном анализе значений RR-интервалов и данных о характере форм желудочковых комплексов);

— анализ вариабельности сердечного ритма (основан на анализе последовательности NN-интервалов, т.е. RR-интервалов между смежными QRS-комплексами, относящимися к фоновому ритму);

— анализ ST-сегмента ЭКГ (выполняется только для QRS-комплексов фонового ритма).

Анализ морфологии QRS-комплексов затрудняется следующими факторами:

- разнообразие форм ЭКГ у разных людей, а также невозможность выбора каких-либо «образцовых» форм сигнала;
- изменчивость форм желудочковых комплексов во времени;
- наличие помех, искажающих форму сигнала;
- вариабельность формы кардиоцикла ЭКГ, связанная с изменением ЧСС, дыханием, изменением состояния или положения тела пациента.

Как правило, процедура анализа (классификации) форм QRS-комплексов включает следующие основные этапы:

- измерение информативных признаков QRS-комплексов;
- формальное разбиение на кластеры по сходству морфологий;
- идентификация кластеров (присвоение каждому из кластеров условного идентификатора «норма», «патология» или «неопределенность»).

В основе любой процедуры кластеризации QRS-комплексов всегда лежит какой-либо метод вычисления признаков формы, позволяющих оценить сходство или различие морфологий комплексов. Наибольшее распространение находят следующие три вида методов:

- анализ параметров формы, вычисляемых во временной области;
- использование взаимной корреляционной функции;
- анализ спектральных параметров.

При использовании параметров формы, вычисляемых во временной области, для каждого из QRS-комплексов анализируемой ЭКГ рассматриваются четыре признака формы [3.4]: длительность T , S ; размах A , mB ; смещение относительно нулевой линии S , mB , и суммарная площадь волн P , $mB \cdot c$ (рис. 3.16). Расчет признаков выполняется по следующим формулам:

$$T = K \Delta T;$$

$$A = \max_{k=1}^K x(k) - \min_{k=1}^K x(k);$$

$$S = \frac{1}{2} \left[\max_{k=1}^K x(k) + \min_{k=1}^K x(k) \right] - x(1);$$

$$P = \sum_{k=1}^K |x(k) - x(1)|,$$

где K — число отсчетов сигнала в пределах QRS-комплекса; $\Delta T = 1/f_d$ — интервал дискретизации сигнала при частоте отсчетов f_d ; $x(k)$ — отсчеты сигнала в пределах данного QRS-комплекса ($k = 1, 2, \dots, K$). При расчете признаков формы значение первого отсчета

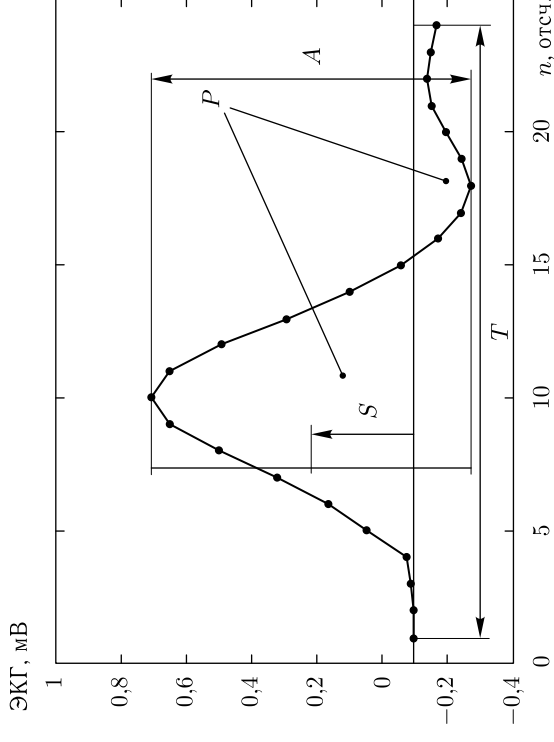


Рис. 3.16. Иллюстрация вычисления признаков формы QRS-комплекса

соответствующего комплекса условно принимается за уровень нулевой линии.

Сопоставление форм двух QRS-комплексов между собой выполняется с помощью критерия близости D_{lij} , рассчитываемого следующим образом:

$$D_{lij} = w_T C_T \cdot |T_i - T_j| + w_A C_A \cdot |A_i - A_j| + w_S C_S \cdot |S_i - S_j| + w_P C_P \cdot |P_i - P_j|,$$

где признаки формы с индексами i и j относятся к соответствующим комплексам; величины C_T , C_A , C_S и C_P — это нормировочные коэффициенты, необходимые для приведения всех признаков формы к безразмерным единицам и для уравнивания их масштабов; w_T , w_A , w_S и w_P — весовые коэффициенты, служащие для учета значимости соответствующих параметров формы для классификации. Чем ближе между собой значения соответствующих признаков формы двух комплексов, тем меньшей будет величина показателя D_{lij} .

На рис. 3.17 показан пример вычисления признаков формы и показателя близости форм D_{lij} для QRS-комплексов одинаковых (δ и ε) и различных (θ) форм, взятых из одной и той же реализации ЭКГ (a). На графике ЭКГ под каждым QRS-комплексом через дефис указаны порядковый номер и условный класс формы. Здесь все QRS-комплексы, кроме комплексов с номерами 7 и 9, относятся к классу «1» (в данном случае он объединяет комплексы фонового ритма, относящиеся к типу «норма»), а комплексы с номерами 7 и 9 — к классу «2»

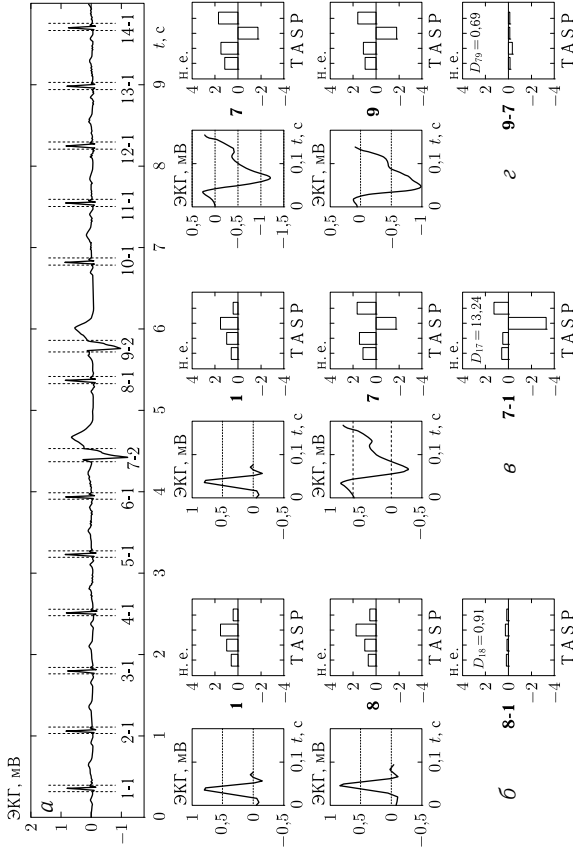


Рис. 3.17. Пример вычисления признаков формы и показателя близости форм D_{ij} для QRS-комплексов одинаковых (б и в) и различных (г) форм, взятых из одной и той же реализации ЭКГ (а)

(в данном случае это желудочковые экстрасистолы, относящиеся к типу «патология»). На рисунках б, в и г слева показаны сравниваемые комплексы, а справа — графически представлены вычисленные для них признаки формы (два верхних графика в правом столбце) и разности, рассчитанные для соответствующих признаков (нижний график в правом столбце). Значения всех признаков выражены в нормированных единицах (н.е.), что соответствует операции нормирования их значений среднеквадратичными отклонениями этих признаков. Приведенная иллюстрация показывает, что для сходных по форме комплексов разности между соответствующими признаками (и, следовательно, значения D_{ij}) относительно невелики по сравнению с разностями для комплексов различной формы.

В случае использования корреляционного метода, для двух QRS-комплексов вычисляется взаимная корреляционная функция (ВКФ) [3.3]:

$$C_{ij}(m) = \begin{cases} \frac{1}{N\sigma_i\sigma_j} \sum_{n=0}^{N-m-1} x_i(n+m)x_j(n), & 0 \leq m \leq M, \\ \frac{1}{N\sigma_i\sigma_j} \sum_{n=0}^{N-|m|-1} x_i(n)x_j(n+m), & -M \leq m < 0, \end{cases}$$

где m — значение сдвига, M — максимальный сдвиг, $x_i(n)$ и $x_j(n)$ — отсчеты двух сравниваемых комплексов (i -го и j -го), N — число отсчетов в каждом из образцов сигнала, σ_i и σ_j — соответствующие стандартные отклонения.

Максимальный сдвиг M выбирается с учетом возможной ошибки в определении условной опорной точки QRS-комплекса, которая может приводить к неточному совмещению сопоставляемых сигналов. Показатель близости форм двух QRS-комплексов (i -го и j -го) рассчитывается как максимум из для всех заданных значений сдвига по формуле

$$D_{ij} = \left[1 - \max_{m=-M}^M C_{ij}(m) \right],$$

благодаря чему обеспечивается то условие, что чем лучше коррелированы два сигнала, тем ближе к нулю будет значение показателя D_{ij} , а случаю наихудшей корреляции ($C_{ij}(m) = -1$) будет соответствовать значение $D_{ij} = 2$.

На рис. 3.18 приведены примеры вычисления ВКФ и показателей близости форм D_{ij} для QRS-комплексов одинаковых и различных классов форм, взятых из одной и той же реализации ЭКГ. Расчет ВКФ выполнялся для фрагментов сигнала длительностью 200 мс (50 отсчетов). Точками на графиках ВКФ отмечены значения функции

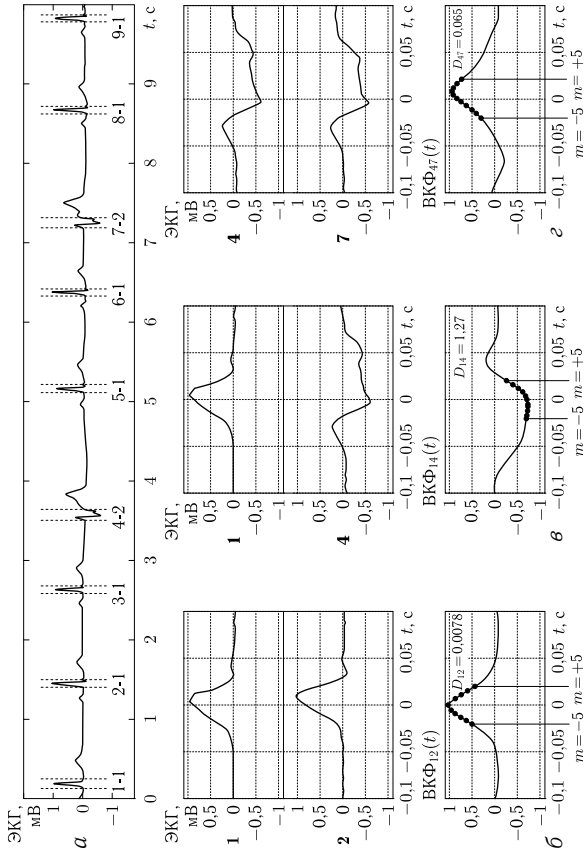


Рис. 3.18. Пример вычисления ВКФ и показателя близости форм D_{ij} для QRS-комплексов одинаковых (б и в) и различных (г) форм, взятых из одной и той же реализации ЭКГ (а)

в диапазоне величины сдвига m от $-M$ до $+M$ ($|M|=5$), которые использовались для определения соответствующих расстояний $D_{\Pi ij}$.

В соответствии с подходом, основанным на спектральных параметрах, вычисляется спектральное разложение QRS-комплекса (дискретное преобразование Фурье)

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{N} nk}, \quad k = 0, \dots, N-1,$$

где $x(k)$ — отсчеты сигнала, а N — размер фрагмента сигнала, по которому рассчитывается ДПФ. Далее вычисляется амплитудный спектр

$$B(k) = \frac{2}{N} \sqrt{[\operatorname{Re} X(k)]^2 + [\operatorname{Im} X(k)]^2}.$$

Компоненты амплитудного спектра могут рассматриваться в качестве признаков формы QRS-комплекса. Так как большая часть энергии ЭКГ в пределах QRS-комплекса сосредоточена в полосе частот от 4 до 50 Гц [3.4], для расчета показателя близости форм используются только спектральные компоненты, относящиеся к этому диапазону частот. Показатель близости форм QRS-комплексов (i -го и j -го) рассчитывается по формуле [3.3]

$$D_{\Pi ij} = \frac{1}{\sqrt{\sigma_i \sigma_j}} \sum_{k=1}^K |B_i(k) - B_j(k)|,$$

где $B_i(k)$ и $B_j(k)$ — спектральные компоненты с номером k для соответствующих комплексов; K — номер элемента разложения для частотного поддиагона, включающего частоту 50 Гц; σ_i и σ_j — соответствующие стандартные отклонения.

На рис. 3.19 приведены примеры вычисления амплитудных спектров и показателя близости форм $D_{\Pi ij}$ для QRS-комплексов одинаковых и различных классов форм, взятых из одной и той же реализации ЭКГ. Расчет амплитудных спектров выполнялся для фрагментов сигнала длительностью 256 мс (64 отсчета).

Каждый из трех рассмотренных методов обладает определенными преимуществами и недостатками по сравнению с двумя другими. Поэтому более высокое качество классификации может быть достигнуто в случае их совместного применения. В качестве интегрального признака, объединяющего показатели D_{1ij} , $D_{\Pi ij}$ и $D_{\Pi ij}$, может быть предложено евклидово расстояние от начала координат в трехмерном пространстве нормированных значений данных показателей [3.3]:

$$D_{ij} = \sqrt{D_{1ij}^2 + D_{\Pi ij}^2 + D_{\Pi ij}^2},$$

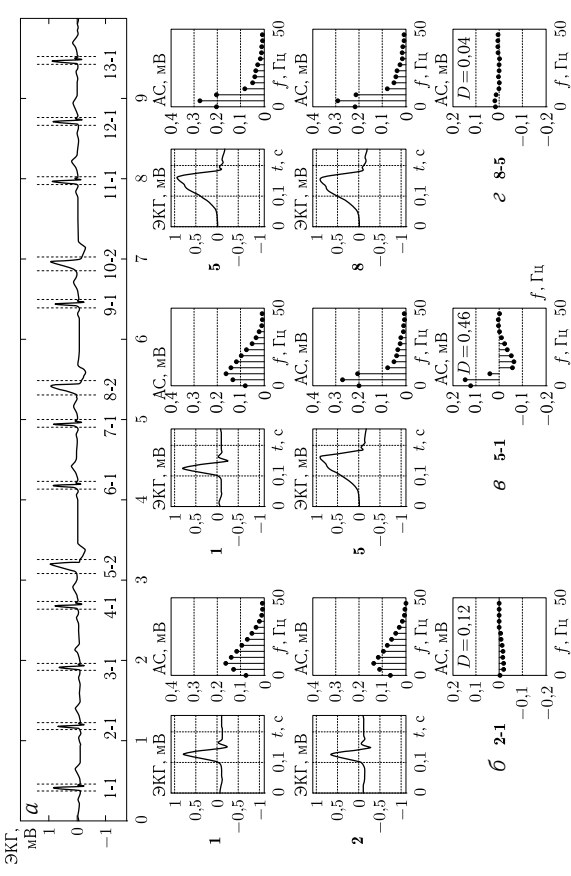


Рис. 3.19. Пример вычисления амплитудных спектров (АС) и показателя близости форм $D_{\Pi ij}$ для QRS-комплексов одинаковых (б и г) и различных (в) форм, взятых из одной и той же реализации ЭКГ (а)

где D_{1ij} , $D_{\Pi ij}$ и $D_{\Pi ij}$ — значения соответствующих показателей, рассчитанных для некоторых двух комплексов с номерами i и j , нормированные среднеквадратичными значениями с целью выравнивания масштабов показателей. Исходя из способа расчета данных показателей, можно предположить, что величина D_{ij} будет тем меньше, чем более сходны между собой сравниваемые комплексы, и наоборот.

На рис. 3.20 показано распределение значений показателя D_{ij} в пространстве признаков D_{1ij} , $D_{\Pi ij}$ и $D_{\Pi ij}$. Точками отмечены значения, полученные для QRS-комплексов, относящихся к одному классу формы, а крестиками — к разным классам. Распределение было получено с использованием верифицированной выборки реальных записей ЭКГ, включающей 50 фрагментов сигнала продолжительностью по 60 с. Из рисунка видно, что точки, соответствующие параметрам комплексов одного класса формы, концентрируются вблизи начала координат, а остальные распределяются на значительном удалении от него. В связи с этим может быть предложено решающее правило в виде порога, определяемого как поверхность сферы с центром в начале координат (см. рис. 3.20). При этом минимальная ошибка классификации достигается для значения радиуса сферы, равного 4,6 [3.3].

Конечной целью анализа сердечного ритма в системах кардиологического наблюдения является распознавание аритмий. Распознавание аритмий основывается на информации о временных интервалах между

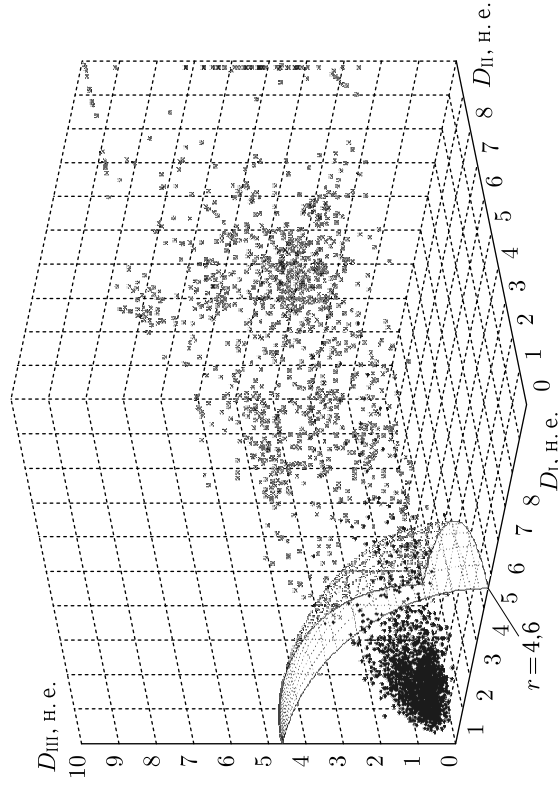


Рис. 3.20. Распределение значений показателя D_{ij} в пространстве признаков $D_{III,ij}$, $D_{II,ij}$ и $D_{I,ij}$. Точками отмечены значения, полученные для QRS-комплексов, относящихся к одному классу формы, а крестиками — к разным классам

обнаруженными QRS-комплексами (RR-интервалах) и типах форм комплексов, получаемой на предыдущих стадиях анализа сигнала. Непосредственно распознавание нарушений ритма выполняется при помощи комплекса логических правил, описывающих аритмические события в терминах соотношений длительностей RR-интервалов и идентификации типов морфологии QRS-комплексов [3.4].

3.7. Анализ ишемических изменений ЭКГ

Одна из наиболее важных задач автоматического анализа ЭКГ состоит в выявлении так называемых *ишемических изменений* кардиоцикла ЭКГ, которые отражают недостаточность кровоснабжения сердечной мышцы [3.4, 3.10]. Ишемические изменения проявляются на ЭКГ в виде трансформации ST-сегмента кардиоцикла (промежуток между окончанием волны S и началом волны T). В норме ST-сегмент выглядит как горизонтальный участок, расположенный приблизительно на уровне изоэлектрической линии (рис. 3.21, а). Ишемические изменения могут носить характер как смещения ST-сегмента вверх или вниз относительно изоэлектрической линии ЭКГ (значимым обычно считается смещение более чем на $\pm 0,1$ мВ), так и изменения формы

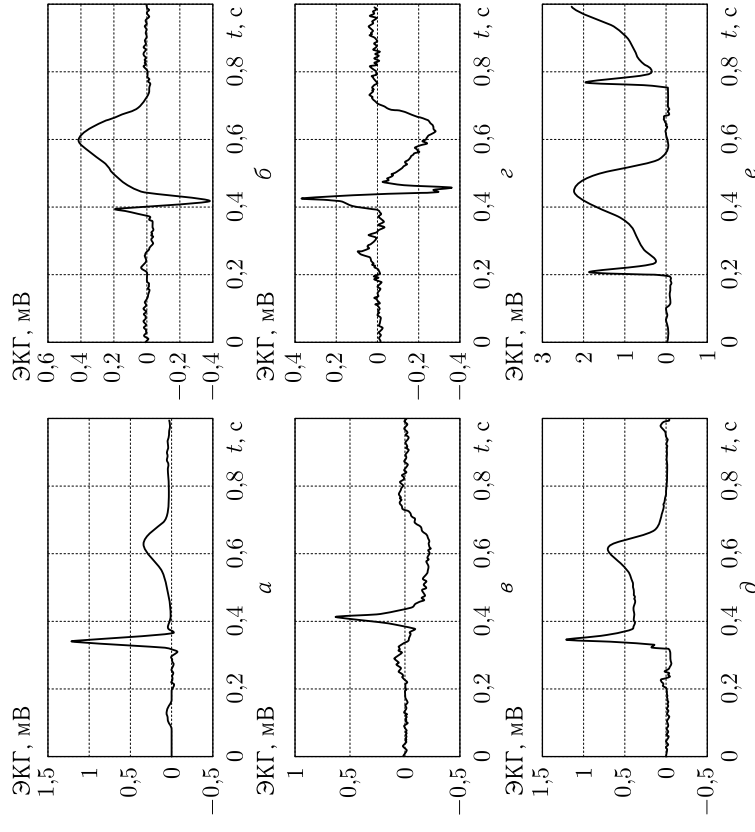


Рис. 3.21. Примеры различных форм ST-сегмента ЭКГ

ST-сегмента (различают горизонтальную, косонисходящую, косовосходящую, вогнутую и выпуклую формы). Примеры кардиоциклов с различными ишемическими изменениями ST-сегмента показаны на рис. 3.21, б–е.

К задачам алгоритмов анализа ST-сегмента относятся оценка величины смещения ST-сегмента относительно изоэлектрической линии ЭКГ и анализ формы ST-сегмента. При этом анализируются только кардиоциклы, относящиеся к фоновому ритму. Решение задачи анализа ST-сегмента ЭКГ затрудняется следующими факторами:

- наличием низкочастотной помехи в виде дрейфа изоэлектрической линии ЭКГ;
- необходимостью предварительного анализа формы QRS-комплекса с целью выбора для анализа ST-сегмента только комплексов фонов ритма;
- сложностью выявления кратковременных переходящих изменений ST-сегмента;
- сложностью точного определения границ QRS-комплекса и ST-сегмента при наличии в сигнале миографических помех.

Процедура анализа ST-сегмента, как правило, включает следующие основные этапы:

- выбор из обнаруженных QRS-комплексов только комплексов, относящихся к фоновому ритму;
- устранение низкочастотной помехи;
- измерение смещения ST-сегмента;
- анализ формы ST-сегмента.

Задача отбора только комплексов фонового ритма решается на этапе классификации форм QRS-комплексов (см. разд. 3.6).

Устранение низкочастотной помехи обычно осуществляется с помощью математического аппарата сплайн-интерполяции [3.4], так как использование фильтрации верхних частот вносит недопустимо большие искажения в форму самого ST-сегмента. При этом для каждого обнаруженного QRS-комплекса определяется условная точка на участке между волной P и началом комплекса (на PQ-сегменте), значение сигнала в которой рассматривается как локальная оценка текущего уровня изоэлектрической линии. Найденные таким образом точки используются в качестве узлов сплайн-интерполяции, в результате которой формируется кусочно-непрерывная функция, состоящая из фрагментов полиномов третьей степени и служащая оценкой низкочастотной помехи в сигнале. Вычитая эту функцию из исходного сигнала, получают ЭКГ, очищенную от дрейфа изоэлектрической линии. Описанный процесс иллюстрируется на рис. 3.22. На верхнем графике показан фрагмент ЭКГ, на котором наблюдается низкочастотная помеха. Круглыми

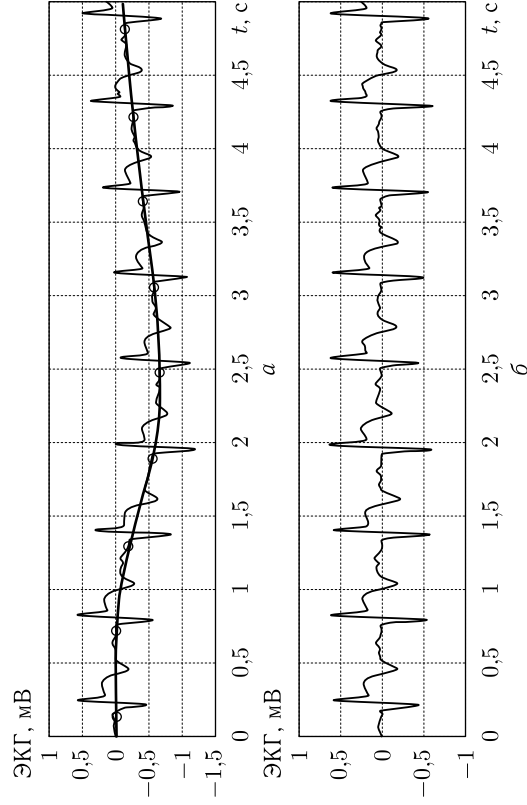


Рис. 3.22. Иллюстрация процедуры устранения дрейфа изоэлектрической линии с использованием сплайн-интерполяции

отмечены найденные узловые точки на PQ-сегменте, а жирной линией показан рассчитанный сплайн. На нижнем графике приведен тот же фрагмент ЭКГ после вычитания сплайна из исходного сигнала. Видно, что дрейф изоэлектрической линии практически полностью устранен.

После устранения низкочастотной помехи для каждого анализируемого комплекса определяется так называемая точка J, за которую принимается точка окончания QRS-комплекса. Смещение ST-сегмента относительно изоэлектрической линии измеряется на определенном расстоянии от точки J (наиболее часто — через 60–80 мс после нее). На рис. 3.23 иллюстрируется процедура вычисления смещения ST-сегмента.

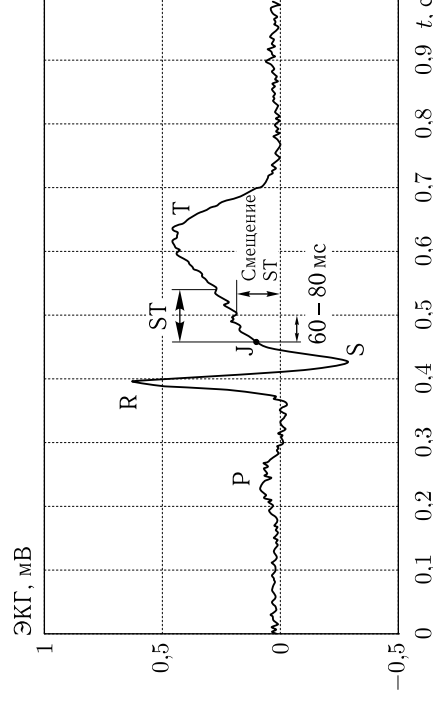


Рис. 3.23. Процедура вычисления смещения ST-сегмента ЭКГ

Анализ характера формы ST-сегмента [3.4] осуществляется с использованием логических правил по положению трех последовательных точек на ST-сегменте (например, по точке J, точке измерения значения смещения и точке начала волны T).

Глава 4

АНАЛИЗ СИГНАЛОВ ВАРИАБЕЛЬНОСТИ СЕРДЕЧНОГО РИТМА, ЭЛЕКТРОЭНЦЕФАЛОГРАММ И СПИРОГРАММ В МЕДИЦИНСКИХ СИСТЕМАХ

4.1. Анализ вариабельности сердечного ритма и артериального давления

Анализ вариабельности сердечного ритма

Анализ ВСР представляет собой метод исследования статистических и частотных свойств сигнала, образуемого последовательно интервалов времени между смежными сокращениями сердца, относящимися к фоновому ритму. Анализ ВСР позволяет получить количественные индикаторы активности различных отделов вегетативной нервной системы человека, что оказывается полезным для решения многих медицинских диагностических задач [4.6, 4.8, 4.9, 4.15, 4.20, 4.27].

Интервалы времени между последовательными сокращениями сердца (RR-интервалы) обычно измеряются по ЭКГ как расстояния между вершинами R-зубцов QRS-комплексов. Если за период регистрации ЭКГ в сигнале наблюдались нарушения сердечного ритма или помехи, вызвавшие ошибки в определении RR-интервалов, то участки сигнала, соответствующие этим событиям, должны исключаться из процедуры анализа. В связи с этим при анализе ВСР в качестве входных данных используется не непосредственно последовательность RR-интервалов, а так называемая последовательность NN-интервалов, т. е. интервалов между смежными QRS-комплексами фонового ритма.

Для анализа ВСР наиболее широко применяются следующие частотные методы. При этом в качестве основных спектральных показателей используются суммарные мощности сигнала в трех диапазонах частот [4.27]: VLF (very low frequency, от 0,003 до 0,04 Гц), LF (low frequency, от 0,04 до 0,15 Гц) и HF (high frequency, от 0,15 до 0,4 Гц), а также ряд показателей, производных от них.

Так как анализируемая последовательность интервалов времени не представляет собой равномерно дискретизованный сигнал, традиционно используемый при частотном анализе, то к ней нельзя

непосредственно применить общепринятые методы цифровой обработки. В связи с этим на первом этапе анализа из последовательности NN-интервалов получают равномерно дискретизованный сигнал (рекомендованная частота дискретизации составляет 4 Гц [4.27]), который рассматривается как описание восстановленной функции управления сердечным ритмом. Именно исследование этой функции, порождающей последовательность регистрируемых RR-интервалов, является предметом анализа ВСР. Для получения равномерно дискретизованного сигнала наиболее часто используется сплайн-интерполяция.

Последовательность NN-интервалов (или RR-интервалов) принято графически отображать в виде ритмограммы (тахограммы), на которой значения интервалов изображаются в виде вертикальных линий, а по оси абсцисс откладываются порядковые номера соответствующих интервалов. На рис. 4.1 приведены примеры ритмограмм для нескольких различных видов сердечного ритма:

- высокая активность в диапазоне LF (рис. 4.1, а);
- высокая активность в диапазонах LF и HF (рис. 4.1, б);
- умеренная активность в диапазонах LF и HF (рис. 4.1, в);
- высокая активность в диапазонах VLF, LF и HF (рис. 4.1, г).

Для получения оценок спектральных показателей вариабельности сердечного ритма используют как непараметрические (классические) методы, основанные на дискретном преобразовании Фурье, так и параметрические методы, в основе которых лежит модель авторегрессии — скользящего среднего (см. гл. 2).

Процедура получения оценок спектральных показателей вариабельности сердечного ритма включает следующие основные этапы:

- 1) определение имеющихся в последовательности RR-интервалов нарушений непрерывности, связанных с помехами или аритмиями;
- 2) получение последовательности NN-интервалов;
- 3) восстановление функции управления сердечным ритмом в виде последовательности равноотстоящих отсчетов;
- 4) расчет оценки СПМ;
- 5) расчет показателей ВСР.

Если в анализируемом фрагменте сигнала встречаются интервалы, связанные с нарушениями ритма, или интервалы, неверно определенные из-за помех, необходимо предварительно скорректировать сигнал. В противном случае в результирующих спектральных оценках могут появиться завышенные значения высокочастотных составляющих, так как искажения, вызванные указанными причинами, обычно проявляются в виде резких всплесков или впадин в последовательности NN-интервалов. В случае относительно коротких промежутков времени между оставшимися NN-интервалами дефектные значения не используются в качестве узлов интерполяции, но при этом сохраняется непрерывность сигнала во времени. Если же искажается достаточно продолжительный отрезок сигнала, то либо к нему добавляются

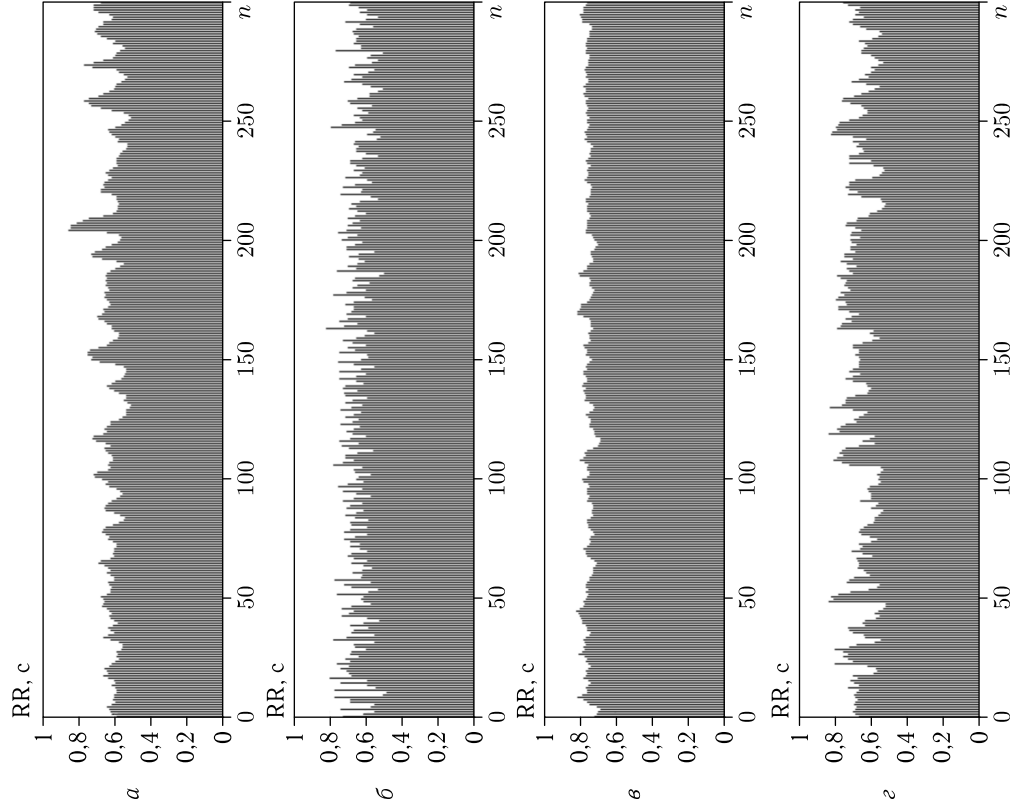


Рис. 4.1. Примеры ритмограмм при различных видах сердечного ритма

промежуточные узлы интерполяции, либо приходится отказываться от анализа данной реализации сигнала.

Равномерно дискретизованный сигнал чаще всего получают с помощью математического аппарата сплайн-интерполяции. На рис. 4.2 иллюстрируется процедура перехода от последовательности NN-интервалов к равномерно дискретизованному сигналу. Верхний график (а) представляет собой фрагмент ритмограммы NN-интервалов, на среднем графике (б) показан процесс получения сплайна, проходящего через узловые точки, соответствующие значениям NN-интервалов, а нижний

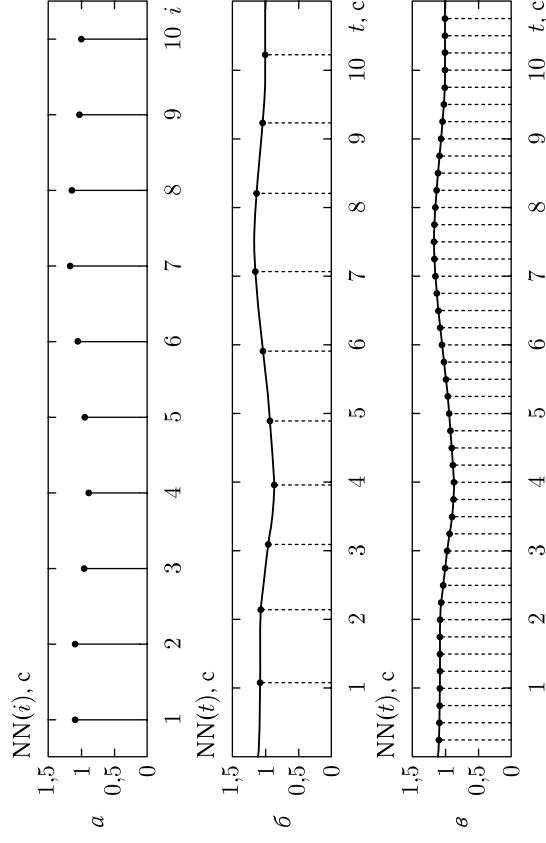


Рис. 4.2. Иллюстрация перехода от последовательности NN-интервалов (а) к равномерно дискретизованному сигналу с частотой 4 Гц (б) с использованием сплайн-интерполяции (б, в)

график (в) демонстрирует расчет по функции сплайна значений равномерно дискретизованного сигнала с частотой отсчетов 4 Гц (или с периодом 0,25 с).

На рис. 4.3, а, б показаны соответственно ритмограмма для фрагмента ЭКГ продолжительностью 5 мин и рассчитанный по ней равномерно дискретизованный сигнал.

На рис. 4.4 показаны полученные для реализации сигнала сердечного ритма, приведенной на рис. 4.3, оценки СПМ и значения рассчитанных по ним частотных параметров ВСП для двух методов спектрального анализа: метода Уэлча (а) и авторегрессионного (б).

Расчет показателей VLF, LF и HF осуществляется интегрированием СПМ с использованием формулы прямоугольников (см. 2.2) по соответствующим диапазонам частот:

$$\text{VLF} = \Delta f \sum_k p(f_k), \quad 0,003 \leq f_k \leq 0,04 \text{ Гц},$$

$$\text{LF} = \Delta f \sum_k p(f_k), \quad 0,04 \leq f_k \leq 0,15 \text{ Гц},$$

$$\text{HF} = \Delta f \sum_k p(f_k), \quad 0,15 \leq f_k \leq 0,4 \text{ Гц},$$

где $\Delta f = f_n/N$ ($f_n = 4$ Гц — частота дискретизации, N — число входных значений ДПФ), а $p(f_k)$ — значение СПМ с индексом k .

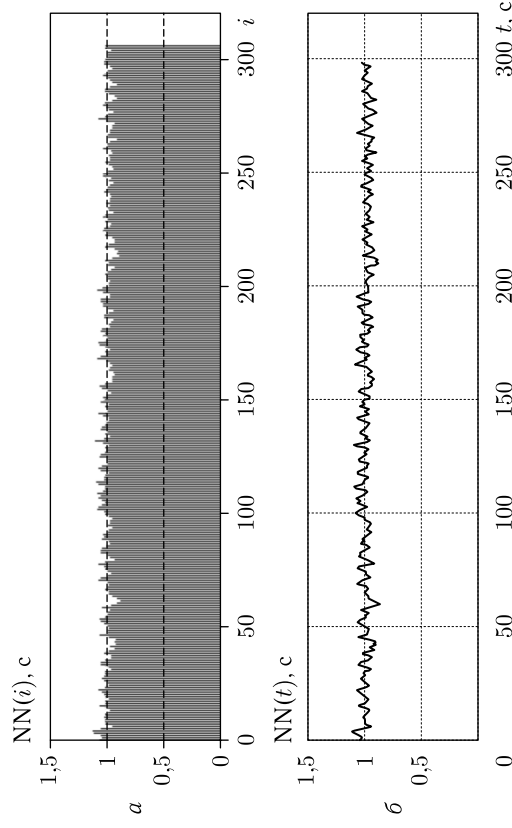


Рис. 4.3. Ритмограмма фрагмента ЭКГ длительностью 5 минут (а) и интерполирующая кривая, полученная с использованием сплайнов (б)

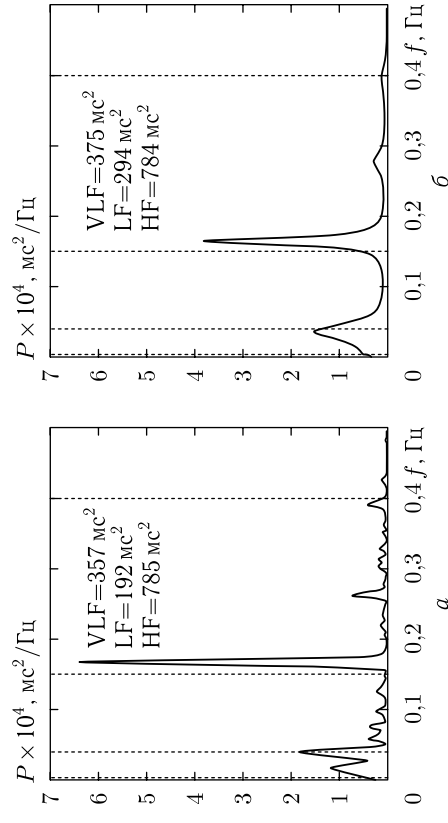


Рис. 4.4. Графики оценки СПМ и значения рассчитанных по ним частотных параметров ВСР для двух методов спектрального анализа: метода Уэлча (а) и авторегрессионного (б)

Заметно, что как вид графиков СПМ, так и значения рассчитанных разными методами параметров несколько различаются между собой, что объясняется случайным характером спектральных оценок, получаемых при использовании любого из методов. Необходимо также обратить внимание на единицы измерения спектральных показателей: СПМ измеряется в $mc^2/\Gamma c$, а мощность — в mc^2 . Это связано с тем,

что в качестве исходных данных для расчета спектров использовались длительности интервалов времени, измеряемые в миллисекундах.

Совместный анализ вариабельности сердечного ритма и артериального давления

С появлением технической возможности одновременной регистрации сигналов сердечного ритма и мгновенного кровяного давления широкое распространение получили математические методы, позволяющие выявить и количественно оценить взаимосвязи между этими двумя сигналами с использованием, в частности, взаимной корреляционной функции и взаимной спектральной плотности мощности сигналов [4.1, 4.9, 4.27, 4.33, 4.40].

Методы расчета взаимной спектральной плотности мощности были рассмотрены в разд. 2.3. На рис. 4.5 показаны фрагменты сигналов сердечного ритма (RR) и артериального давления (АД), которые были получены с использованием сплайн-интерполяции из синхронно записанных последовательностей значений RR-интервалов ЭКГ и мгновенных значений артериального давления. В сигналах хорошо просматриваются сходные по своему характеру высокочастотные колебания.

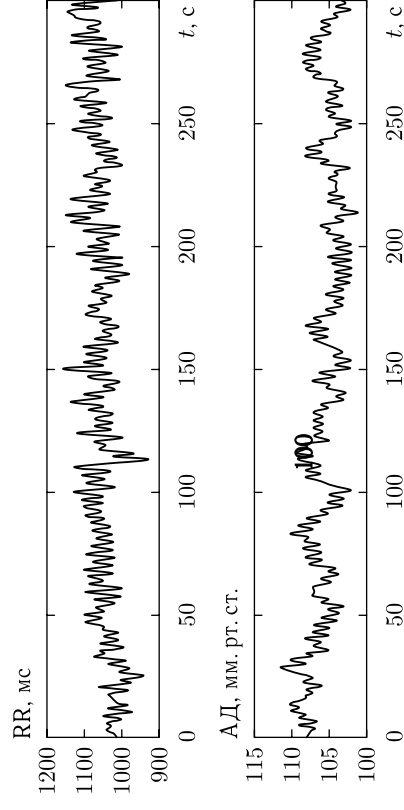


Рис. 4.5. Фрагменты двух синхронно записанных сигналов сердечного ритма (а) и артериального давления (б)

На рис. 4.6 приведены примеры графиков оценок СПМ сигналов сердечного ритма и артериального давления, показанных на рис 4.5, а также график ВСРМ этих сигналов. Значения ВСРМ выражены в $mc \cdot mm \text{ рт.ст.}/\Gamma c$. На графике отчетливо просматриваются несколько пиков, свидетельствующих о существовании выраженных взаимосвязей между колебаниями сигналов в определенных частотных диапазонах. Видно, что ВСРМ имеет относительно высокие значения только на тех частотах, на которых в спектрах обоих сигналов присутствуют хорошо выраженные составляющие.

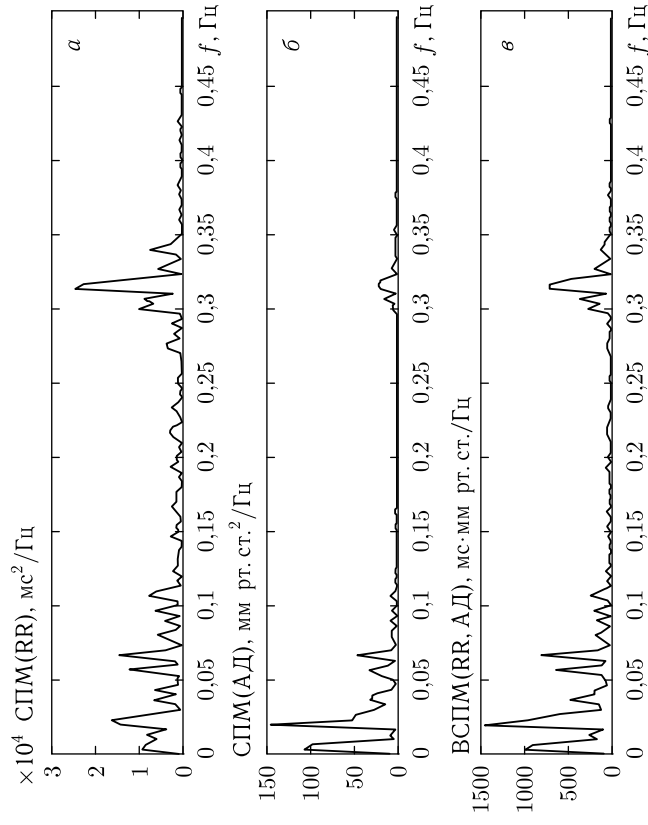


Рис. 4.6. Примеры графиков оценок СПМ сигналов сердечного ритма (а) и артериального давления (б), а также график ВСПМ этих сигналов (в)

На рис. 4.7 приведены примеры графиков оценок фазового спектра сигналов сердечного ритма и артериального давления, показанных на рис. 4.5, а также график взаимного фазового спектра (ВФС) этих сигналов. Значения взаимного фазового спектра выражены в градусах. Из графиков видно, что ВФС принимает устойчивые значения на тех самых участках частотного диапазона, где наблюдаются относительно высокие значения ВСПМ. В частности, это наиболее хорошо заметно в диапазоне частот от 0,29 до 0,36 Гц, соответствующем дыхательной активности человека. Например, на частоте 0,3 Гц фазовый сдвиг составляет примерно 120 град. Так как в данном случае в качестве первого из сигналов использовался сигнал сердечного ритма, а в качестве второго — сигнал мгновенного артериального давления, то положительное значение фазы свидетельствует о запаздывании изменений ритма относительно изменений давления, что соответствует существующим предположениям о физиологических механизмах регуляции сердечного ритма [4.14]. Количественно это запаздывание можно оценить следующим образом (см. раздел 2.3):

$$\Delta T(f_k) = \frac{\arg P_{xy}^o(f_k)}{360^\circ f_k} = \frac{120^\circ}{360^\circ \cdot 0,3} \approx 1,1 \text{ с.}$$

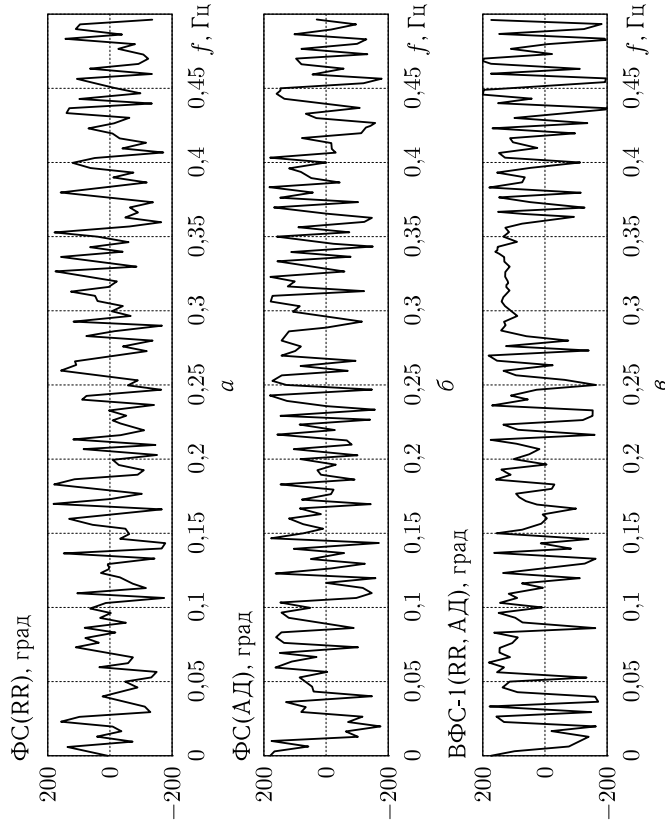


Рис. 4.7. Примеры графиков оценок фазового спектра сигналов сердечного ритма (а) и артериального давления (б), а также график взаимного фазового спектра этих сигналов (в)

То есть на частоте дыхания колебания сердечного ритма здесь отстают от дыхательных волн примерно на 1 с.

4.2. Анализ ВСР на основе модели управления водителем сердечного ритма

Методы автоматического анализа ВСР находят применение при исследовании активности различных физиологических систем, управляющих работой сердца. Основным методом оценки влияния таких физиологических факторов, как артериальное давление, терморегуляция и дыхание, является спектральный анализ ВСР. В стандартных методах принято измерять мощность сигнала в частотных диапазонах VLF, LF, HF и проводить текущий анализ ее изменения во времени. По результатам такого анализа можно получить физиологическую интерпретацию уровней регуляции сердечного ритма пациента, оперативно распознавая вид и степень тяжести его состояния, и даже прогнозировать степень риска внезапной смерти кардиологических больных.

Один из важнейших этапов обработки ритмограммы для задач спектрального анализа — непрерывное (или равномерное) представление ВСР. Дискретность сигнала сердечного ритма вызывает неопределенность поведения функции между отсчетами. Устранить подобную неопределенность позволяет достаточно большое число методов, к которым относятся интерполяция кусочно-полиномиальными функциями (нулевого, первого и третьего порядков), представление сигнала в виде счетных событий, интервальной последовательности и другие традиционные методы [4.26, 4.27]. Их особенность состоит в том, что в основном они восстанавливают функцию, представляющую собой огибающую зарегистрированной последовательности кардиointервалов. Существует и другой подход, который основан на использовании модели управления водителем сердечного ритма. В этом случае за исходную функцию, по которой оценивается частотный состав ВСР, принимается функция, управляющая частотой сердечных сокращений (ЧСС) и, как следствие, содержащая в себе наиболее полную информацию об уровнях регуляции вегетативной нервной системы (ВНС).

Благодаря изучению взаимосвязи функции автономной управляющей активности и последовательности сердечных событий была создана математическая модель, называемая интегральным частотно-импульсным модулятором (IPFM — integral pulse frequency modulator) [4.29]. Эта модель учитывает физиологическую природу и биологические механизмы функционирования водителя сердечного ритма, что имеет большое значение для распознавания и интерпретации спектральных компонент функции управления ритмом сердца.

Электрофизиологические процессы зарождения импульсов возбуждения в синоатриальном (СА) узле и проведения их в систему миокарда находятся под влиянием ряда регуляторных факторов. Нормальное возбуждение сердца начинается в СА-узле. Когда мембранный потенциал волокон узла достигает порогового значения, генерируется потенциал действия, который активирует другие узловые волокна и способствует распространению возбуждения в сердце. Существует по меньшей мере три фактора, определяющих частоту разрядов водителя сердечного ритма. К ним относятся:

- 1) крутизна диастолической деполаризации;
- 2) уровень потенциала покоя;
- 3) уровень порогового потенциала.

Их влияние на работу водителя ритма иллюстрируется на рис. 4.8. При этом решающее значение имеет изменение порогового потенциала и потенциальной деполаризации, а изменение порогового потенциала и потенциала покоя оказывает на ЧСС незначительное влияние, которым при разработке модели можно пренебречь. Разные отделы ВНС оказывают противоположное действие на изменение ЧСС. Так, стимуляция симпатических волокон, которые иннервируют сердце, увеличивает частоту разрядов водителя ритма путем увеличения крутизны диастолической деполаризации, в то время как стимуляция парасимпатических волокон

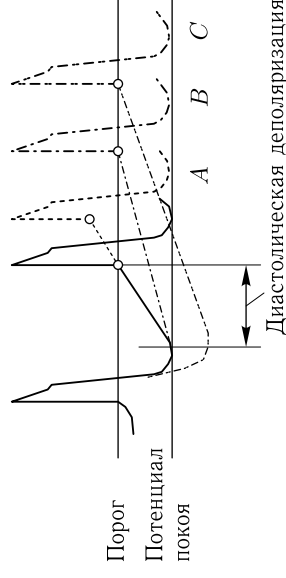


Рис. 4.8. Изменение частоты появления мембранного потенциала водителя сердечного ритма под влиянием изменения порогового значения — А, уменьшения крутизны диастолической деполаризации — В, увеличения максимального диастолического потенциала — С

дает обратный эффект. Поэтому именно анализ изменения скорости деполаризации лежит в основе исследования различных влияний ВНС на сократительную работу сердца.

Модель управления водителем ритма

Согласно IPFM-модели, существует функция $m(t)$, которая влияет на электрическую активность сердца. Эта управляющая функция представляет собой результат деятельности обеих ветвей ВНС. Для оценки вегетативного влияния на работу сердца необходимо соотносить последовательность сердечных событий с этим управляющим сигналом $m(t)$.

На рис. 4.9 представлена модель управления частотой водителя сердечного ритма посредством изменения крутизны диастолической деполаризации. Данная модель отражает лишь диастолическую фазу мембранного потенциала, так как значение имеет только время появления потенциала действия.

Как следует из IPFM-модели, диастолическая деполаризация моделируется путем интегрирования входной функции $m(t)$. Как только

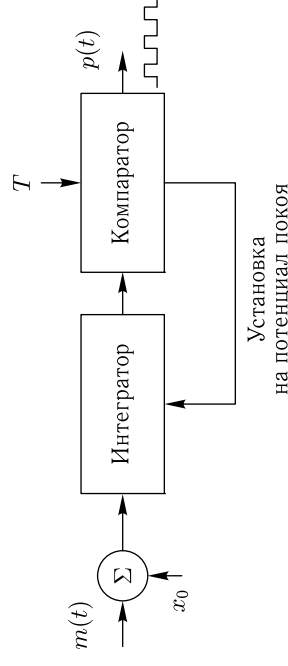


Рис. 4.9. IPFM-модель автономного управления частотой водителя сердечного ритма

значение интеграла достигает фиксированного порога T , соответствующего пороговому значению мембранного потенциала, генерируется импульс, и интегратор устанавливается на значение мембранного потенциала покоя.

Входная функция $m(t)$ и выходная посылка импульсов $p(t)$ математически связаны друг с другом следующим образом. Последовательность $p(t)$ представляется как поток единичных импульсов, расположенных в моменты времени t_i , $i = 0, 1, 2, \dots, k$, в которые наблюдаются желудочковые комплексы электрокардиосигнала:

$$p(t) = \sum_i \delta(t - t_i). \quad (4.1)$$

Моменты времени t_{i-1}, t_i , определяющие два последовательных сердечных события в (4.1), связаны следующим интегральным соотношением:

$$\int_{t_{i-1}}^{t_i} [m(t) + x_0] dt = T, \quad (4.2)$$

где T — постоянная интегрирования; x_0 — постоянная величина, обеспечивающая положительное значение интеграла на временном промежутке (t_{i-1}, t_i) .

Из выражения (4.2) следует, что данная модель позволяет получить математическую зависимость между временем t_i появления сокращения сердца и входной функцией $m(t)$. При этом принимаются следующие допущения и упрощения.

А. Соотношения между автономной активностью симпатического и парасимпатического отделов ВНС и крутизной диастолической деполаризации объединены в непрерывную управляющую функцию $m(t)$, задаваемую на входе модели.

Б. Изменение крутизны диастолической деполаризации является единственным фактором изменения ЧСС.

В. В математическом описании отсутствует в явном виде моделирование рефрактерного периода, хотя для максимального приближения PRFM-модели к своему физиологическому аналогу необходимо его учитывать, поскольку он влияет на формирование длительности кардицикла. Эта задача решается путем косвенного учета рефрактерного периода в параметрах модели: пороговом значении компаратора T и постоянной составляющей x_0 .

Алгоритм восстановления функции управления $m(t)$

Восстановление функции $m(t)$ основано на вычислении интегральной функции $M(t_i)$, определяемой следующим образом. Пусть начальное время $t_0 = 0$, тогда получаем эквивалентное (4.2) выражение

$$M(t_i) = \int_0^{t_i} [m(t) + x_0] dt = i \cdot T, \quad (4.3)$$

где t_i , $i = 1, 2, \dots, N$, — моменты времени, связанные с появлением каждого i -го сокращения сердца.

Функция $M(t_i)$ имеет дискретный вид, поскольку определена лишь в моменты времени t_i . По последовательности $M(t_i)$, $i = 1, 2, \dots, N$, можно восстановить непрерывную функцию $M(t)$, применяя линейную или кубическую интерполяцию, например кусочно-полиномиальными функциями. Последующее дифференцирование $M(t)$ позволяет получить непрерывное описание функции $m(t)$. Производя равномерную дискретизацию восстановленного сигнала, можно решать задачу спектрального анализа сердечного ритма, ожидая объективности оценки спектральной плотности мощности (СПМ) и, следовательно, возможности распознавания функциональных состояний пациента.

На рис. 4.10 в графическом виде последовательно представлены этапы преобразования ритмограммы в непрерывную функцию $m(t)$. Для наглядности все построения проиллюстрированы на модельном примере.

Первый этап обработки связан с анализом интервальной последовательности RR_i , $i = 1, 2, \dots, N$. Ритмограмма (рис. 4.10, а) преобразуется в нерегулярную последовательность длительностей RR -интервалов (рис. 4.10, б), для которой $RR(t_i) = t_i - t_{i-1}$.

Далее формируется двумерный массив, содержащий N точек с координатами $(t_i, M(t_i))$, $i = 1, 2, \dots, N$ (рис. 4.10, в). Значение $M(t_i)$ в моменты времени t_i вычисляется как произведение постоянной величины T на значение целочисленной переменной i .

Следующий шаг (рис. 4.10, в) — построение непрерывной кривой $M(t)$ по опорным точкам $(t_i, M(t_i))$, реализуемое с помощью кубических сплайнов.

Завершающий этап — дифференцирование функции $M(t)$ и нахождение искомой непрерывной функции $m(t)$ (рис. 4.10, з). Далее осуществляется взятие равномерных отсчетов с заданной частотой (4 Гц) и реализуется спектральный анализ полученного дискретного представления функции $m(t)$.

Блок-схема алгоритма восстановления функции $m(t)$ по ритмограмме приведена на рис. 4.11. Здесь используются следующие обозначения: RR_i — длительность i -го RR -интервала, $i = 1, 2, \dots, N$; N — размер массива RR -интервалов в отсчетах; t_0 — начальный момент времени; t_i — момент времени появления i -го сокращения сердца; $M(t_0)$ — начальный отсчет массива $M(t_i)$, $i = 1, 2, \dots, N$.

Высокого качества интерполяции по заданным точкам $M(t_i)$ можно достичь с помощью глобального интерполяционного сплайна, представленного по базису из кубических B -функций [4.19]. Однако организация его вычислений осложняется необходимостью определения большого числа неизвестных, получаемых в процессе решения. Наилучшие результаты с точки зрения вычислительных затрат дает метод интерполяции кубическими сплайнами, в котором определение параметров кубического сплайна сводится к решению системы линейных уравнений,

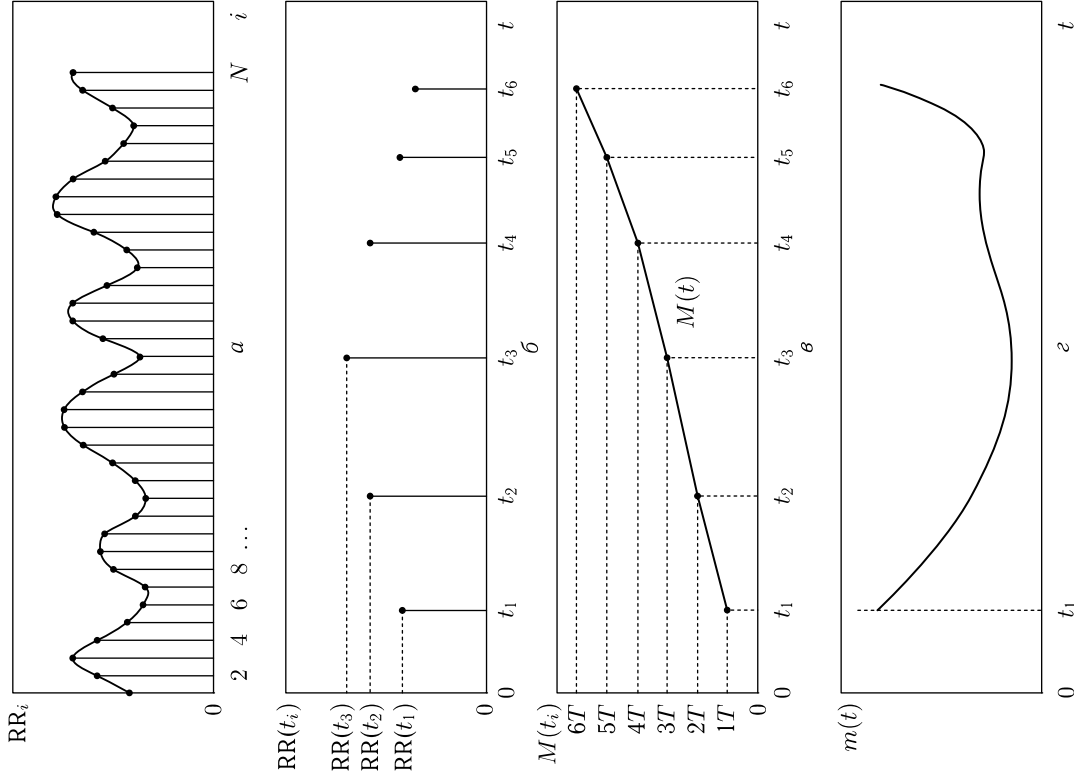


Рис. 4.10. Реконструкция функции управления $m(t)$ по ритмограмме: a — ритмограмма; $б$ — построение нерегулярной последовательности $RR(t_i)$, где $i = 1, 2, \dots$; $в$ — формирование массива $(t_i, M(t_i))$, $i = 1, 2, \dots$, и построение непрерывной функции $M(t)$; $г$ — воспроизведение функции $m(t)$

представленной достаточно простой трехдиагональной матрицей [4.10]. В этом случае возможно организовать вычисление функции управления $m(t)$ в ходе непрерывного наблюдения за динамикой сердечного ритма.

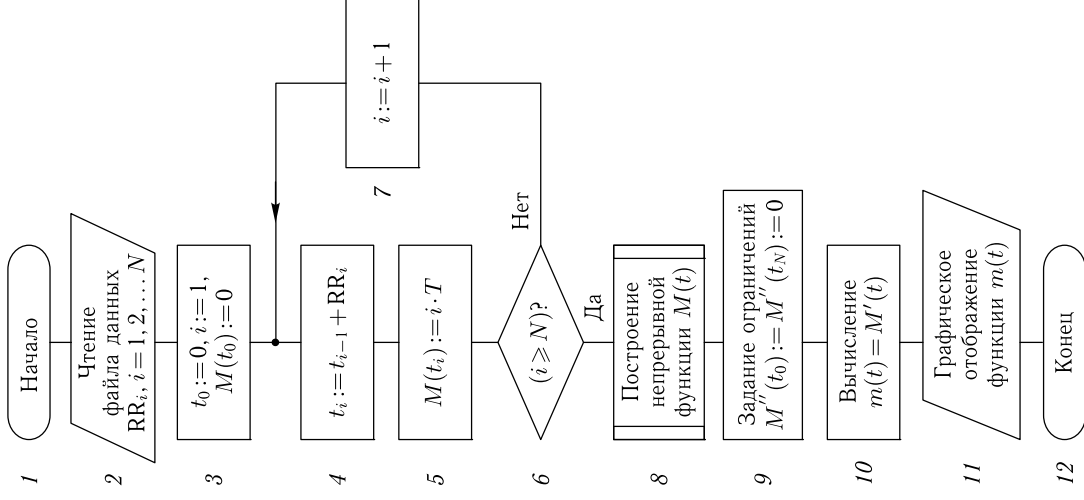


Рис. 4.11. Блок-схема алгоритма воспроизведения функции управления водителем сердечного ритма

Результаты экспериментов

Проведен сравнительный анализ различных методов преобразования сигнала сердечного ритма к равномерной временной шкале [4.14, 4.34, 4.35]. Исследования включали ряд экспериментов, выполненных на модельных и реальных сигналах.

Эксперименты на модельных сигналах проводились с целью оценки эффективности отображения спектрального состава сигналов с заранее заданными частотными свойствами [4.14, 4.34]. Модельные сигналы $m_1(t)$, $m_2(t)$ и $m_3(t)$, представленные с частотой дискретизации 250 Гц, включали одну, две и три гармонические компоненты с частотами $f_1 = 0,10$ Гц, $f_2 = 0,03$ Гц и $f_3 = 0,25$ Гц, соответствующими диапазонам LF, VLF и HF. С помощью алгоритма IPRFM формировались последовательности единичных импульсов. Затем использовался конкретный способ преобразования ритмограммы, осуществлялся переход к равномерно дискретизованному сигналу и вычислялась оценка СПМ.

В качестве показателей эффективности были выбраны:

- интенсивность утечки S_L (определяется как относительное значение суммарной мощности спектральных компонентов, находящихся за пределами изолирующих окон, центрированных относительно частот заданных гармонических функций);
- число ложных спектральных компонент n_α с амплитудами более чем α (1, 5 и 10 % от полной мощности сигнала).

Используя эти показатели, необходимо было определить метод непрерывного представления последовательности импульсов, наиболее точно воспроизводящий частотный состав исходного сигнала.

Сравнительный анализ рассмотренного метода и традиционно используемых в практических разработках способов преобразования сигнала сердечного ритма показал [4.14, 4.34], что метод спектрального анализа, основанный на реконструкции функции управления, обеспечивает более высокую эффективность распознавания частотных составляющих $m(t)$ в HF-, LF-, VLF-областях. Значения S_L при использовании данного метода составили 0,76, 0,69 и 2,46 % для сигналов $m_1(t)$, $m_2(t)$ и $m_3(t)$ соответственно, а показатель $n_\alpha = 6, 4$ и 3 для сигнала $m_3(t)$. Близкие результаты показал метод, основанный на кубической интерполяции неравномерной последовательности кардиочиклов, но значения S_L оказались выше и составили соответственно 0,92, 1,15 и 4,63 %.

В ходе экспериментов, выполненных на реальных записях ЭКГ, установлено [4.35], что переход от спектрального анализа функций, описывающих ритмограмму, к анализу функции управления ЧСС улучшает условия распознавания состояний, связанных с разным уровнем вегетативной регуляции. В качестве характеристики баланса симпатических, соответствующих LF-диапазону, и парасимпатических, соответствующих HF-диапазону, влияний использовался показатель $k = P_{LF}/P_{HF}$, где P_{LF} и P_{HF} — значения суммарной мощности в диапазоне низких LF и высоких HF частот. Анализ более чем 60 реализаций ЭКГ длительностью 5 мин, полученных в клиниках Санкт-Петербурга и представляющих два класса функциональных состояний сердечно-сосудистой системы, включал оценку показателя k двумя способами: с использованием традиционного метода построения

интервальной последовательности (равномерной последовательности) и на основе рассмотренного метода воспроизведения функции $m(t)$.

Результаты экспериментов показали, что критерий Фишера J , принимающий максимальное значение в условиях наилучшей классификации двух состояний, увеличивает значение в 2 раза при переходе к анализу функции автономного управления.

4.3. Оценка глубины анестезии по ЭЭГ на основе спектральной энтропии

Контроль глубины анестезии в ходе проведения хирургических операций — важнейшая задача врача-анестезиолога. При этом важно поддерживать необходимую глубину наркоза и не допускать передозировки анестетиков. Результаты многочисленных исследований [4.5, 4.28, 4.41] показывают, что одним из наиболее информативных методов инструментального контроля глубины анестезии служит автоматический анализ электроэнцефалограммы (ЭЭГ) пациента. При переходе пациента из состояния бодрствования в состояние глубокого наркоза в ЭЭГ наблюдаются характерные изменения как в спектральных характеристиках сигнала, так и в степени его хаотичности. Поэтому наиболее часто для автоматического анализа глубины анестезии по ЭЭГ предлагаются следующие два подхода: анализ частотных свойств сигнала [4.28, 4.30, 4.41] и оценка энтропийных показателей [4.22, 4.23, 4.30, 4.41].

Существуют различные способы вычисления энтропии сигнала. Во временной области рассматривают, например, аппроксимированную энтропию [4.22, 4.30] или энтропию Шеннона [4.22, 4.23]. В частотной области может быть рассчитана спектральная энтропия (СЭ) [4.42], важным преимуществом которой является то, что вклад в энтропию составляющих, лежащих в любом заданном диапазоне частот, может быть выделен как самостоятельная величина.

Основой для расчета СЭ служит оценка спектральной плотности мощности сигнала (СПМ). Наиболее часто для получения оценки СПМ сигнала используется дискретное преобразование Фурье, позволяющее получить из последовательности отсчетов равномерно дискретизованного с частотой f_d сигнала $x(n) = x(nT)$, где $T = 1/f_d$ — интервал дискретизации, такое же количество комплексных величин X_k [4.16]:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)nk},$$

где N — число отсчетов в анализируемом фрагменте сигнала, а каждый из элементов преобразования $X(k)$ соответствует частоте $f(k) = kf_d/N$.

Вычисление СЭ основано на энтропии Шеннона [4.23]. Причем эта характеристика применяется не к самому сигналу, а к его спектру [4.42]. СЭ в некотором диапазоне частот $[f_1, f_2]$ для заданного фрагмента сигнала может быть вычислена при помощи следующей последовательности шагов.

Путем возведения в квадрат амплитуды каждого из элементов $X(k)$ преобразования Фурье сигнала $x(n)$ рассчитывается соответствующее значение спектра мощности $P(k)$:

$$P(k) = X(k)X^*(k),$$

где $X^*(k)$ представляет собой комплексно сопряженную величину для элемента разложения Фурье $X(k)$.

Далее спектр мощности нормируется, для чего рассчитывается такая константа нормировки C_n , что сумма нормированного спектра мощности в пределах заданного диапазона частот $[f_1, f_2]$ равняется единице. Тогда значения нормированного спектра мощности $P_n(k)$ рассчитываются как

$$\sum_{f_i=f_1}^{f_2} P_n(f_i) = C_n \sum_{f_i=f_1}^{f_2} P(f_i) = 1.$$

Спектральная энтропия, соответствующая диапазону частот $[f_1, f_2]$, рассчитывается как сумма

$$E[f_1, f_2] = \sum_{f_i=f_1}^{f_2} P_n(f_i) \lg \left(\frac{1}{P_n(f_i)} \right).$$

После этого вычисляется нормированный на диапазоне от 0 (полная регулярность) до 100 (максимальная нерегулярность) показатель спектральной энтропии $E_n[f_1, f_2]$, для чего рассчитанное ранее значение $E[f_1, f_2]$ делится на коэффициент $\lg(N[f_1, f_2])$, где $N[f_1, f_2]$ равно общему числу частотных компонентов в диапазоне $[f_1, f_2]$:

$$E_n[f_1, f_2] = \frac{E[f_1, f_2]}{\lg(N[f_1, f_2])} \cdot 100.$$

На рис. 4.12 приведены примеры сигналов ЭЭГ одного и того же пациента для четырех различных стадий общей анестезии: *a* — бодрствование непосредственно перед началом анестезии (состояние 1); *б* — глубокий наркоз, достигаемый через несколько минут после начала анестезии (состояние 2); *в* — этап хирургической операции через 1–2 ч после начала анестезии, когда состояние наркоза поддерживается за счет периодического введения дополнительных доз анестезирующих препаратов (состояние 3); *г* — момент времени сразу после пробуждения пациента (состояние 4).

На рис. 4.13 приведены графики СПМ (в логарифмическом масштабе), рассчитанной с использованием алгоритма БПФ для тех же

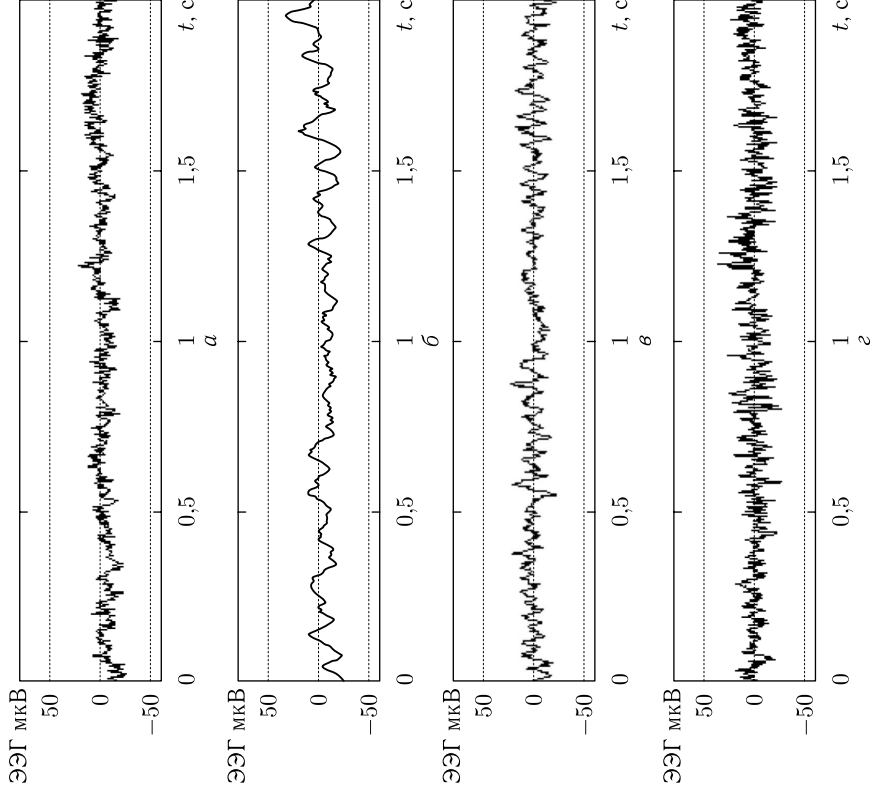


Рис. 4.12. Примеры ЭЭГ одного и того же пациента на различных стадиях хирургической операции: *a* — состояние 1; *б* — состояние 2; *в* — состояние 3; *г* — состояние 4

фрагментов ЭЭГ. Как по самим сигналам ЭЭГ, так и по графикам СПМ хорошо заметно, что в состоянии глубокого наркоза преобладают частоты в диапазоне от 0 до 30 Гц, что соответствует типичному частотному диапазону ЭЭГ. В состоянии бодрствования можно отметить существенное возрастание вклада более высокочастотных составляющих, что объясняется наличием миографической активности. Также на графиках СПМ приведены значения показателя СЭ, рассчитанные для диапазона частот от 1 до 47 Гц. Можно отметить, что в состояниях 1 и 4, соответствующих бодрствованию пациента, показатель СЭ имеет более высокие значения, чем в состояниях 2 и 3, соответствующих различным стадиям анестезии.

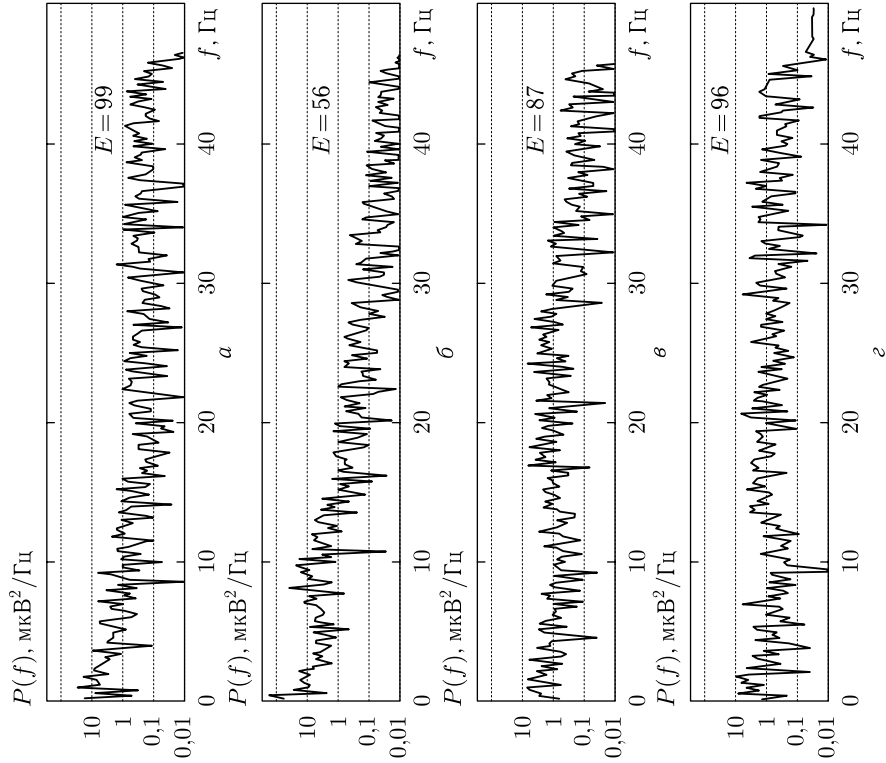


Рис. 4.13. Графики СПМ, рассчитанные по ЭЭГ, приведенным на рис. 1, и соответствующие значения показателя спектральной энтропии на различных стадиях хирургической операции: а — состояние 1; б — состояние 2; в — состояние 3; г — состояние 4

Для повышения информативности показателя спектральной энтропии в областях значений, соответствующих глубокому наркозу и переходу к состоянию бодрствования, необходимо дополнительное преобразование с использованием нелинейной шкалы, эмпирически выведенное выражение для которой имеет вид

$$E_s = \frac{2400}{120 - E} - 20,$$

где E и E_s — соответственно значения спектральной энтропии до и после применения нелинейной шкалы. На рис. 4.14 приведен график данной шкалы.

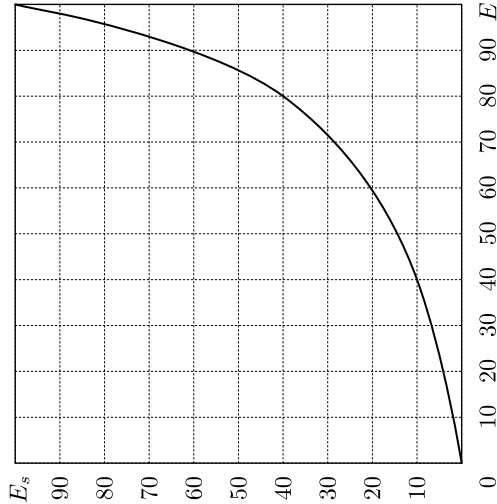


Рис. 4.14. График нелинейной функции преобразования шкалы значений спектральной энтропии

Таблица 4.1

Состояние пациента	E — вычисленное значение СЭ	Значение СЭ E_s , полученное с помощью нелинейной шкалы
1	99	94
2	56	18
3	87	53
4	96	80

В таблице 4.1 приведены вычисленные значения СЭ и значения, полученные в результате применения нелинейной шкалы. Видно, что в последнем случае достигается гораздо более отчетливая дифференциация случайных состояний пациента.

Исследования, проведенные с использованием объемной выборки записей ЭЭГ, зарегистрированных в ходе хирургических операций, продемонстрировали, что показатель глубины анестезии, рассчитанный на основе одной только спектральной энтропии, хотя и позволяет с достаточно высокой точностью различать состояния бодрствования и наркоза, но в то же время не всегда дает возможность надежной дифференциации различных стадий анестезии [4.17, 4.18]. В связи с этим возникает необходимость анализа дополнительных количественных показателей глубины анестезии, которые могут быть рассчитаны по сигналу ЭЭГ. Ниже представлены результаты исследования [4.31, 4.32],

направленного на создание практического алгоритма, использующего комбинацию альтернативных признаков, характеризующих различные проявления в сигнале, связанные с применением общей анестезии.

Для проведения данного исследования был сформирован набор верифицированных записей ЭЭГ, полученных в ходе хирургических операций. Набор содержал 54 записи сигнала ЭЭГ общей продолжительностью более 140 ч. Использовалась частота дискретизации 500 Гц и разрядность 24 бита. Для верификации записи с точки зрения оценки глубины анестезии использовались поминутно регистрировавшиеся показания контрольного прибора (монитор анестезии BIS A-2000XP, «Asprect Medical»), подключенного к тому же пациенту с использованием отдельного комплекта электродов. В качестве показателя использовался выдаваемый прибором на индикацию так называемый BIS-индекс [4.41], принимающий значения в диапазоне от 0 до 100.

Исследование проводилось с использованием специально подготовленной выборки фрагментов сигнала, взятых из описанного выше набора. Выборка включала 23 группы по 4 фрагмента. Длительность каждого фрагмента 60 с. Фрагменты каждой группы выбирались из одной записи ЭЭГ. Четыре фрагмента в группе соответствовали четырем различным состояниям пациента, характерным для применения общей анестезии и описанным в табл. 4.2.

Таблица 4.2

Критерии выбора фрагментов ЭЭГ

Условный номер состояния	Стадия операции	Состояние пациента	Значение BIS-индекса
1	Перед применением анестезии	Бодрствование до применения анестезии	95–100
2	Через несколько минут после применения первой дозы анестетика	Глубокая анестезия	15–25
3	Через 1–2 ч после первого применения анестезии	Стабильное состояние анестезии	55–65
4	Пробуждение после окончания операции	Бодрствование сразу после выхода из состояния анестезии	85–95

На рис. 4.15. Показаны четыре десятисекундных фрагмента, взятых из одной и той же группы записей и соответствующих четырем описанным в табл. 4.2 состояниям пациента.

Переход от одной стадии анестезии к другой сопровождается изменениями в частотном составе ЭЭГ [4.41]. Для использования этого

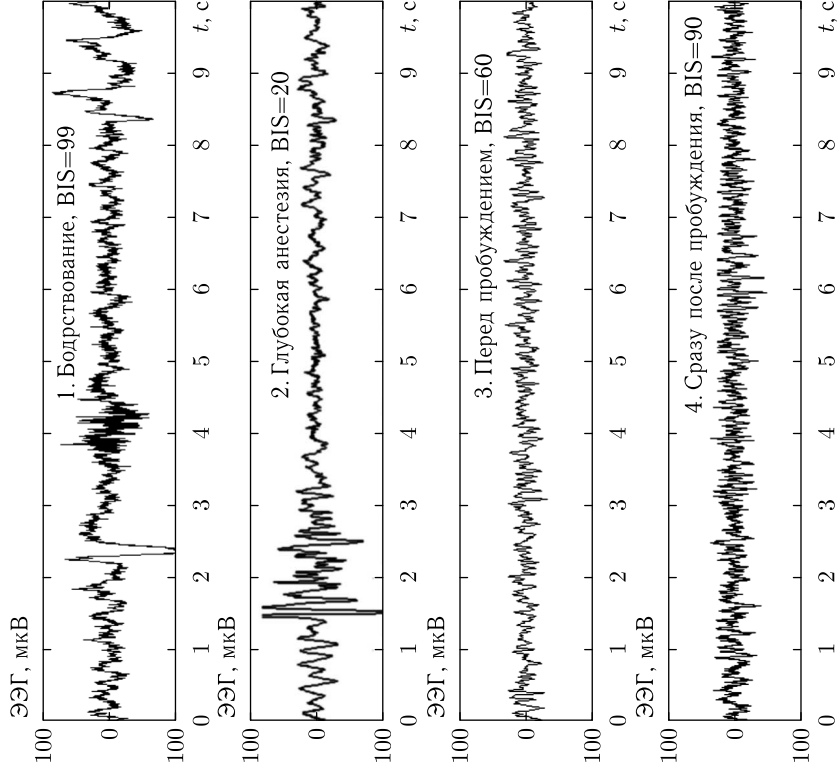


Рис. 4.15. Четыре десятисекундных фрагмента, взятых из одной и той же группы записей и соответствующих четырем различным состояниям пациента

свойства на практике необходимо получить некоторые количественные оценки этих изменений. Для экспериментальных исследований использовались фрагменты ЭЭГ, соответствующие двум следующим стадиям анестезии:

- глубокий наркоз, достигаемый через несколько минут после начала анестезии (состояние 2, BIS = 20);
- этап хирургической операции через 1–2 ч после начала анестезии (состояние 3, BIS = 60).

Рассчитывался показатель

$$R_p = \sqrt{\frac{P_{\text{low}}}{P_{\text{high}}}},$$

где P_{low} — суммарная мощность в диапазоне частот $f_1 \leq f \leq f_2$ и P_{high} — суммарная мощность в верхней части диапазона частот $f_3 \leq f \leq f_4$.

Оценки спектральной плотности мощности (СПМ) рассчитывались с использованием метода Уэлча [4.16] (см. разд. 2.3), суть которого заключается в усреднении оценок СПМ, полученных для некоторого числа перекрывающихся сегментов записи сигнала. Перед использованием преобразования Фурье для каждого сегмента выполнялась операция удаления линейного тренда.

Таким образом, исследовались следующие параметры данной процедуры:

f_1, f_2, f_3, f_4 — пределы выбираемых частотных диапазонов;

T_{fr} — суммарная длительность фрагмента сигнала для вычисления СПМ;

t_w — длительность сегмента сигнала (окна) для расчета БПФ;

ΔT — шаг сдвига окна по времени.

В качестве количественной оценки различия между некоторой исследуемой парой показателей «1» и «2» использовался следующий критерий, известный из теории линейного дискриминантного анализа [4.4]:

$$J_{1,2} = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2},$$

где m_1, m_2 — средние значения рассчитанных показателей и s_1^2, s_2^2 — соответствующие значения дисперсий. Чем выше разделяющая способность анализируемого показателя, тем более высокие значения критерия $J_{1,2}$ должны получаться.

В случае исследования двух состояний анестезии «2» и «3» оптимизация параметров процедуры вычисления показателя R_p выполнялась с использованием описанного выше критерия, обозначаемого здесь $J_{2,3}$. Оптимизация выполнялась с использованием метода наискорейшего спуска [4.25]. В результате был выбран следующий набор параметров: $f_1 = 0$ Гц, $f_2 = 0,8$ Гц, $f_3 = 7$ Гц, $f_4 = 16$ Гц, $T_{\text{fr}} = 30$ с (15 000 отсчетов сигнала при частоте дискретизации $f_d = 500$ Гц), $T_w = 8,192$ с (4096 отсчетов), $\Delta T = 1$ с (500 отсчетов). Максимальная достигнутая величина критерия оптимизации при этом составила $J_{2,3} = 3,46$, что свидетельствует о достаточно хорошей разделяющей способности показателя R_p .

Пример графиков оценок СПМ для двух анализируемых состояний анестезии показан на рис. 4.16. Закрашенные области под кривой СПМ соответствуют значениям суммарной мощности, используемым для вычисления показателя R_p . Как можно видеть, значения этого показателя, рассчитанные для двух разных состояний анестезии, существенно отличаются.

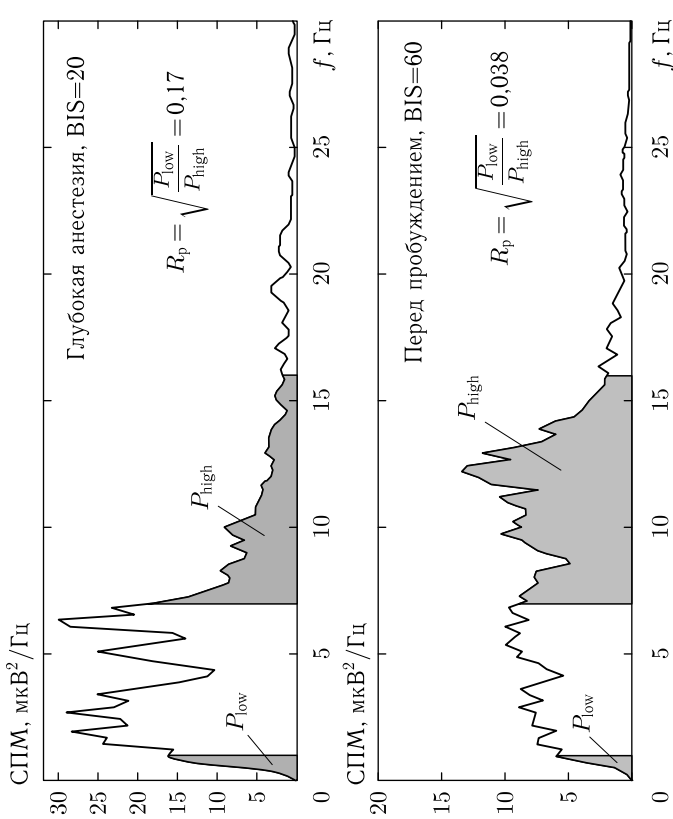


Рис. 4.16. Пример графиков СПМ для двух анализируемых состояний анестезии

В состоянии глубокой анестезии в сигнале ЭЭГ часто наблюдается эффект, получивший название «вспышка-подавление» («burst-suppression») [4.41], который проявляется в чередовании сегментов сигнала с очень низкой амплитудой («подавление») и коротких высокоамплитудных всплесков сигнала («вспышка»). В качестве количественной оценки выраженности данного эффекта используют так называемое «отношение вспышка-подавление» («burst-suppression ratio»), вычисляемое как отношение суммарной продолжительности T_S участков типа «подавление» к аналогичной величине T_B для участков типа «вспышка», подсчитанное за определенный промежуток времени:

$$R_{BS} = \frac{T_S}{T_B}.$$

Сегмент сигнала $x(n)$, расположенный в промежутке между отсчетами, имеющими индексы n_1 и n_2 , считается участком типа «подавление» в случае выполнения условия

$$|x(n)| < X_{\text{lim}}, \quad n_1 \leq n \leq n_2,$$

где $X_{\text{lim}} = k s_x$ — пороговое значение, s_x — стандартное отклонение анализируемого фрагмента сигнала, а k — эмпирически подобранный

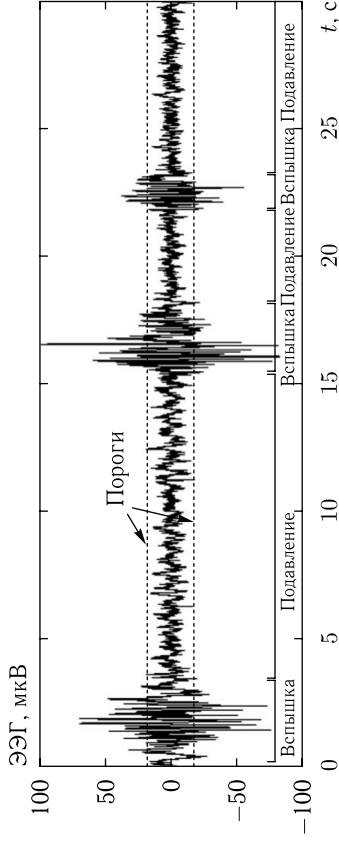


Рис. 4.17. Процесс сегментации фрагмента ЭЭГ

коэффициент. Процесс сегментации фрагмента ЭЭГ на участки типа «подавление» и «вспышка» иллюстрируется на рис. 4.17.

Исследовались следующие параметры процедуры сегментации:

k — коэффициент, используемый для расчета X_{lim} ;

T_{lr} — общая длительность анализируемого фрагмента сигнала;

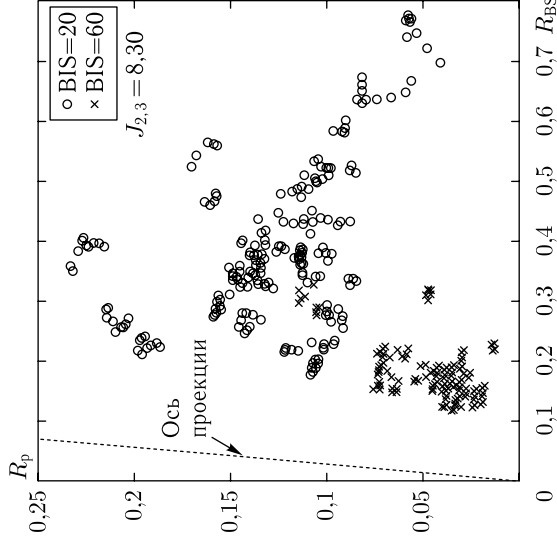
T_w — размер сегмента сигнала (окна) для расчета показателя R_{BS} ;

ΔT — шаг сдвига окна по времени.

В результате оптимизации был выбран следующий набор параметров: $k = 0,8$, $T_{lr} = 40$ с (20 000 отсчетов сигнала при частоте дискретизации $f_d = 500$ Гц), $T_w = 5$ с (2500 отсчетов), $\Delta T = 1$ с (500 отсчетов). Достигнутая максимальная величина критерия $J_{2,3} = 2,12$ свидетельствует о наличии выраженной разделяющей способности отношения «вспышка–подавление» R_{BS} .

Максимальные значения критерия J , достигнутые для оптимальных комбинаций параметров обоих рассмотренных выше методов, достаточно велики, чтобы продемонстрировать наличие некоторой разделяющей способности этих параметров с точки зрения дифференциации различных стадий анестезии, но в то же время не настолько велики, чтобы обеспечить надежность и стабильность этой дифференциации.

На рис. 4.18 представлено совместное распределение значений двух предложенных показателей R_p и R_{BS} для стадий анестезии, соответствующих значениям BIS-индекса 20 и 60. Это распределение, с одной стороны, демонстрирует отсутствие какой-либо корреляции между показателями, а с другой — показывает, что точки на плоскости, соответствующие двум различным состояниям, формируют практически изолированные области на плоскости. С использованием методов линейного дискриминантного анализа [4.25] может быть найдено положение оптимальной оси проекции (пунктирная линия на графике), которая соответствует максимуму используемого критерия различимости состояний $J_{2,3} = 8,3$. Проекция точек распределения на эту ось могут быть найдены с использованием выражения $R_{BS,p} = k_{BS}R_{BS} + k_pR_p$,

Рис. 4.18. Совместное распределение значений показателей R_p и R_{BS} для стадий анестезии, соответствующих значениям BIS-индекса 20 и 60

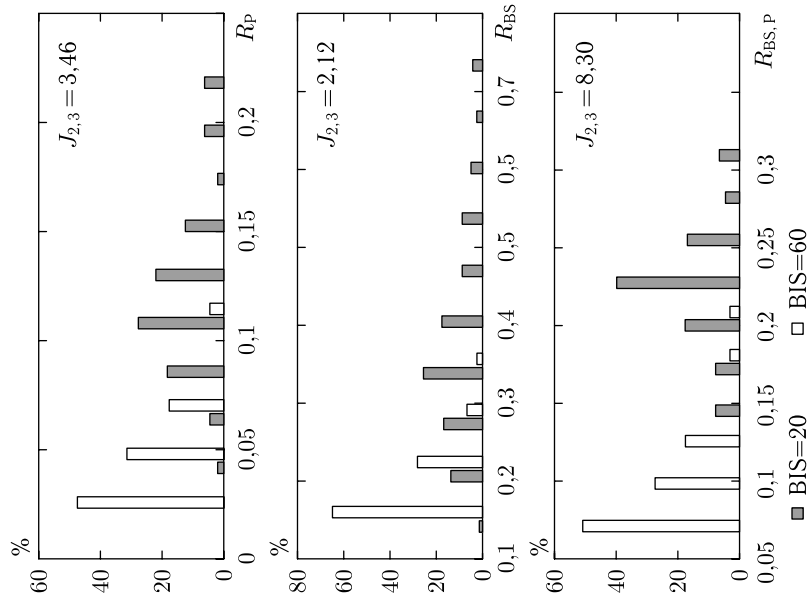
где $k_{BS} = 0,27$ и $k_p = 0,96$ — коэффициенты, определяющие положение оси проекции.

На рис. 4.19 приведены гистограммы распределения значений показателей R_p , R_{BS} и $R_{BS,p}$. Можно отметить, что значение критерия $J_{2,3} = 8,3$ для нижнего из распределений значительно выше, чем значения критериев, рассчитанных для каждого двух других случаев. Это означает, что комбинационное использование параметров R_p и R_{BS} позволяет получить достаточно хороший показатель глубины анестезии.

В результате проведения описанного исследования были получены два основных показателя глубины анестезии: оценка спектральной энтропии $\bar{E}_s$, обеспечивающая надежную дифференциацию между состояниями бодрствования и наркоза, и комбинационный показатель $R_{BS,p}$, демонстрирующий способность к различению стадий анестезии. Была предложена следующая комбинация этих двух показателей:

$$\begin{aligned} E &= \bar{E}_s + C(I_{BS,p}, \bar{E}_s), \\ I_{BS,p} &= (1 - \bar{R}_{BS,p}) \cdot 100, \end{aligned}$$

где $\bar{E}_s$ и $\bar{R}_{BS,p}$ — показатели $\bar{E}_s$ и $R_{BS,p}$, усредненные по пяти последним значениям, а $C(I_{BS,p}, \bar{E}_s)$ — экспериментально определенная поправка (график на рис. 4.20). Параметры предложенных правил получены в результате минимизации по методу наименьших квадратов

Рис. 4.19. Гистограммы значений показателей R_P , R_{BS} и $R_{BS,P}$

расхождений между получаемыми значениями E и соответствующими показаниями контрольного прибора.

На основе представленного здесь исследования был разработан практический алгоритм, реализованный в виде пакета программ на языке программирования C++. Кроме описанных выше методов анализа ЭЭГ, практический вариант алгоритма включает также ряд процедур, повышающих его стабильность и устойчивость к шумам, в частности текущий анализ уровня шумов, предварительная цифровая фильтрация, усреднение и медианная фильтрация рассчитываемых показателей, устранение изолированных всплесков в выходных значениях.

Для тестирования алгоритма был использован весь набор записей ЭЭГ, описанный выше (54 записи общей продолжительностью более 140 ч). Статистические параметры, характеризующие расхождения между рассчитанными значениями показателя E и соответствующими показаниями контрольного прибора (BIS-монитора), представлены

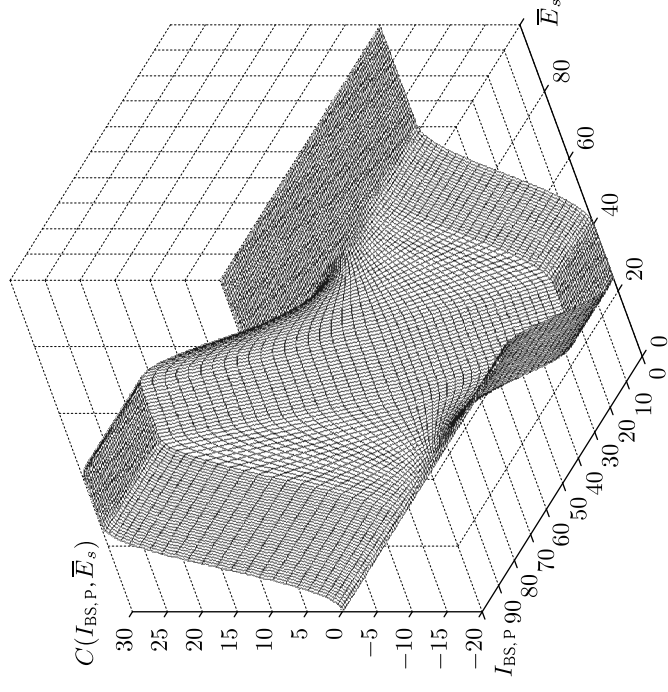
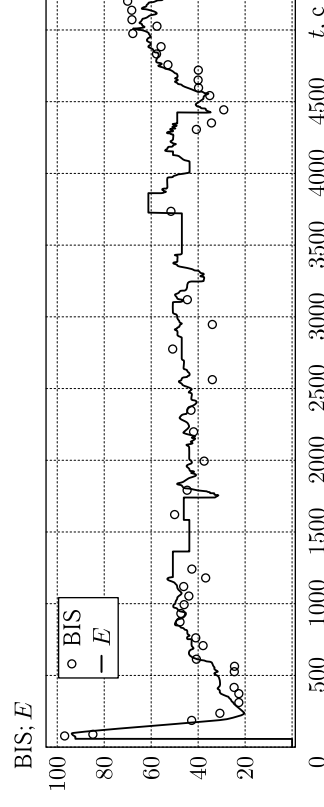
Рис. 4.20. График нелинейной двумерной поправки $C(I_{BS,P}, \overline{E}_s)$, используемой для расчета показателя глубины анестезии E 

Рис. 4.21. Пример сравнения показаний BIS-монитора (кружки) и выходных значений тестируемого алгоритма (сплошная линия)

в табл. 4.3. На рис. 4.21 представлен пример сравнения показаний BIS-монитора (кружки) и выходных значений тестируемого алгоритма (сплошная линия) для одной из записей ЭЭГ тестового набора.

Анализ результатов тестирования алгоритма демонстрирует адекватность выходных показаний глубине анестезии. Большая часть

Таблица 4.3

Результаты тестирования алгоритма

Параметр	Максимум	Минимум	Среднее
Коэффициент корреляции	0,91	0,67	0,78
Стандартное отклонение, мкВ	11,9	6,5	9,7
Максимальная ошибка, мкВ	33,2	15,6	23,1

выявленных ошибок была вызвана спонтанно возникающими скачкообразными изменениями выходных значений из-за высокого уровня помех во входном сигнале. Таким образом, дальнейшие усилия по совершенствованию алгоритма должны быть в основном направлены на повышение его устойчивости к помехам.

4.4. Оценка параметров дыхания при искусственной вентиляции легких

Аппараты искусственной вентиляции легких (ИВЛ) относятся к основным жизнеобеспечивающим приборам, которыми оснащены все отделения реанимации и интенсивной терапии. Главное назначение аппаратов ИВЛ — снабжение легких пациента необходимым для дыхания газовой смесью и вывод из них углекислого газа и других компонентов воздушной смеси. Данная функция аппаратов ИВЛ носит название «протезирование дыхания» и абсолютно необходима для сохранения жизни пациентам в раннем послеоперационном периоде, в коматозном состоянии и других клинических случаях, когда человек не способен дышать самостоятельно [4.36].

Искусственная вентиляция легких — проблема сложная как с технической точки зрения, так и с позиций физиологии. Кроме того, задачи, решаемые аппаратами ИВЛ, и условия их применения весьма разнообразны. Подробно эти вопросы описываются в специальной литературе [4.11, 4.21, 4.36]. Здесь мы рассмотрим только наиболее общие вопросы, связанные с применением методов цифровой обработки сигналов в аппаратуре для ИВЛ.

Важнейшими параметрами, характеризующими процесс искусственной вентиляции легких, являются поток, объем и давление газа [4.11]. Наиболее часто поток (F) принято выражать в литрах в минуту, объем (V) — в миллилитрах, а давление (P) — в сантиметрах водного столба (см вод. ст.), хотя находят применение и альтернативные единицы измерения [4.11, 4.39]. Характер изменения указанных трех параметров в течение дыхательного цикла зависит от установленного режима ИВЛ, состояния органов дыхания пациента, наличия у пациента спонтанной (собственной) дыхательной активности и ряда других факторов. Рассмотрим два наиболее часто применяемых режима ИВЛ:

режим с заданным давлением и режим достижения заданного объема. Для упрощения будем считать, что спонтанная дыхательная активность пациента отсутствует, то есть дыхание происходит только под управлением аппарата ИВЛ.

В режиме с контролируемым давлением основной управляемый параметр — давление. В начальный момент цикла дыхания открывается впускной клапан, через который газ подается в легкие, и начинается наполнение легких газом (фаза вдоха). При этом аппарат контролирует величину давления и не позволяет ей превысить установленное значение. Фаза вдоха завершается по истечении заданного промежутка времени, после чего впускной клапан закрывается и одновременно открывается выпускной клапан, благодаря чему газ начинает выходить из легких за счет упругости грудной клетки (фаза выдоха). Время, отведенное на фазу выдоха, также является устанавливаемым параметром. На рис. 4.22 ($a, б, в$) показаны кривые изменения потока, давления и объема при данном режиме ИВЛ.

В показанном примере были установлены следующие основные параметры режима ИВЛ:

- частота дыхания 5 циклов в минуту (продолжительность цикла 12 с);
- отношение: время вдоха/время выдоха 1:2;
- максимальное давление 10 см вод. ст.;
- положительное давление конца выдоха (ПДКВ) 4 см вод. ст.

Помимо кривых, показывающих изменение потока, давления и объема, широкое применение находят двумерные графики, попарно связывающие мгновенные значения объема и давления (рис. 4.22, $г$) и потока и объема (рис. 4.22, $д$), называемые петлями «объем–давление» и «поток–объем» соответственно. Специалисты могут по характеру этих изображений получить важную и достаточно полную информацию как о ходе ИВЛ, так и о состоянии органов дыхания пациента. В связи с этим во всех промышленно выпускаемых аппаратах ИВЛ предусматривается визуальное отображение данных петель наряду с одномерными (скалярными) графиками каждого из трех параметров в отдельности. Следует отметить, что при построении петель информация о ходе процесса ИВЛ во времени теряется.

В режиме достижения заданного объема управляемыми параметрами являются поток и объем. В начальный момент цикла дыхания открывается впускной клапан и начинается наполнение легких газом (фаза вдоха). При этом аппарат контролирует величину потока и не позволяет ей превысить установленное значение. Фаза вдоха завершается по достижении установленного объема газа, который рассчитывается в текущем режиме путем интегрирования значений потока, после чего впускной клапан закрывается и фаза вдоха завершается. Далее может следовать короткий промежуток времени, в течение которого закрыты оба клапана (так называемая «инспираторная пауза»). После

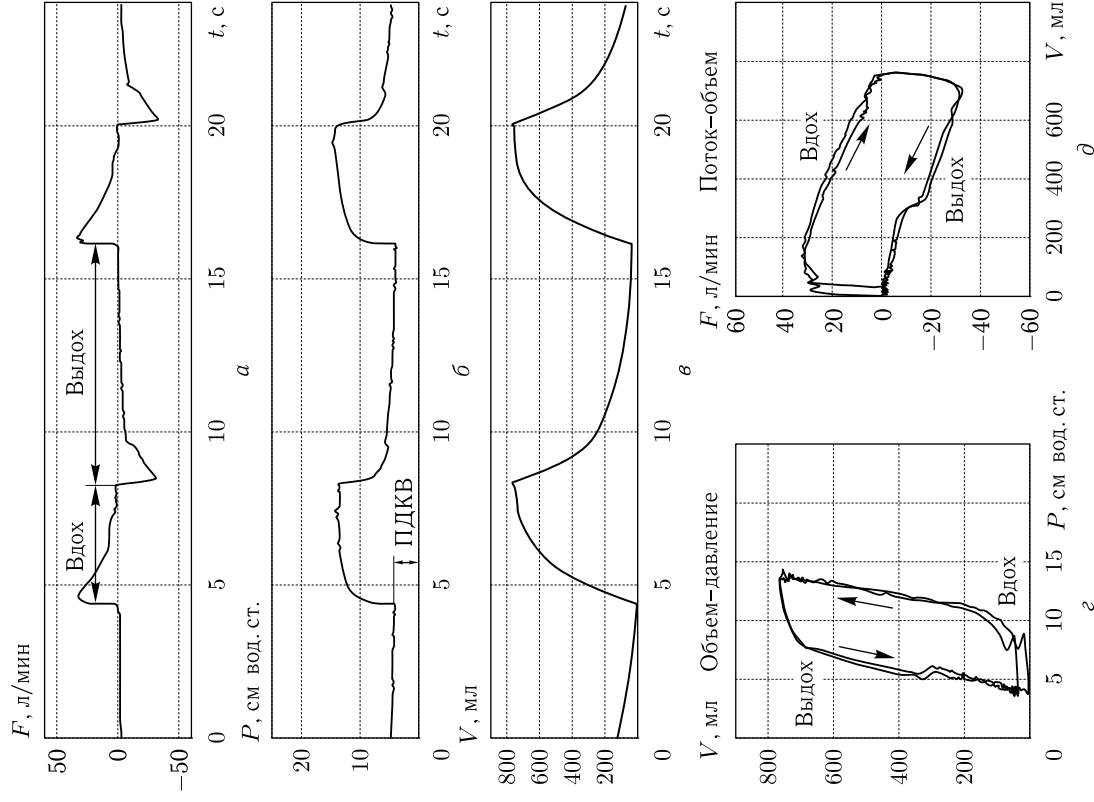


Рис. 4.22. Пример графиков изменения потока (а), давления (б) и объема (в), а также петель «объем-давление» (г) и «поток-объем» (г) в режиме ИВЛ с контролируемым давлением

этого открывается выпускной клапан, и начинается фаза выдоха, которая завершается по прошествии определенного промежутка времени.

На рис. 4.23 (а, б, в) показаны кривые изменения потока, давления и объема при данном режиме ИВЛ, а также петли «объем-

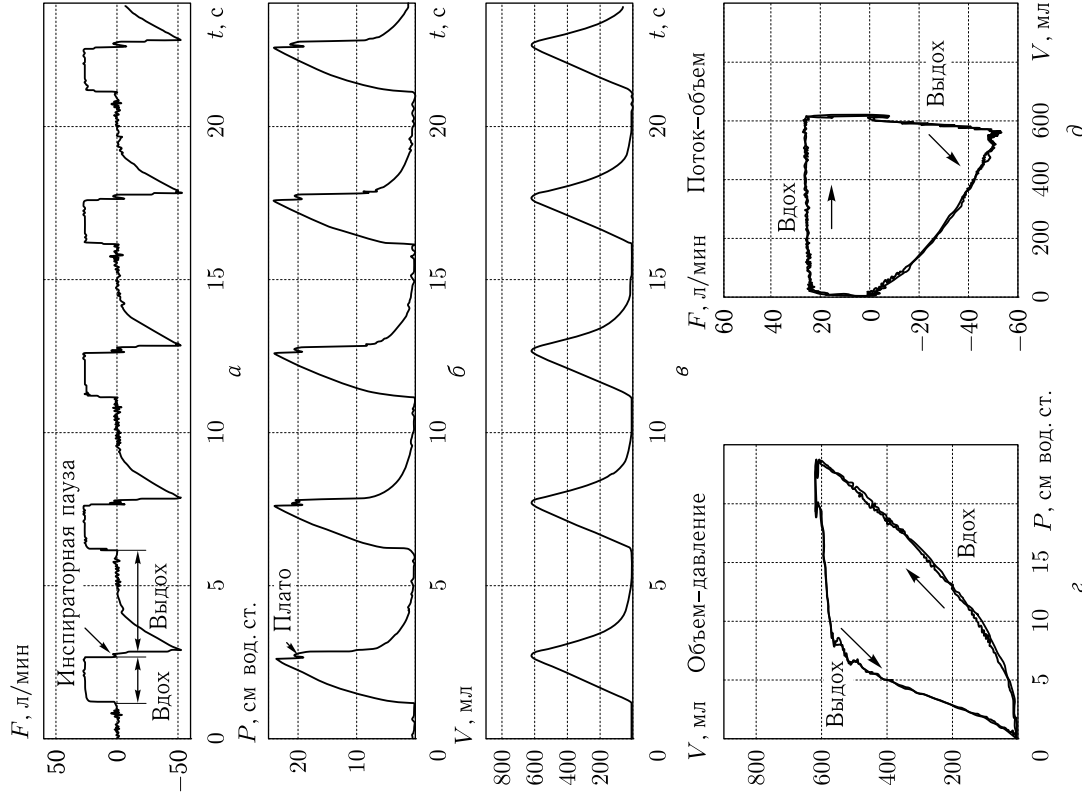


Рис. 4.23. Пример графиков изменения потока (а), давления (б) и объема (в), а также петель «объем-давление» (в) и «поток-объем» (г) в режиме ИВЛ с заданным объемом

давление» (в) и «поток-объем» (г). В данном примере были установлены следующие основные параметры режима ИВЛ:

- частота дыхания 12 циклов в минуту (продолжительность цикла 5 с);
- отношение: время вдоха/время выдоха 1:2;

- максимальный объем 600 мл;
 - положительное давление конца выдоха (ПДКВ) 0 см вод. ст.
- Отметим характерный вид кривой давления в фазе выдоха: в промежутке времени между закрытием впускного клапана и открытием впускного (т.е. во время инспираторной паузы) происходит процесс выравнивания давления в системе аппарат-легкие, сопровождающийся экспоненциальным самопроизвольным снижением давления до некоторого значения, которому соответствует небольшой уступ на кривой давления, называемый «плато» (см. рис. 4.23, б).
- Основными спирометрическими показателями, отражающими состояние дыхательной системы при ИВЛ, являются: растяжимость, или комплаенс, C , а также сопротивление дыхательных путей, или резистанс, R . Растяжимость определяется как способность к изменению объема на единицу изменения давления:

$$C = \frac{\Delta V}{\Delta P},$$

где ΔV и ΔP — соответственно приращения объема и давления газа за некоторый промежуток времени [4.11]. Комплаенс выражается в миллилитрах, деленных на сантиметр водного столба. Резистанс определяется величиной давления, которое необходимо приложить для проведения по дыхательным путям единицы газового объема в единицу времени:

$$R = \frac{\Delta P}{F},$$

где F — величина потока газа. Резистанс выражается в сантиметрах водного столба, умноженных на секунду.

В случае использования режима искусственной вентиляции в режиме достижения заданного объема, расчет резистанса и комплаенса может базироваться на оценке изменения давления в промежутке времени между закрытием клапана вдоха и открытием клапана выдоха. Если обозначить уровень давления, на котором завершился предыдущий выдох P_{expir} , пиковое давление, достигаемое к концу вдоха — P_{peak} , а давление в конечной точке плато (инспираторной паузы) — P_{plato} , то указанные два показателя могут быть рассчитаны следующим образом [4.11]:

$$C = V_T / (P_{\text{plato}} - P_{\text{expir}}),$$

$$R = (P_{\text{peak}} - P_{\text{plato}}) / F.$$

Здесь V_T — максимальный объем в момент завершения выдоха, а F — установленная для данного режима величина потока.

При известных значениях аэродинамического сопротивления дыхательных путей R и растяжимости легких C может быть рассчитана постоянная времени, также являющаяся одним из важнейших показателей, характеризующих процесс ИВЛ:

$$\tau = CR.$$

Исходными данными для алгоритма вычисления параметров, таких как комплаенс, резистанс и постоянная времени, служат синхронно зарегистрированные сигналы, а также отметки времени, соответствующие моментам начала вдоха, окончания вдоха и начала выдоха [4.7]. На первом этапе обработки сигналы потока, давления и объема подвергаются предварительной фильтрации с использованием цифрового фильтра нижних частот. С целью обеспечения устойчивости вычисления параметров вентиляции, методом синхронного накопления выполняется формирование текущей усредненной модели каждого из трех сигналов. Усреднение выполняется по нескольким последним дыхательным циклам. При этом каждый раз оценивается среднеквадратичное отклонение очередного цикла от текущей усредненной модели. В случае если это отклонение превышает заданный порог, соответствующий цикл не используется для усреднения. Это позволяет обеспечить устойчивость процедуры усреднения сигналов к возможным случайным отклонениям и помехам. По полученной усредненной кривой, используя отметки времени, соответствующие моментам переключения клапанов аппарата, определяют необходимые для расчета величины P_{expir} , P_{peak} и P_{plato} , по которым рассчитываются искомые параметры C , R и τ .

В тех режимах вентиляции, когда в сигнале давления отсутствует плато (т.е. длительность инспираторной паузы равняется нулю), описанный выше способ расчета использовать не удается. Кроме того, данный способ позволяет получить только одно значение каждого из этих параметров, соответствующее инспираторной паузе. В то же время известно, что на протяжении всего цикла дыхания эти параметры не остаются постоянными, и информация об их динамике представляет большую диагностическую ценность [4.11, 4.36].

Более универсален подход, основанный на предположении, что общее давление представляет собой сумму давления в дыхательных путях и давления в легких [4.36]:

$$P(t) = P_1(t) + P_2(t) = F(t)R + \frac{V(t)}{C}.$$

Тогда для двух моментов времени t_1 и t_2 имеем систему из двух уравнений

$$\begin{cases} P(t_1) = F(t_1)R + \frac{V(t_1)}{C}, \\ P(t_2) = F(t_2)R + \frac{V(t_2)}{C}. \end{cases}$$

Измерив поток, давление и объем в моменты времени t_1 и t_2 и решив данную систему уравнений, можно вычислить комплаенс C и резистанс R .

Оценка значений комплаенса и резистанса осложняется неустойчивостью сигнала, помехами и наличием спонтанного дыхания. При этом само значение комплаенса меняется на протяжении фазы вдоха. Поэтому его необходимо оценивать по близко расположенным

точкам. Для удаления помех из сигнала можно воспользоваться усреднением, фильтрацией или аппроксимацией полиномом. Однако использование фильтра нижних частот для устранения помех приводит к сглаживанию участков начала и окончания вдоха, то есть к искажению

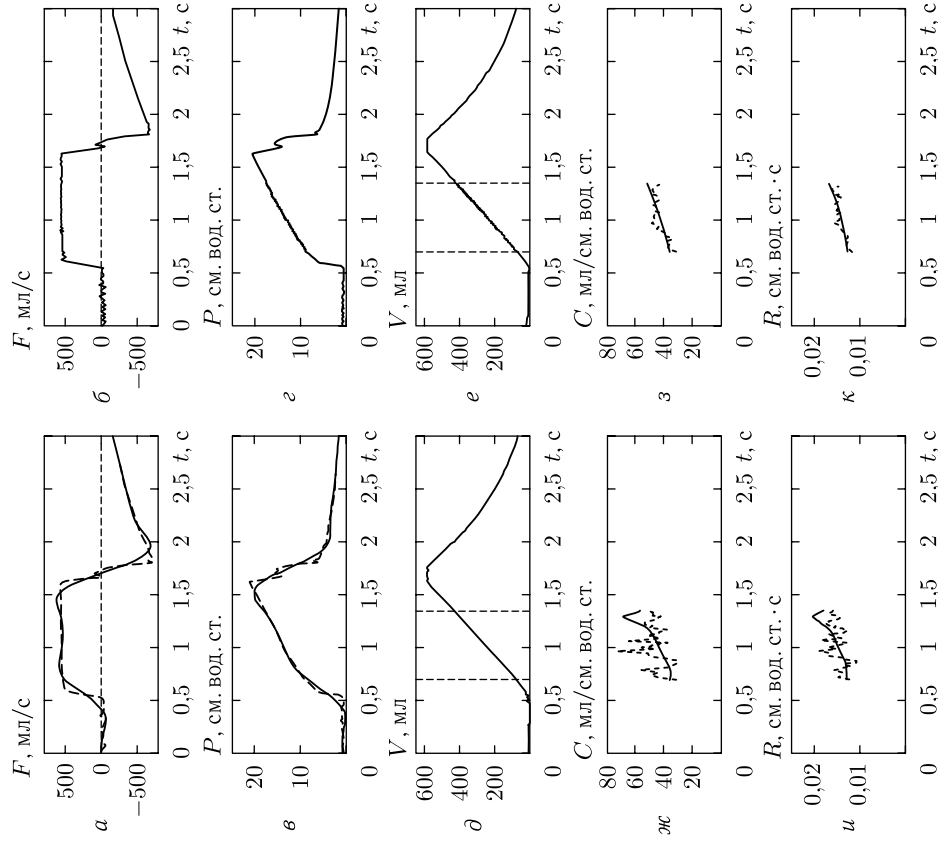


Рис. 4.24. Пример расчета комплайенса и резистанса: а, в, д — сигналы потока, давления и объема, пунктир — исходные сигналы, сплошная линия — сигналы после ФНЧ; б, г, е — сигналы потока, давления и объема после процедуры синхронного усреднения; жирной линией показан фрагмент аппроксимации полиномом степени 2; ж, и — рассчитанные значения комплайенса C и резистанса R , пунктир — по исходным сигналам, сплошная линия — с использованием ФНЧ; з, к — рассчитанные по усредненному сигналу значения комплайенса и резистанса, пунктир — по усредненным сигналам, сплошная линия — с использованием полиномиальной аппроксимации

сигнала, которое может вызвать появление ошибок расчета параметров дыхания. При синхронном усреднении в сигнале остаются те помехи, которые связаны с особенностями аппаратуры и одинаковы для всех циклов вдоха. Оба этих недостатка могут быть устранены при использовании полиномиальной аппроксимации. Однако при этом встает вопрос о выборе оптимального набора конкретных параметров данной процедуры: степени полинома, длительности аппроксимируемого фрагмента, шага перемещения фрагмента. В работе [4.7] показано, что наилучшие результаты удается получить при следующем наборе параметров алгоритма:

- число усредняемых циклов дыхания 5–10;
- длительность аппроксимируемого фрагмента 0,4–0,6 с;
- шаг перемещения фрагмента 0,05–0,1 с;
- дистанция между парой точек для решения системы уравнений 0,06–0,08 с;
- порядок аппроксимирующего полинома 2.

На рис. 4.24 показан пример использования аппроксимации полиномом в сравнении с методами, основанными на фильтрации и усреднении. Как видно из данного рисунка, в случае применения полиномиальной аппроксимации обеспечивается наилучшая стабильность вычисления комплайенса и резистанса.

4.5. Автоматический анализ формы спирографических петель по их сигнатуре

Контроль состояния пациента в режиме искусственной вентиляции легких (ИВЛ) основан на динамической оценке ряда спирографических показателей, характеризующих эффективность газообмена. К ним в первую очередь относятся такие параметры вентиляции, как жизненная емкость легких, сопротивление дыхательных путей, растяжимость легких. Современные аппараты ИВЛ, обеспечивающие функцию «протезирования дыхания», должны также выполнять диагностическую функцию, позволяющую на ранних стадиях обнаруживать развитие состояний, угрожающих жизни пациента. К последним можно отнести, например, отек легких, обструктивные нарушения в бронхолегочной системе больного [4.11, 4.39].

Ранняя диагностика патологий органов дыхания возможна лишь в ходе непрерывного контроля за состоянием пациента. Она основана на автоматическом анализе основных параметров внешнего дыхания, а также на обнаружении существенных отклонений в заданных режимах ИВЛ. Такой анализ осуществляется по спирографическим критериям, включающим как скалярные функции (давление, поток и объем), так и двумерные функции, представленные в виде петель «объем–давление» (ОД) и «поток–объем» (ПО). В исследованиях, посвященных автоматическому анализу спирограмм при ИВЛ, рассматриваются

процедуры измерения параметров вентиляции легких по скалярным кривым, а петли привлекаются только для визуального пояснения наблюдаемых изменений [4.24, 4.37, 4.38]. В то же время установлено, что этот вид спирограмм представляет наибольший практический интерес для задач мониторинга в процессе проведения респираторной поддержки [4.11].

В работах [4.12, 4.13] предложены методы и алгоритмы непрерывного контроля функций внешнего дыхания по спирографическим петлям, в частности, по петлям ОД. Они основаны на динамическом анализе ряда числовых характеристик, которые связаны с основными спирографическими параметрами (compliance and resistance). Применение их в аппаратах ИВЛ позволяет более точно регулировать параметры вентиляции и своевременно диагностировать развитие патологий.

Петли «объем—давление»

Петля ОД является графической формой описания функциональной зависимости дыхательного объема V от давления P в контуре системы дыхания. В ходе регистрации кривых петель образуется на протяжении каждого дыхательного цикла. Она одновременно отражает влияние двух физиологических параметров: растяжимости легких C (compliance) и сопротивления дыхательных путей R (resistance) [4.11, 4.13]. Типичная кривая ОД при ИВЛ с управляемым объемом для одного дыхательного цикла изображена на рис. 4.25.

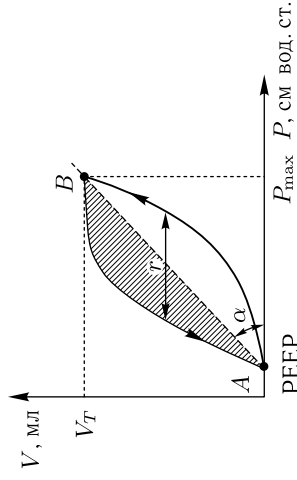


Рис. 4.25. Модельная петля «объем—давление»

Точка A соответствует началу вдоха (конец выдоха), точка B — концу вдоха (началу выдоха), а величина V_T показывает дыхательный объем легких. Ввиду того что существует запаздывание изменения объема V относительно давления P , график имеет вид петли гистерезиса. Нижняя ветвь петли связана с работой дыхания по растяжению эластичных тканей легких во время вдоха, а верхняя — с работой дыхания по преодолению сопротивления дыхательных путей на выдохе. Поскольку давление в конце выдоха должно поддерживаться на заданном уровне, кривая дыхания смещена вдоль оси давления P на положительную величину РЕЕР (positive end-expiratory pressure).

Линия, соединяющая две характерные точки кривой (A , B), задает направление основной оси петли. Угол наклона оси, равный α , характеризует динамическую растяжимость дыхательной системы $C = \tan \alpha$, а ширина петли r — сопротивление дыхательных путей R . При этом расстояние от оси петли до ее восходящей и нисходящей ветвей (на рис. 4.25 указано стрелками) отражают инспираторное R_{in} и экспираторное R_{ex} сопротивления соответственно. Кроме горизонтальной протяженности петли r с изменением величины сопротивления дыхательных путей связан другой интегральный параметр — площадь петли S_p в фазе выдоха (на рис. 4.25 отмечена штриховкой).

Наблюдение за динамикой формы петли, положением кривой на плоскости, углом наклона α и шириной петли r позволяет косвенно судить об изменениях основных спирографических показателей C и R , описывающих работу дыхания пациента [4.11, 4.39].

На рис. 4.26 в качестве примера изображены две спирограммы пациентов, зарегистрированные в условиях клиники, для случаев нормы ОРДС (рис. 4.26, a) и наличия острого респираторного дистресс-синдрома — ОРДС (рис. 4.26, b).

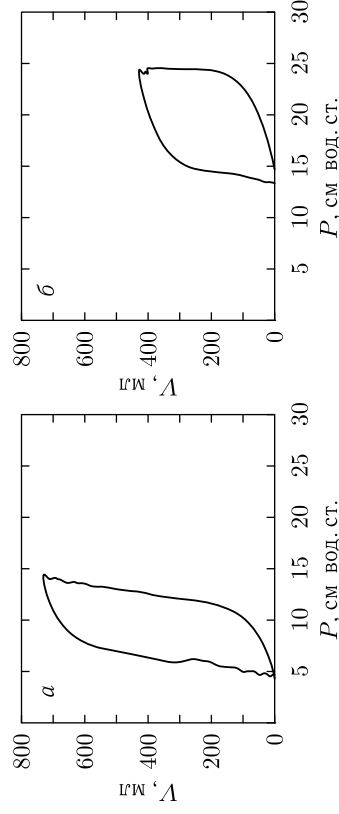


Рис. 4.26. Пример спирограмм пациентов в режиме искусственной вентиляции легких: a — норма; b — острый респираторный дистресс-синдром

Можно заметить, что при патологии изменяются форма, размеры петли, расположение ее на плоскости. На кривой ОД наблюдается отклонение петли к горизонтальной оси (снижается растяжимость легких C) и расширение ее вдоль оси давления (повышается сопротивление дыхательных путей R).

Анализ формы петель по их сигнатуре

Существуют различные способы описания формы петель. Они основаны на использовании цепных кодов, аппроксимации фигур полиномами разных степеней, описании границ набором числовых признаков (площадь, длина, направление главных осей замкнутой фигуры и т. д.), представлении двумерных кривых в виде сигнатур [4.2]. Сигнатуры

имеют то преимущество, что описание двумерных петель сводится к более простым одномерным функциям.

Наиболее простой способ получения сигнатуры петли состоит в построении зависимости расстояния от центра петли (центр — некоторая средняя точка объекта, например его центр тяжести) до границ кривой в виде функции угла θ с применением равномерной дискретизации по углу (во всех экспериментах $\Delta\theta = 1^\circ$). Анализ изменений формы петель по сигнатуре предполагает независимость ее от поворота и размера петли. Первое условие достигается выбором фиксированной начальной точки отсчета сигнатуры, например точки начала дыхательного цикла (точка А, рис. 4.25). Тогда можно сравнивать формы петель по скалярным кривым, поскольку они будут синхронизированы с фазами дыхательного цикла. Инвариантность функции к размеру петли обеспечивается нормировкой сигнатуры по максимальному значению или диапазону изменения ее ординат.

Используя цифровое изображение петли, а также функцию сигнатуры, можно вычислить ряд параметров, характеризующих форму и расположение петли на плоскости, провести сравнительный анализ регистрируемой и опорной петли. В качестве опорных петель необходимо выбрать те, которые отражают установленный врачом режим вентиляции легких. В процессе непрерывного контроля опорные кривые могут периодически обновляться, что связано с необходимостью корректировки параметров вентиляции.

Алгоритм динамического анализа петли «объем—давление» по их сигнатурам

Алгоритм динамического анализа петли ОД и распознавания отклонений в режиме ИВЛ можно представить следующим образом.

Вначале проводится нормировка каждой из координатных осей (V и P), с тем чтобы, во-первых, перейти к безразмерным величинам и, во-вторых, получить расположение петель на плоскости в соответствии с графическим отображением линии нормальной динамической растяжимости легких. Как известно, эта линия должна располагаться так, чтобы значению растяжимости 65 мл/см вод. ст. соответствовал угол наклона α оси петли 45° [4.39]. Поэтому при построении сигнатуры нами использована нормировка в виде $\bar{V}_n = V/V^*$ и $\bar{P}_n = P/P^*$, где нормирующие величины $V^* = 1300$ мл, $P^* = 20$ см вод. ст.

Далее выполняется обработка нормированных петель (построение сигнатуры, ее нормировка и смещение по оси θ), вычисление и сравнительный анализ параметров, характеризующих особенности формы спирографических кривых. В качестве этих параметров выбраны:

- направление основной оси анализируемой петли ОД ($\text{tg } \alpha$) и угловое отклонение от оси опорного дыхательного цикла ($\Delta\alpha$);
- среднеквадратичное отклонение σ_k сигнатуры;
- ширина петли r и оценка площади петли ОД в фазе выдоха S_5 ;
- выраженность артефактов L_a .

Режим слежения позволяет фиксировать отклонения этих признаков на значение, превышающее некоторый заданный в процентном отношении порог. Как показали эксперименты, сравнение петель целесообразно проводить по усредненным кривым, полученным путем их синхронного накопления. В этом случае значительно снижается влияние помех на качество анализа.

Наклон оси петли $\text{tg } \alpha$ является оценкой динамической растяжимости легких $C = \text{tg } \alpha$. Если измерять угловое отклонение $\Delta\alpha$, то можно оценивать степень изменения динамического комплайенса. Введение порога для $\Delta\alpha$ позволяет обнаруживать патологические изменения, связанные с резким увеличением или уменьшением растяжимости легких.

Среднеквадратичное отклонение сигнатуры используется в качестве оценки изменения формы петли относительно опорной кривой. Оно вычисляется в виде

$$\sigma_k = \sqrt{\frac{1}{n} \sum_{\theta} (r_{k\theta}^* - r_{k\theta})^2},$$

где $r_{k\theta}^*$ — отсчеты сигнатуры опорной петли, $r_{k\theta}$ — отсчеты сигнатуры для k -го порядкового дыхательного цикла, $\theta = 1-360^\circ$, n — размер выборки ($n = 360$).

Фактически данный параметр оценивает ошибку, связанную с отклонением в форме петель. Введение порога для величины σ_k позволяет выявить существенные отклонения в заданных режимах ИВЛ.

Сопровождение дыхательных путей можно оценивать одним из двух способов. Первый основан на измерении ширины петли r на уровне середины осевой линии. Вторым способом предполагает вычисление площади фигуры S_5 . Учитывая трудности интерпретации измеренных отклонений в случае изменения размеров петли (что связано, например, с необходимостью увеличения V_T при ИВЛ), необходимо использовать дополнительную нормировку петель, например по максимальным значениям объема V_T и давления P_{max} опорной петли. В этом случае опорная петля, условно считающаяся «нормой», располагается по линии нормальной динамической растяжимости, а изменения параметров r и S_5 подлежат сравнению даже при изменении параметров управления вентиляцией легких.

Показатель выраженности артефактов рассчитывался в виде $L_a = I^2/S$, где L — длина контура петли, S — площадь. Динамическая оценка его важна для выявления значительных колебаний контура петли, что связано с влиянием различных факторов (изменение тонуса мышц пациента, наличие препятствий в дыхательном контуре).

Экспериментальные исследования

Проведено моделирование алгоритма автоматического анализа петли ОД с применением специально разработанной программы,

в основе которой лежит функция signature [4.3]. В качестве входных сигналов использовались:

- модельные кривые, полученные путем перевода спирографических изображений в цифровую форму (12 записей) [4.39];
- сигналы, полученные на аппарате ИВЛ механическим моделированием изменений растяжимости C и сопротивления R (23 записи);
- реальные записи спирограмм, зарегистрированные в клинике в условиях ИВЛ (спирограммы 15 пациентов).

Реальные данные включали как выборку записей без патологий (5 записей), так и сигналы с тяжелыми нарушениями органов дыхания — ОРДС (10 записей) и обструкцией дыхательных путей (5 записей). Эти нарушения были рассмотрены в качестве основных при разработке алгоритма распознавания патологий.

Процесс обработки изображений, исходно представленных в виде битовых матриц, включал в себя следующие этапы: получение описания изображения в виде траектории единичной толщины; приведение описания изображений к фактическому масштабу, указанному для исходных графиков; нахождение узловых точек кривых; интерполяция петель кубическими сплайнами и получение сигналов в виде последовательностей равномерно дискретизованных отсчетов. По двумерным нормированным кривым определялись две опорные точки (начало и конец вдоха), по которым оценивался угол наклона петли α . Далее вычислялась функция сигнатуры и по ней путем сравнения одномерных описаний двух петель (анализируемой и опорной) определялись величины σ_k и r .

В качестве примера на рис. 4.27 и 4.28 представлены реальные петли ОД. На рис. 4.27, а, в, рис. 4.28, а, в приведены спирографические кривые; на рис. 4.27, б, г, рис. 4.28, б, г — сигнатуры петель. На каждом рисунке нижние кривые соответствуют нормальной форме спирограмм и являются опорными. Верхние петли, имеющие патологические отклонения, сравниваются с ними.

На рис. 4.27 приведены петли ОД для случая снижения величины COMPLAINT, которое вызывает поворот верхней петли вправо. Это изменение обусловлено тем, что для достижения данного уровня объема дыхания (650 мл) требуется большее давление. Подобный вид петель можно наблюдать на поздних стадиях ОРДС. Наблюдаемое снижение величины COMPLAINT может идти как постепенно — при развитии легочных заболеваний, так и внезапно — при острых нарушениях. На рис. 4.28 показан случай расширения петли ОД, вызванного повышением сопротивления дыхательных путей на выдохе, причем инспираторные участки двух кривых аналогичны. Возможными причинами таких отклонений может быть появление нарушений проходимости (обструкция дыхательных путей) [4.39].

Из сравнения числовых характеристик петель следует, что в первом случае (рис. 4.27) наблюдается снижение растяжимости легких ($\Delta\alpha = -15^\circ$), а во втором (рис. 4.28) наклон петли сохраняется

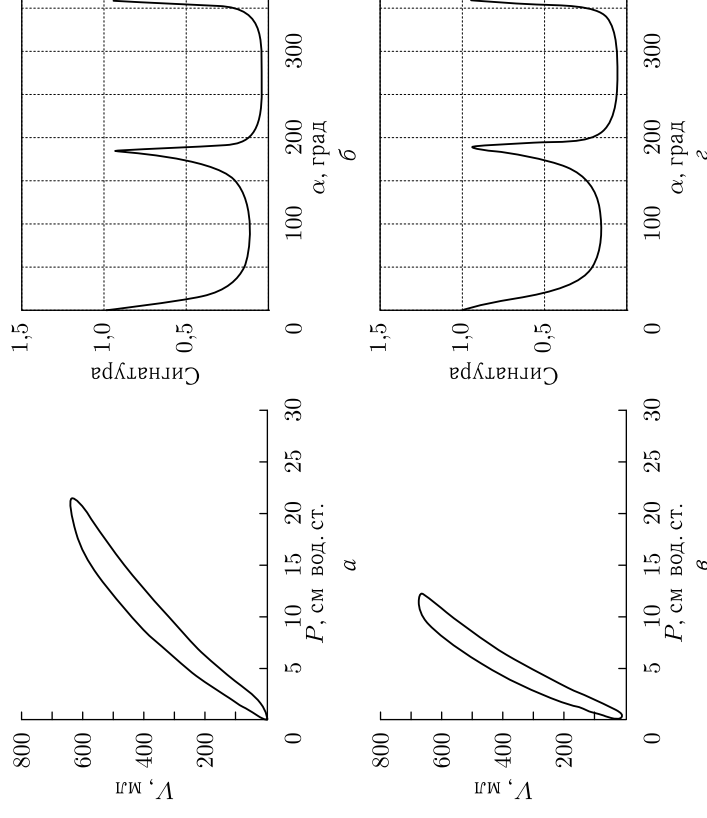


Рис. 4.27. Результаты сравнения петель ОД для случая снижения величины COMPLAINT: $\sigma = 0,08$; $\Delta\alpha = -15^\circ$; $\Delta r = 0,05$; $\Delta S_3 = 0,03$

($\Delta\alpha = 2^\circ$), но изменяется форма петли. Чем больше отклонения в форме петли, тем больших значений достигает среднеквадратичное отклонение. В нашем случае величина $\sigma = 0,31$ превышает пороговое значение, равное 0,2. Это связано с резким расширением петли ($\Delta r = 0,96$) и увеличением ее площади ($\Delta S_3 = 0,08$).

Распознавать эти опасные нарушения на ранних стадиях можно, используя полученные характеристики. Резкое уменьшение угла наклона α осевой линии петли ОД свидетельствует об увеличении жесткости легких и указывает на развитие ОРДС. Резкое увеличение σ , а также значений параметров r и S_3 свидетельствует о развитии обструкции дыхательных путей.

Проведены эксперименты по оценке эффективности распознавания различных форм петель (реальные записи спирограмм) и обнаружению значимых изменений динамической растяжимости C и сопротивления R дыхательных путей (модельные сигналы). При моделировании спирограмм задавались разные значения COMPLAINT и резистанса, что обеспечило возможность контроля изменений этих параметров. Результаты экспериментов на реальных записях показали, что во всех 15 случаях были безошибочно распознаны нарушения — ОРДС

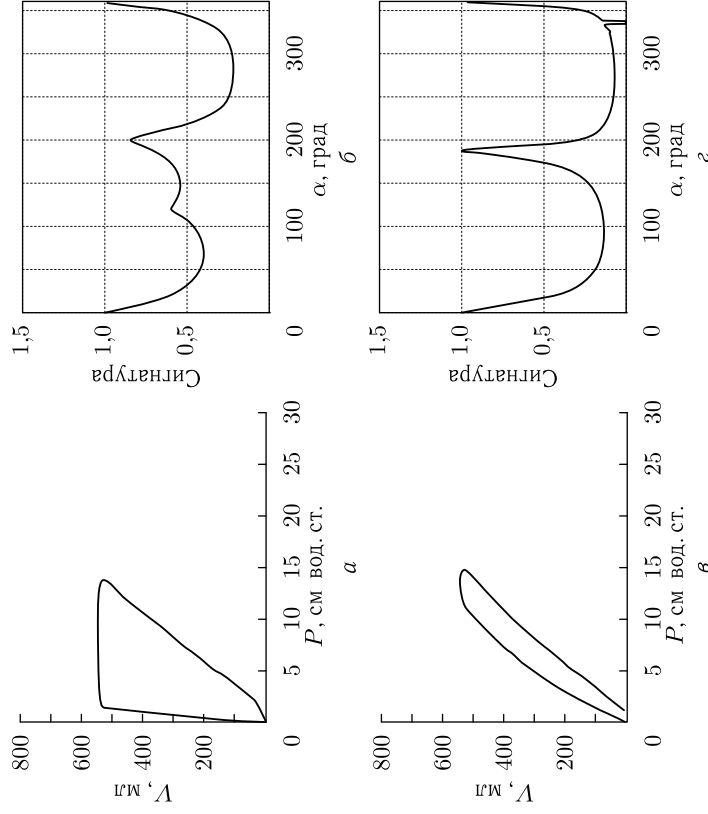


Рис. 4.28. Результаты сравнения петель ОД для случая повышения величины резистанса: $\sigma = 0,31$; $\Delta\alpha = 2^\circ$; $\Delta r = 0,96$; $\Delta S_3 = 0,08$

и обструкция дыхательных путей. В ходе экспериментов на модельных сигналах (30 испытаний) удалось безошибочно распознать кривые с измененной формой петли ОД, а также с отклонениями в динамической растяжимости легких. И только в двух случаях из 30 (6,7%) не удалось зафиксировать отклонения величины сопротивления дыхательных путей. Анализ записей показал, что эти ошибки связаны с заданием слишком малых величин управляемого потока.

Кроме того, проведен анализ статистической зависимости между заданными величинами R и вычисленными оценками резистанса: r , r_3 , S_3 . Значения парных коэффициентов корреляции k оказались достаточно высокими для всех трех параметров ($k > 0,85$), однако наибольшее значение $k = 0,91$ получено для параметра r . Следовательно, именно данный параметр наиболее важен для контроля значимых изменений резистанса по петлям ОД.

В результате проведенных исследований разработан комплекс программ автоматического анализа спирограмм, который предназначен для использования в аппаратах ИВЛ для расширения их диагностических возможностей.

Список литературы

Список литературы к гл. 1

- 1.1. Рангайян Р.М. Анализ биомедицинских сигналов. Практический подход / Пер. с англ. А. Н. Калинин. Под ред. А. П. Немирко. — М.: Физматлит, 2007. — 440 с.
- 1.2. Гублер Е.В. Информатика в патологии, клинической медицине и педиатрии. — Л.: Медицина, 1990.
- 1.3. Дюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях. — СПб.: Питер, 2003. — 528 с.
- 1.4. Кардиомониторы. Аппаратура непрерывного контроля ЭКГ / Под ред. А. Л. Барановского и А. П. Немирко. — М.: Радио и связь, 1993. — 248 с.
- 1.5. Ту Дж., Гонсалес Р. Принципы распознавания образов / Пер. с англ. под ред. Ю. И. Журавлева. — М.: Мир, 1978. — 411 с.
- 1.6. Дуда З., Харт П. Распознавание образов и анализ сцен / Пер. с англ. — М.: Мир, 1976. — 511 с.
- 1.7. Ватник В.Н., Червоненкис А.Я. Теория распознавания образов. Стохастические проблемы обучения. — М.: Наука, 1974.
- 1.8. Нильсон Н. Обучающиеся машины / Пер. с англ. — М.: Мир, 1967. — 180 с.
- 1.9. Фу К. Структурные методы в распознавании образов / Пер. с англ. — М.: Мир, 1977. — 320 с.
- 1.10. Гублер Е.В. Вычислительные методы анализа и распознавания патологических процессов. — Л.: Медицина, 1978. — 290 с.
- 1.11. Генкин А.А., Медведев В.И. Прогнозирование психофизиологических состояний. Вопросы методологии и алгоритмизации. — Л.: Наука, 1973. — 144 с.
- 1.12. Немирко А.П. Цифровая обработка биологических сигналов. — М.: Наука, 1984. — 144 с.
- 1.13. Айвазян С.А., Бержея З.И., Староверов О.В. Классификация многомерных наблюдений. — М.: Статистика, 1974. — 240 с.
- 1.14. Loog M., Duin R.P.W., Naeb-Umbach R. Multiclass linear dimension reduction by weighted pairwise Fisher criteria // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001. Vol. 23, № 7. P. 762–766.
- 1.15. Маннло Л.А. Преобразование пространства спектральных признаков с использованием весовых функций в задачах распознавания сигналов // Изв. вузов России. Сер. Радиоэлектроника. 2007. Вып. 2. С. 37–42.
- 1.16. Маннло Л.А., Немирко А.П. Сокращение размерности пространства спектральных признаков в многоклассовой задаче распознавания сигналов // Математические методы распознавания образов ММО-13: 13-я Всерос. конф. Ленингр. обл., г. Зеленогорск, 30 сент. — 6 окт. 2007 г.: Сб. докл. — М.: МАКС Пресс, 2007. С. 356–359.

- 1.17. Васильев В.И. Распознающие системы. — Киев: Наукова думка, 1983. — 422 с.
- 1.18. Перовзванский А.А. Распознавание абстрактных образов как задача линейного программирования // Изв. Академии наук СССР. Сер. Техническая кибернетика. 1965. № 4. С. 41–44.
- 1.19. Таха Х.А. Введение в исследование операций. 6-е изд. / Пер. с англ. — М.: Вильямс, 2001. — 912 с.
- 1.20. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. — М.: Наука. Гл. ред. физ.-мат. лит., 1983. — 464 с.
- 1.21. Ластед Л. Введение в проблему принятия решений в медицине. — М.: Мир, 1971.
- 1.22. Фрайтаг Г., Гуде В., Якоби Х. и др. Введение в технику работы с таблицами решений / Пер. с нем. под ред. Д. А. Постелова. — М.: Энергия, 1979. — 88 с.
- 1.23. Немирко А.П. Микропроцессорные медицинские диагностические системы: Учеб. пособие. — Л.: ЛЭТИ, 1984. — 64 с.
- 1.24. Хамби Э. Программирование таблиц решений / Пер. с англ. — М.: Мир, 1976. — 86 с.
- 1.25. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000 Jun 13; 101(23): E215-20.
- 1.26. Phisonet. <http://www.physionet.org>. 2016.
- 1.27. Nemirko A.P. Transformation of feature space based on Fisher's linear discriminant // Pattern Recognition and Image Analysis. 2016. V. 26, № 2. P. 257–261.
- 1.28. Iris Data Set. UCI Machine Learning Repository. <https://archive.ics.uci.edu/ml/datasets/Iris>, 2016.

Список литературы к гл. 2

- 2.1. Айфичер Э.С., Джервис Б.У. Цифровая обработка сигналов: практический подход / Пер. с англ. — М.: ИД «Вильямс», 2004. — 992 с.
- 2.2. Анищенко В.С. Знакомство с нелинейной динамикой. 3-е изд. — М.: Издательство ЛКИ, 2008. — 224 с.
- 2.3. Головкин В.А. Нейросетевые методы обработки хаотических процессов // Научная сессия МИФИ-2005: Лекции по нейроинформатике. — М.: МИФИ, 2005. — С. 43–91.
- 2.4. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения / Пер. с англ. — М.: Мир, 1971 (вып. 1); 1972 (вып. 2).
- 2.5. Дюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях. СПб.: Питер, 2003. — 528 с.
- 2.6. Зенков Л.Р. Клиническая электроэнцефалография с элементами эпипелтологии. — М.: МЕДпресс-Информ, 2004. — 367 с.
- 2.7. Немирко А.П., Манило Л.А., Калинин А.Н., Волкова С.С. Использование показателей энтропии ЭЭГ-сигнала для распознавания стадий наркоза // Интеллектуализация обработки информации: 8-я Международ. конф. Кипр, г. Пафос, 17–24 окт. 2010 г.: Сб. докл. — М.: МАКС Пресс, 2010. С. 454–457.

- 2.8. Барановский А.Л., Калинин А.Н., Манило Л.А. и др. Кардиомониторы. Аппаратура непрерывного контроля ЭКГ / Под ред. А. Л. Барановского и А. П. Немирко. — М.: Радио и связь, 1993. — 248 с.
- 2.9. Малинецкий Г.Г., Потапов А.Б., Подлазов А.В. Нелинейная динамика: подходы, результаты, надежды. Изд. 3-е. — М.: Книжный дом «Либроком», 2011. — 280 с.
- 2.10. Манило Л.А. Автоматический анализ электрокардиосигнала в задачах распознавания нарушений сердечного ритма // Биотехнические системы в медицине и биологии / Под общ. ред. проф. Е. П. Попечителя. — СПб.: Политехника, 2002. — С. 61–68.
- 2.11. Манило Л.А. Авторегрессионные модели случайных процессов в задачах распознавания нарушений сердечного ритма // Изв. СПбГЭТУ «ЛЭТИ». Сер. Биотехнические системы в медицине и экологии. 2004. Вып. 2. С. 1–8.
- 2.12. Манило Л.А., Волкова С.С. Анализ параметров аппроксимированной энтропии в задаче оценки глубины наркоза по ЭЭГ // Биомедицинская радиоэлектроника. Биомедицинские технологии и радиоэлектроника. 2012. № 1. С. 58–61.
- 2.13. Манило Л.А., Зозуля Е.П. Автоматическое распознавание мерцательной аритмии с использованием оценок аппроксимированной энтропии // Информационно-управляющие системы. 2006, № 1 (20). С. 21–27.
- 2.14. Манило Л.А., Зозуля Е.П. Исследование возможности применения аппроксимированной энтропии для анализа биосигналов // Изв. СПбГЭТУ «ЛЭТИ». Сер. Биотехнические системы в медицине и экологии. 2007. Вып. 1. С. 3–9.
- 2.15. Манило Л.А., Зозуля Е.П. Оценка корреляционной размерности временных рядов в задачах анализа аритмий // Биомедицинская радиоэлектроника. 2009, № 11. С. 32–39.
- 2.16. Манило Л.А. Идентификация биосигналов методами оценки условной энтропии // Изв. СПбГЭТУ «ЛЭТИ». Сер. Биотехнические системы в медицине и экологии. 2005. Вып. 2. С. 53–59.
- 2.17. Манило Л.А., Немирко А.П. Аппроксимация энтропии Колмогорова при анализе хаотических процессов на конечных выборках // Математические методы распознавания образов ММРО-14: 14-я Всерос. конф. Владимирская обл., г. Суздаль, 21–26 сент. 2009 г.: Сб. докл. — М.: МАКС Пресс, 2009. С. 405–407.
- 2.18. Манило Л.А. Распознавание фибрилляции предсердий в кардиологических системах диагностики и наблюдения // Биотехносфера. 2009, № 2. С. 41–45.
- 2.19. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения / Пер. с англ. — М.: Мир, 1990. — 584 с.
- 2.20. Микрокомпьютерные медицинские системы: проектирование и применение / Под ред. У. Томпкинса, Дж. Уэбстера. Пер. с англ. — М.: Мир, 1983. — 544 с.
- 2.21. Мун Ф. Хаотические колебания: Вводный курс для научных работников и инженеров / Пер. с англ. — М.: Мир, 1990. — 312 с.
- 2.22. Немирко А.П., Манило Л.А., Калинин А.Н., Волкова С.С. Сравнительный анализ применения различных оценок энтропии ЭЭГ-сигнала

- для распознавания стадий наркоза // Биотехносфера. 2010. Вып. 3. С. 3–10.
- 2.23. *Немирко А.П., Манило Л.А., Калинин А.Н., Волкова С.С.* Энтропийные методы оценки уровня анестезии по ЭЭГ-сигналу // Информационно-управляющие системы. 2010, № 3. С. 69–74.
- 2.24. *Рабинер Л., Голд Б.* Теория и применение цифровой обработки сигналов. — М.: Мир, 1978. — 848 с.
- 2.25. *Рангайян Р.М.* Анализ биомедицинских сигналов / Пер. с англ. А. Н. Калининченко под ред. А. П. Немирко. — М.: Физматлит, 2007. — 440 с.
- 2.26. *Хемминг Р.В.* Цифровые фильтры. — М.: Сов. радио, 1980. — 224 с.
- 2.27. *Шеннон К.* Работы по теории информации и кибернетике / Пер. с англ. — М., 1965. — 438 с.
- 2.28. *Шустер Г.* Детерминированный хаос: Введение. — М.: Мир, 1988. — 240 с.
- 2.29. Heart rate variability. Standards of measurements, physiological interpretation, and clinical use // Circulation. 1996. Vol. 93 (5). P. 1043–1065.
- 2.30. *Manilo L.A., Volkova S.S.* Recognition of the deep anesthesia stage from parameters of the approximated entropy of EEG signal // Pattern Recognition and Image Analysis. 2013. Vol. 23, № 1. P. 92–97.
- 2.31. *Manilo L.A.* Detection of biological signals with chaotic properties through assessment of conventional entropy // 9-th Intern. Conf. «Pattern Recognition and Image Analysis: New Information Technologies» (PRIA-9-2008). Nizhnii Novgorod. 2008. Vol. 2. P. 11–14.
- 2.32. *Porta A., Baselli G., Liberati D.* et al. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow // Biological Cybernetics. 1998. Vol. 78. P. 71–78.
- 2.33. Nonlinear Biomedical Signal Processing. Vol. 2. Dynamic Analysis and Modelling / Ed. by Metin Akay. — New York: IEEE, 2001. — 341 p.
- 2.34. *Pincus S.M.* Approximate entropy as a measure of system complexity // Proc. Natl. Acad. Sci. USA 88: 1991. P. 2297–2301.
- 2.35. *Гонсалес Р., Вудс Р.* Цифровая обработка изображений / Пер. с англ. — М.: Техносфера, 2005. — 1072 с.

Список литературы к гл. 3

- 3.1. *Табак В.Я., Черныш А.М., Немирко А.П., Манило Л.А.* Динамика спектральных характеристик ЭКГ при развитии фибрилляции желудочков сердца // Анестезиология и реаниматология. 1980, № 1. С. 71–74.
- 3.2. *Дюк В., Эммануэль В.* Информационные технологии в медико-биологических исследованиях. — СПб.: Питер, 2003.
- 3.3. *Калининченко А.Н.* Оценка разделяющей способности методов классификации форм ЭКГ // Изв. СПбГЭТУ «ЛЭТИ». Сер. Биотехнические системы в медицине и экологии. 2006. Вып. 1. С. 21–30.
- 3.4. *Барановский А.Л., Калинин А.Н., Манило Л.А.* и др. Кардиомониторы. Аппаратура непрерывного контроля ЭКГ / Под ред. А. Л. Барановского и А. П. Немирко. — М.: Радио и связь, 1993. — 248 с.
- 3.5. *Манило Л.А.* Упорядочение спектральных признаков по эмпирическим оценкам межгруппового расстояния в задачах классификации биосигналов // Изв. вузов России. Радиоэлектроника. 2006. Вып. 3. С. 20–29.

- 3.6. *Манило Л.А.* Линейный дискриминант Фишера в задачах распознавания биосигналов по частотным свойствам // Докл. XII Всерос. конф. «Математические методы распознавания образов» — ММРО-12, Москва, 20–26 нояб. 2005. — М.: МАКС Пресс, 2005. С. 371–374.
- 3.7. *Манило Л.А.* Построение решающих функций в пространстве спектральных признаков для систем кардиологического наблюдения // Изв. СПбГЭТУ «ЛЭТИ». Сер. Биотехнические системы в медицине и экологии. 2006. Вып. 1. С. 13–21.
- 3.8. *Марпл-мл. С.Л.* Цифровой спектральный анализ и его приложения / Пер. с англ. — М.: Мир, 1990. — 584 с.
- 3.9. Микрокомпьютерные медицинские системы: Проектирование и применение / Под ред. У. Томпкинса, Дж. Уэбстера; пер. с англ. — М.: Мир, 1983. — 544 с.
- 3.10. *Рангайян Р.М.* Анализ биомедицинских сигналов / Пер. с англ. А. Н. Калининченко под ред. А. П. Немирко. — М.: Физматлит, 2007. — 440 с.
- 3.11. *Ту Дж., Гонсалес Р.* Принципы распознавания образов / Пер. с англ. — М.: Мир, 1978. — 411 с.
- 3.12. *Уидроу Б., Смирнз С.* Адаптивная обработка сигналов. — М.: Радио и связь, 1989. — 440 с.
- 3.13. *Clayton R.H., Murray A.* Comparison of techniques for time-frequency analysis of the ECG during human ventricular fibrillation // IEE Proc. on Sci., Measurement and Technol. 1998, N 145. P. 301–306.
- 3.14. *Manilo L.A.* Measurement and computerized analysis of signals in biomedical systems // Proc. of the V Intern. Symp. on Test and Measurement. 2003. Vol. 1. China, June 1–5. P. 867–870.
- 3.15. *Manilo L.A., Nemirko A.P.* Forming the spectral signs when classifying the electrocardiosignals within the frequency range // Pattern Recognition and Image Analysis. 2005. Vol. 15, N 4. P. 668–671.
- 3.16. *Manilo L.A., Nemirko A.P.* Recognition of biomedical signals based on their spectral description data analysis // Pattern Recognition and Image Analysis. 2016. Vol. 26, N 4. P. 782–788.
- 3.17. MIT-BIH Arrhythmia Database. Available from MIT-BIH database distribution // Massachusetts Institute of Technology. <http://www.physio-net.org/physiobank/database/mitdb>; <http://ecg.mit.edu>.
- 3.18. *Nemirko A.P., Manilo L.A., Degtyareva I.N.* Recognition of heart ventricular fibrillation on the basis of analysis of the electrocardiogram in the frequency domain // Pattern Recognition and Image Analysis. 2001. V. 11, № 2. P. 353–355.

Список литературы к главе 4

- 4.1. *Богачев М.И., Мамонтов О.В., Конради А.О., Ульяницкий У.Д.* Оценка спонтанного артериального барорефлекса методом совместного анализа показателей кратковременной изменчивости артериального давления и сердечного ритма // Артериальная гипертензия. 2007. Т. 13, №1.
- 4.2. *Гонсалес Р., Вудс Р.* Цифровая обработка изображений / Пер. с англ. — М.: Техносфера, 2005. — 1072 с.
- 4.3. *Гонсалес Р., Вудс Р., Эддинс С.* Цифровая обработка изображений в среде MATLAB / Пер. с англ. — М.: Техносфера, 2006. — 616 с.

- 4.4. Дуда З., Харт П. Распознавание образов и анализ сцен / Пер. с англ. — М.: Мир, 1976. — 511 с.
- 4.5. Зенков Л.Р. Клиническая электроэнцефалография с элементами эпипелтологии. — М.: МЕДпресс-Информ, 2004. — 367 с.
- 4.6. Калининко А.Н., Юрева О.Д. Влияние частоты дискретизации ЭКГ на точность вычисления спектральных параметров вариабельности сердечного ритма // Информационно-управляющие системы. 2008, № 2. С. 46–49.
- 4.7. Калининко А.Н., Немирко А.П., Саламонова И.С. Использование полиномиальной интерполяции для измерения показателей искусственной вентиляции легких // Биомедицинская радиоэлектроника. 2013, № 11. С. 71–74.
- 4.8. Калининко А.Н., Коляденко М.И. Исследование алгоритмов оценки стационарности сердечного ритма // Изв. СПбГЭТУ «ЛЭТИ». Сер. Биотехнические системы в медицине и экологии. 2006. Вып. 2. С. 101–105.
- 4.9. Калининко А.Н. О точности и достоверности спектральных методов расчета показателей вариабельности сердечного ритма // Информационно-управляющие системы. 2007, № 6. С. 41–48.
- 4.10. Барановский А.Л., Калининко А.Н., Манило Л.А. и др. Кардиомониторы. Аппаратура непрерывного контроля ЭКГ / Под ред. А. Л. Барановского и А. П. Немирко. — М.: Радио и связь, 1993. — 248 с.
- 4.11. Лебединский К.М., Мазурок В.А., Нефедов А.В. Основы респираторной поддержки. — СПб.: Изд-во МАПО, 2006. — 213 с.
- 4.12. Манило Л.А., Немирко А.П., Саламонова И.С. Автоматический анализ формы спиреографических петель по их сигналам // Машинное обучение и анализ данных. 2013. Т. 1, № 6. С. 687–694.
- 4.13. Манило Л.А., Немирко А.П., Саламонова И.С. Автоматический анализ формы спиреографических петель при искусственной вентиляции легких // Изв. СПбГЭТУ «ЛЭТИ». 2013, № 8. С. 73–78.
- 4.14. Манило Л.А., Родина Н. И. Новый подход к спектральному анализу вариабельности сердечного ритма // Известия СПбГЭТУ «ЛЭТИ». Сер. Биотехнические системы в медицине и экологии. 2003. Вып. 1. С. 16–20.
- 4.15. Михайлов В.М. Вариабельность ритма сердца. Опыт практического применения. — Иваново: НейроСофт, 2000. — 200 с.
- 4.16. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения / Пер. с англ. — М.: Мир, 1990. — 584 с.
- 4.17. Немирко А.П., Манило Л.А., Калининко А.Н., Волкова С.С. Сравнительный анализ применения различных оценок энтропии ЭЭГ-сигнала для распознавания стадий наркоза // Биотехносфера. 2010. Вып. 3. С. 3–10.
- 4.18. Немирко А.П., Манило Л.А., Калининко А.Н., Волкова С.С. Энтропийные методы оценки уровня анестезии по ЭЭГ-сигналу // Информационно-управляющие системы. 2010, № 3. С. 69–74.
- 4.19. Павлидис Т. Алгоритмы машинной графики и обработки изображений / Пер. с англ. — М.: Радио и связь, 1986. — 400 с.
- 4.20. Рангайян Р.М. Анализ биомедицинских сигналов / Пер. с англ. А. Н. Калининко под ред. А. П. Немирко. — М.: Физматлит, 2007. — 440 с.

- 4.21. Царенко С.В. Практический курс ИВЛ. — М.: Медицина, 2007. — 160 с.
- 4.22. Bein B. Entropy / Best Practice & Research Clinical Anaesthesiology. 2006. V. 20, № 1 P. 101–109.
- 4.23. Bruhn J. et al. Shannon entropy applied to the measurement of the electroencephalographic effects of desflurane // Anesthesiology. 2001. V. 95. P. 30–35.
- 4.24. Karason S., Sondergaard S., Lundin S., Stenquist O. Continuous on-line measurements of respiratory system, lung, and chest wall mechanics during mechanic ventilation // Intensive Care Medicine. 2001. V. 27. P. 1328–1339.
- 4.25. Gill P.E., Murray W., Wright M.H. Practical Optimization. — L.: N.Y.: Academic Press, 1981. — 401 p.
- 4.26. Guimaraes H.N., Santos R.A.S. A comparative analysis of preprocessing techniques of cardiac event series for the study of heart rhythm variability using simulated signals // Braz. J. Biol. Res. 1998. V. 31(1). P. 421–430.
- 4.27. Heart rate variability. Standards of measurements, physiological interpretation, and clinical use // Circulation. 1996. V. 93 (5). P. 1043–1065.
- 4.28. Hughes S., Griffiths R. Anaesthesia monitoring techniques // Anaesth. and Intensive Care Medicine. 2002. P. 477–480.
- 4.29. Hyndman B.W., Mohn R.K. A model of the cardiac pacemaker and its use in decoding the information content of cardiac intervals // Automedica. 1975. V. 1. P. 239–252.
- 4.30. Kaul H.L., Bharti N. Monitoring depth of anaesthesia // Indian J. Anaesth. 2002. V. 46(4). P. 323–332.
- 4.31. Kalinichenko A.N., Manilo L.A., Nemirko A.P. Analysis of anesthesia stages based on the EEG entropy estimation // Pattern Recognition and Image Analysis. 2015. V. 25, № 4. P. 680–689.
- 4.32. Kalinichenko A., Manilo L., Nemirko A. Recognition of anesthesia stages based on nonlinear EEG analysis // Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry, and Food Industry. Proceedings of the 9th International Conference, MDA 2014, St. Petersburg, Russia, July 2014 / Ed. by P. Perner. — Ibai-Publishing. P. 25–31.
- 4.33. La Rovere M.T., Pinna G.D., Raczak G. Baroreflex sensitivity: measurement and clinical implications // Ann. Noninvasive Electrocardiol. 2008 Apr. 13(2). P. 191–207.
- 4.34. Manilo L.A., Rodina N.I. Investigation of a model of the cardiac rhythm pacemaker control for the spectral analysis of a rhythmogram // Pattern Recognition and Image Analysis. 2001. V. 11, № 2. P. 342–344.
- 4.35. Manilo L.A., Rodina N.I. Estimation of the frequency properties of rhythmograms in problems of recognition of physiological states // Pattern Recognition and Image Analysis. 2003. V. 13, № 2. P. 298–301.
- 4.36. Mehta S., Hill N.S. Noninvasive ventilation // Am. J. Respir. Crit. Care Med. 2001. Feb. 163(2). P. 540–577.
- 4.37. Kheirani S., Polese G., Aliverti A. et al. On-line monitoring of lung mechanics during spontaneous breathing: a physiological study // Respiratory Medicine. 2010. V. 104. P. 463–471.

- 4.38. *Stenquist O.* Practical assessment of respiratory mechanics // *British J. of Anaesth.* 2003. V. 91 (1). P. 92–105.
- 4.39. *Waugh J.B., Deshpande V.M., Brown M.K., Harwood R.J.* Rapid Interpretation of Ventilator Waveforms. — Upper Saddle River, New Jersey: Pearson/Prentice Hall, 2007. — 151 p.
- 4.40. *Steight P., La Rovere M.T., Mortara A., Pinna G., Maestri R., Leuzzi S., Bianchini B., Tavazzi L., Bernardi L.* Physiology and pathophysiology of heart rate and blood pressure variability in humans. Is power spectral analysis largely an index of baroreflex gain? // *Clinical Science.* 1995. V. 88. P. 103–109.
- 4.41. *Tong S., Thakor N.V.* Quantitative EEG Analysis Methods and Clinical Applications. — Boston: Artech House, 2009. — 440 p.
- 4.42. *Viertio-Oja H., Maja V., Sarkela M., Talja P.* et al. Description of the entropy algorithm as applied in the Datex-Ohmeda S/5 Entropy Module // *Acta Anaesthesiol. Scand.* 2004. V. 48. P. 154–161.