



НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
МЕЖДУНАРОДНЫЙ НАУЧНО-УЧЕБНЫЙ ЦЕНТР
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Л.С. ФАЙНЗИЛЬБЕРГ

**ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ
ОБРАБОТКИ
СИГНАЛОВ
СПОЖНОЙ ФОРМЫ**
Теория и практика

КИЕВ НАУКОВА ДУМКА 2008

В монографии изложены основы теории и практики построения прогрессивных информационных технологий, предназначенных для анализа и интерпретации сигналов сложной формы. Основное внимание уделено методологическим основам построения таких технологий. Предложены и исследованы оригинальные методы и эффективные вычислительные алгоритмы извлечения диагностической информации из реализаций, искаженных внутренними и внешними возмущениями. Приведены примеры практического применения этих результатов при синтезе прикладных информационных технологий оценки состояния технических и биологических объектов по сигналам с локально сосредоточенными диагностическими признаками.

Для читателей, интересующихся теорией и практикой современных информационных технологий обработки сигналов.

У монографії викладено основи теорії і практики побудови прогресивних інформаційних технологій, призначених для аналізу та інтерпретації сигналів складної форми. Головну увагу приділено методологічним основам побудови таких технологій. Запропоновано і досліджено оригінальні методи та ефективні обчислювальні алгоритми видобування діагностичної інформації з реалізацій, спотворених внутрішніми і зовнішніми збуреннями. Наведено приклади практичного застосування цих результатів при синтезі прикладних інформаційних технологій оцінювання стану технічних і біологічних об'єктів за сигналами з локально зосередженими діагностичними ознаками.

Для читачів, що цікавляться теорією і практикою сучасних інформаційних технологій обробки сигналів.

Печатается по решению ученого совета Международного научно-учебного центра информационных технологий и систем НАН Украины и МОН Украины (протокол № 13 от 14 декабря 2007 г.)

Рецензенты: доктор технических наук, профессор *А.А. Туник*,
доктор технических наук *Г.С. Теслер*

Научно-издательский отдел физико-математической
и технической литературы

Редактор *С.Е. Ноткина*

ISBN 978-966-00-0805-2

© Л.С. Файнзильберг, 2008
© НПП «Издательство “Наукова думка”
НАН Украины», дизайн, 2008

ВВЕДЕНИЕ

Профессиональная деятельность и частная жизнь человека постоянно связаны с необходимостью принятия тех или иных решений. Понятно, что решения могут быть эффективными лишь в том случае, когда они опираются на достоверную и своевременную информацию. Немаловажна также форма представления информации, поскольку человеческий мозг не способен адекватно оценивать большие объемы пусть даже достоверной, но не структурированной информации.

Информация наряду с материей и энергией является первичным понятием нашего мира и поэтому в строгом смысле не может быть определена. Можно лишь перечислить ее основные свойства:

- информация дает *новые* сведения об объектах, явлениях и процессах, которые не были известны до ее получения, иными словами, информация снижает неопределенность нашего представления об окружающей среде;
- информация преподносится в форме *материальных* носителей сообщений, которые должны быть понятны конечному пользователю.
- информация может быть *объектом* операций восприятия, накопления, преобразования, использования и передачи.

Хотя сам термин *информация* существует достаточно долго и интуитивно понятен лю-

бому человеку, только в первой половине XX в. появились научные дисциплины, изучающие структуру и закономерности протекания информационных процессов. Во второй половине XX в. стремительно развились средства вычислительной техники и электроники, появились оригинальные методы численной обработки данных, произошла эволюция языков программирования высокого уровня. Сформировалось новое научное направление — информатика, которая применяется во всех областях человеческой деятельности.

Сбываются предсказания академика В.М. Глушкова и других ученых о том, что в XXI в. в технически развитых странах мира основная масса информации будет храниться и обрабатываться с помощью электронных вычислительных машин.

Современные информационные технологии (ИТ) [71] служат информационной поддержкой при принятии решений и управлении в науке, экономике, промышленности, медицине и других областях. Продуктом ИТ является информация, подготовленная в соответствии с потребностями конкретных пользователей в той или иной предметной области.

Одно из перспективных направлений использования современных ИТ — диагностика состояния технических и биологических объектов. Обычно такие технологии основаны на компьютерной обработке доступных сигналов, которые порождает объект в процессе своего функционирования. Предполагается, что такой сигнал несет информацию о количественных или качественных показателях текущего состояния объекта, а задача информационной технологии — извлечь эту информацию из сигнала.

Общим закономерностям извлечения информации из наблюдений посвящены фундаментальные научные труды Р.Л. Хартли, К. Шеннона, Н. Винера, А.Н. Колмогорова, В.А. Котельникова, А.Я. Хинчина, А.А. Харкевича и других основоположников теории информации. Значительный вклад в развитие теории цифровой обработки сигналов внесли Т.К. Винцюк, Я.П. Драган, В.К. Задирака, Ю.П. Кунченко, В.А. Омельченко и многие другие украинские ученые.

В тех случаях, когда информация о состоянии объекта равномерно «распределена» на интервале наблюдений, для ее извлечения могут быть привлечены классические методы обработки сигналов, в частности ортогональные разложения случайных сигналов по системе базисных функций. Тогда коэффициенты

функционального ряда могут использоваться в качестве диагностических признаков, характеризующих состояние объекта.

Задача существенно усложняется, если информацию о состоянии объекта несут диагностические признаки, сосредоточенные на небольших фрагментах интервала наблюдения сигнала, а сам сигнал имеет достаточно «сложную» форму, которую невозможно описать моделью с небольшим количеством параметров. Классическими примерами таких сигналов являются электрокардиограммы (ЭКГ), магнитокардиограммы (МКГ), реограммы и другие биологические сигналы, локальные фрагменты которых несут ценную информацию о состоянии сердечно-сосудистой системы живого организма.

Сигналы сложной формы с локально сосредоточенными признаками встречаются и в других приложениях. Например, в металлургии, минералогии, петрографии и химии широко используют так называемый метод термического анализа. В соответствии с этим методом физико-химические свойства многокомпонентного материала определяют по характерным фрагментам сигнала, несущего информацию об изменении температуры образца в процессе его кристаллизации. Список подобных примеров можно было бы продолжить.

Классические методы мало пригодны при обработке сигналов сложной формы, поскольку коэффициенты функционального ряда, построенного на всей области определения, почти не чувствительны к вариациям амплитудно-временных признаков, сосредоточенных на локальных фрагментах наблюдаемого сигнала. В то же время именно изменение таких признаков несет соответствующую диагностическую информацию о текущем состоянии объекта.

В последние годы развиваются альтернативные подходы к анализу локализованных свойств сигналов, в частности метод вейвлет-преобразований, ориентированный на оценку локализованных во времени частотных компонент сигнала. Получили известность также методы обнаружения разладки статистических свойств случайных последовательностей, лингвистический подход к анализу структуры сигналов сложной формы и ряд других новых научных направлений.

Однако, по оценке специалистов, существующие прикладные ИТ обработки сигналов сложной формы все еще не обеспечивают требуемую достоверность результатов диагностики. В значительной мере это обусловлено несовершенством математических

моделей, описывающих механизм порождения таких сигналов в условиях возмущений, которые в реальных ситуациях не могут быть сведены лишь к аддитивному шуму. Требуют дальнейшего развития также научно обоснованные подходы к построению математических методов и компьютерных алгоритмов выделения полезного сигнала сложной формы по искаженным реализациям. Необходимо усовершенствовать формальные критерии, позволяющие по экспериментальным данным оценивать диагностическую эффективность локально сосредоточенных признаков.

Практически отсутствуют монографии методологического плана, в которых бы охватывался весь комплекс научных и прикладных вопросов, возникающих перед конструктором новой ИТ анализа и интерпретации сигналов с локально сосредоточенными признаками.

Настоящая монография преследует цель восполнить этот пробел. В ней представлена информация об оригинальных математических моделях, методах, вычислительных алгоритмах и программных компонентах, ориентированных на обработку сигналов сложной формы. Тем самым строится своеобразный «мост» между теорией и практикой построения прогрессивных ИТ извлечения диагностической информации из таких сигналов.

Книга состоит из шести глав, которые желательно читать последовательно.

В первой главе формулируется общая проблема построения ИТ для оценки состояния объектов по наблюдаемым сигналам, которые порождает исследуемый объект в процессе своего функционирования. Рассмотрена рабочая гипотеза, позволяющая обосновать принципиальную возможность построения таких технологий. Проведена сравнительная оценка известных подходов к анализу свойств сигналов, которые могут быть использованы в качестве потенциально полезных диагностических признаков для оценки состояния объектов различной физической природы.

Во второй главе рассматриваются методологические основы построения ИТ, ориентированных на извлечение диагностической информации из искаженных сигналов сложной формы. Развивается стохастическая модель порождения таких сигналов в условиях внутренних и внешних возмущений, которые приводят к неравномерным изменениям области определения полезного сигнала. Предлагается обобщенная схема извлечения локализованной диагностической информации из искаженных реализаций сигналов сложной формы.

В третьей главе рассматривается оригинальный метод восстановления полезного сигнала сложной формы по последовательности искаженных реализаций. Метод основан на переходе от дискретной последовательности скалярных значений сигнала к последовательности векторов в нормированном фазовом пространстве, усреднение которых с использованием хаусдорфовой метрики позволяет оценить эталон полезного сигнала в фазовом пространстве и во временной области. Приводятся результаты теоретических и экспериментальных исследований этого метода на основе предложенной стохастической модели порождения сигнала. Наглядно иллюстрируется преимущество метода перед традиционным подходом, основанным на усреднении искаженных реализаций во временной области.

В четвертой главе формулируются основные принципы построения инструментальной системы СИДИГРАФ, предназначенной для интерактивного синтеза прикладных ИТ обработки сигналов сложной формы. Ядро системы построено на двух объектно-ориентированных классах, представляющих собой обобщенную модель носителя данных и обобщенную модель виртуальной обработки данных. На базе наследников этих классов реализован широкий набор вычислительных компонент, использующих традиционные и оригинальные процедуры предварительной обработки и анализа сигналов сложной формы. На примере решения задачи подавления аддитивных внешних помех различного типа демонстрируется эффективность оригинальных вычислительных процедур.

Пятая глава посвящена актуальной задаче, с которой неизбежно сталкивается конструктор прикладной ИТ на этапе интерпретации сигналов сложной формы: исследуются формальные условия полезности диагностических признаков и методы интеграции частных решений группы независимых тестов в условиях противоречий; показывается существование граничных значений вероятностей ошибок пропуска цели и ложной тревоги при решении задач диагностики в статистической постановке; описывается усовершенствованный метод оценки эффективности диагностических тестов, основанный на анализе экспериментальной *ROC*-кривой; формулируется общее условие бесполезности релевантных признаков и доказывается ряд теорем, определяющих достаточные условия полезности диагностических признаков.

В шестой главе рассмотрены примеры практического использования полученных теоретических результатов при синтезе ИТ оценки состояния технических и биологических объектов по сигналам сложной формы. Описаны принципы построения созданных под руководством автора данной монографии прикладных компьютерных систем «ТЕРМОГРАФ», «ЧУГУН», «FAZAWIN», «ФАЗАГРАФ», «ИКАР» и «MAGWIN». Приведены результаты испытаний и использования этих систем в Украине, России, Венгрии и Германии. Намечены перспективы развития информационных технологий обработки сигналов сложной формы для решения актуальных прикладных задач.

Автор считает своим приятным долгом выразить признательность академику В.И. Скурихину и профессору В.И. Гриценко за поддержку и взаимопонимание, а также рецензентам монографии доктору технических наук, профессору А.А. Тунику и доктору технических наук Г.С. Теслеру за ценные замечания. Автор благодарен своим молодым коллегам, которые принимали участие в разработке прикладных систем обработки сигналов сложной формы.

*Линия жизни не только твоя,
Всех предыдущих твоих поколений,
Ты — как итог этих чудных ветвлений,
В данный момент заслужил бытия.*

*С миром тобою они говорят,
Ты — слабый шанс в этом свете остаться,
Будь же гуманным, сумей постараться,
Не обрывай на себе этот ряд.*

Татьяна Потапова

ПРОБЛЕМА ОЦЕНКИ СОСТОЯНИЯ ОБЪЕКТОВ В УСЛОВИЯХ НЕПОЛНОЙ ИНФОРМАЦИИ

1.1. Общая формулировка проблемы

В задачах управления состояние объекта принято характеризовать совокупностью значений некоторых величин, определяющих его поведение [119, с. 26]. В задачах технической и медицинской диагностики чаще всего изучается качественная характеристика состояния объекта путем отнесения его текущего состояния к одному из возможных классов (диагнозов) [28].

Используя понятие шкалы измерения [198], введем обобщенный показатель состояния объекта [207], под которым в первом случае будем понимать вектор $s = (s_1, \dots, s_N)$ с числовыми компонентами $s_1, \dots, s_N$, а во втором — значение индикаторной переменной $s = 1, \dots, M$, характеризующей принадлежность текущего состояния технического или биологического объекта к одному из M классов заданного множества $V = \{V_1, \dots, V_M\}$.

Довольно часто на практике непосредственное определение s невозможно из-за отсутствия соответствующих датчиков и индикаторов. Например, агрессивность и высокая температура среды препятствуют созда-

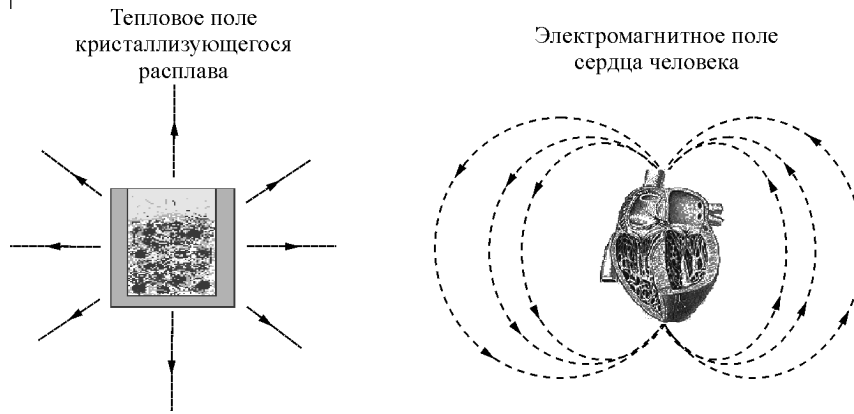


Рис. 1.1. Примеры полей, несущих информацию о состоянии объекта

нию датчиков для непосредственного определения химического состава и других параметров состояния расплавленного металла, что мешает построению эффективных систем управления металлургическими процессами [192]. Невозможно также при массовых профилактических обследованиях определять состояние внутренних органов человека на основе прямых (инвазивных) методов диагностики, например метода коронарографии для диагностики заболеваний сердечно-сосудистой системы человека. Этот метод достаточно дорогой и, самое главное, небезопасный для обследуемого пациента. Список подобных примеров можно было бы продолжить.

В этих ситуациях не остается ничего иного, как косвенно оценивать s , используя доступную информацию. Решение такой задачи является главной целью информационных технологий (ИТ), рассматриваемых в монографии. Какая же рабочая гипотеза может быть положена в основу синтеза таких технологий?

Предположим, что некоторый технический или биологический объект при фиксированном состоянии $s \in S$ порождает физическое поле (рис. 1.1), которое под действием процессов, происходящих внутри объекта, изменяется во времени. Представим это поле в виде

$$y_0 = \Psi_s(u, t), \quad (1.1)$$

где t — время; u — точка пространства; Ψ_s — некоторая функция.

Если зафиксировать точку пространства $u = u_0$ и измерить величину поля в этой точке, то функция $\Psi_s(u, t)$ вырождается [65] в скалярную функцию времени

$$y_0 = \Psi_s(u_0, t) \equiv y_0(t), \quad (1.2)$$

а при измерении поля (1.1) в нескольких фиксированных точках $u_1, u_2, \dots, u_p$ пространства — в векторную функцию $y_0(t) = (y_1(t), \dots, y_p(t))$.

Поскольку сигнал (1.2) зависит от s , то, с формальной точки зрения, ИТ должна обеспечивать решение обратной задачи [207]: требуется оценить неизвестное значение s по результатам измерения сигнала $y_0(t)$ на отрезке наблюдения $t_1 \leq t \leq t_2$ (рис. 1.2).

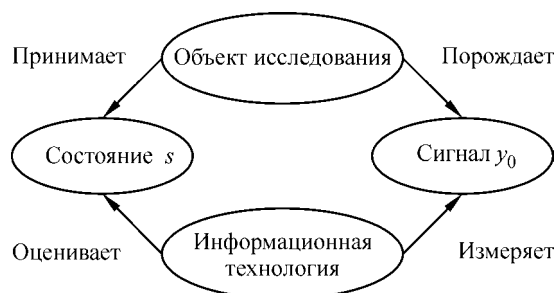


Рис. 1.2. Абстрактная модель ИТ обработки сигнала

Однако на практике функция $\Psi_s(\cdot)$ чаще всего не известна. Обычно конструктору прикладной ИТ приходится сталкиваться с достаточно сложными объектами и процессами, а знания в рассматриваемой предметной области слабо структурированы и формализации поддаются лишь отдельные фрагменты общей постановки [23]. Для упрощения изучения физических закономерностей часто рассматривают «идеализированные» условия, в результате чего полученные таким способом модели приводят к значительным погрешностям, которые не могут быть использованы на практике [105, с. 5]. И хотя есть все основания полагать, что модель (1.1) существует, получить ее только на основе физических представлений сложно [95, с. 7], а иногда и просто невозможно.

Так, уже давно известно, что электромагнитное поле сердца, зарегистрированное в течение сердечных циклов, несет важную ин-

формацию о состоянии сердечно-сосудистой системы живого организма. Однако до сего времени не получены «физические» модели, адекватно описывающие процесс порождения таких полей при фиксированных патологиях сердца [137, с. 25].

Есть все основания полагать, что температурное поле, регистрируемое вокруг кристаллизующегося образца, содержит ценную информацию о свойствах, структуре и химическом составе образца [25, 150]. В то же время процесс кристаллизации представляет собой теснейшее переплетение физико-химических, гидродинамических и тепловых явлений. Сложность этих процессов препятствует построению модели вида (1.1), адекватно описывающей температурное поле, порождаемое при фиксированном химическом составе металла [93].

Поскольку подобные примеры довольно часто встречаются в технике, медицине и других областях приложения, будем рассматривать модель (1.1) лишь как рабочую гипотезу, позволяющую обосновать принципиальную возможность построения ИТ для косвенной оценки состояния объекта по наблюдаемому сигналу. При этом следует понимать, что в условиях ограниченной априорной информации конструктору прикладной ИТ не остается ничего иного, как:

- вводить и проверять на имеющихся экспериментальных данных гипотезы, опираясь на ограниченные, а порой и противоречивые знания экспертов предметной области;
- методом проб и ошибок искать подходящую технологическую цепочку вычислительных процедур, обеспечивающую оптимальный в некотором смысле переход от исходного «сырья» (наблюдаемого сигнала) к информационному продукту (количественным и качественным оценкам состояния объекта).

Понятно, что в условиях неполной априорной информации о физических закономерностях процесса порождения сигнала $y_0(t)$ проблема построения ИТ не может быть сведена лишь к математической задаче, допускающей формальное решение. Поэтому будем далее рассматривать эту проблему в более широком смысле как методологию решения практических задач на основе *формальных* и *неформальных* приемов выбора и настройки эффективных вычислительных процедур обработки сигналов [214]. Тем более что сам термин «технология» происходит от греческого слова «*téchne*» (искусство, мастерство, умение), а задача технологии как науки состоит в выявлении закономерностей с целью

определения и использования на практике наиболее эффективных и экономных производственных процессов [188].

Разумеется, разрешение сформулированной проблемы должно опираться на имеющийся к настоящему времени огромный научный потенциал в области цифровой обработки сигналов, методов распознавания образов и моделирования сложных систем по экспериментальным данным. Однако анализ доступной научно-технической информации показывает, что этот потенциал часто недоиспользуется в существующих прикладных системах.

И это не случайно. Во-первых, прикладная ИТ может быть построена лишь на основе конструктивных вычислительных методов, эффективность которых обоснована не только теоретическими предпосылками и ограничениями, которые порой сложно проверить по имеющимся данным, но и практикой их использования на реальных сигналах. Однако часто принятые ограничения могут не соответствовать практической задаче.

Во-вторых, подавляющее большинство вычислительных алгоритмов, описываемых в научных публикациях, даже если они подтверждены примерами использования на модельных и реальных данных, содержит скрытые от читателя «ноу-хау», известные только их разработчикам. Не является исключением из этого правила и ряд процедур обработки сигналов, реализованных в известных системах компьютерной математики, обзор которых дан в обширных монографиях [63, 82, 174] и других аналогичных изданиях.

Практически нет монографий методологического плана, в которых бы охватывался весь комплекс вопросов разработки методов, вычислительных алгоритмов и программного обеспечения для оценки состояния сложных систем по наблюдаемым сигналам. Поэтому необходимо установить общие закономерности формирования информационного продукта в таких технологиях, по крайней мере, для некоторого класса технических и биологических объектов, реализовать базовый набор инструментальных средств, обеспечивающих эффективную обработку реальных сигналов, и тем самым построить мост между теоретическими положениями и прикладными методами построения ИТ. Важно провести систематизацию и сравнительную оценку наиболее известных методов и вычислительных алгоритмов в рассматриваемой области знаний. Прежде чем перейти к решению этой задачи приведем одну цитату из монографии [23].

«В современной науке весьма распространена такая ситуация, когда некий Исследователь берется за трудную Проблему и решает ее или показывает, что она не может быть решена. Действия его в данном случае высоко оцениваются окружающими и выглядят чрезвычайно привлекательно. Однако практика требует и других исследований. Не менее важным и нужным оказывается проводить сравнительный анализ давно известных объектов: задач, моделей, методов, схем и т. п. — с тем, чтобы сгруппировать их, наметить пути к упорядоченному, обоснованному исследованию и использованию. Подобные методологические проблемы намного менее притягательны и почитаемы. Более того, нередко соответствующие исследования оцениваются по набившему оскомину стереотипу: ничего экстраординарного, объекты давно известны, просто переставлены по-новому! Все же такие исследования нужно проводить, ибо это не что иное, как попытки системного анализа соответствующего научно-практического направления. Именно в результате подобной инвентаризации и могут обнаруживаться «белые пятна», нерешенные задачи и даже новые направления исследований».

Приведенные соображения в полной мере соответствуют представлениям автора и послужили хорошим стимулом в проведении исследований, результаты которых описываются далее.

1.2. Методы построения моделей сложных систем по экспериментальным данным

Пусть требуется *косвенно* оценить численное значение скалярной величины s по результатам *прямых* измерений других величин $x_1, \dots, x_N$ в предположении, что объективно существует зависимость

$$s = s(x) \equiv s(x_1, \dots, x_N), \quad x \in X. \quad (1.3)$$

При известной функции $s(\cdot)$ задача тривиальна. В тех же случаях, когда эта функция не известна, возникает проблема ее восстановления по выборке экспериментальных данных, или, как принято говорить, проблема идентификации [73, 173].

Строго говоря, если множество X бесконечно, то *подтвердить* правомерность гипотезы (1.3) можно, лишь проверив бесконечное число утверждений

$$\forall (\mu, v) \quad \mu \in X, \quad v = s(\mu),$$

где (μ, ν) — пара значений переменных x и s соответственно, что невозможно без привлечения *дополнительной* информации помимо той, которая содержится в выборке [95, с. 51]. Еще сложнее обосновать утверждение

$$s = s(x) + \xi, \quad x \in X, \quad (1.4)$$

где ξ — случайная величина, пусть даже с нулевым математическим ожиданием и конечной дисперсией. Сложность здесь состоит в том, что, даже чтобы *опровергнуть* гипотезу (1.4), требуется привлечение аппарата статистической проверки гипотез, и, по существу, ее нельзя отвергнуть окончательно, а лишь на некотором уровне значимости используемого статистического критерия [95, с. 52].

Тем не менее подобными «рабочими» гипотезами приходится пользоваться при решении прикладных задач. Например, при использовании одного из классических методов — регрессионного анализа [2, 11, 32, 56, 73, 76, 80, 121, 162, 173], прямо или косвенно постулируется, что взаимосвязь s с вектором $x = (x_1, \dots, x_N)$ можно описать функциональной зависимостью

$$s = \eta(x, c, \xi), \quad (1.5)$$

где c — вектор параметров функции η , а ξ — вектор случайных факторов (помех, возмущений и т. п.).

При таких допущениях идентификация модели сводится к построению уравнения регрессии $\hat{s} = f(x)$, связывающего средние значения величины s со значениями вектора x . Для этого используют результаты активного или пассивного эксперимента (выборку $W \geq N$ наблюдений), которые можно представить в виде матрицы значений входных переменных $x_1, \dots, x_N$ и вектора значений выходной переменной s [55]:

$$X_W = \begin{bmatrix} x_{11} & \dots & x_{1N} \\ \dots & \dots & \dots \\ x_{W1} & \dots & x_{WN} \end{bmatrix}; \quad S_W = \begin{bmatrix} s_1 \\ \dots \\ s_W \end{bmatrix}.$$

Принято рассматривать задачу идентификации в «широком» смысле, когда необходимо найти вид (структуру) функции $f(\cdot)$ и оценить вектор параметров c [76, 272], и в более узком смысле — как задачу параметрической идентификации [2, 32, 332], когда

структура зависимости считается известной и необходимо лишь оценить ее параметры.

Обычно искомую функцию представляют взвешенной суммой [11]:

$$\hat{s} = c^T \varphi(x) \equiv c_1 \varphi_1(x) + \dots + c_K \varphi_K(x), \quad (1.6)$$

где T — знак транспонирования; $\varphi(x) = (\varphi_1(x), \dots, \varphi_K(x))$ — вектор базисных функций $\varphi_1(x), \dots, \varphi_K(x)$, которые считаются заранее известными, а $c = (c_1, \dots, c_K)$ — вектор неизвестных параметров, подлежащих определению по выборке. Чаще всего оптимальное значение c_{opt} вектора параметров $c_1, \dots, c_K$, обеспечивающее минимум квадратичной функции потерь

$$R_{\Sigma^2} = \sum_{i=1}^W (s_i - \sum_{j=1}^K c_j \varphi_j(x))^2, \quad (1.7)$$

находят, используя известный метод наименьших квадратов (МНК) [2, 76, 120], в соответствии с которым

$$c_{\text{opt}} = (F_W^T F_W)^{-1} F_W^T S_W,$$

где

$$F_W = \begin{bmatrix} \varphi_1(x_{11}, \dots, x_{1G}) & \dots & \varphi_K(x_{11}, \dots, x_{1G}) \\ \dots & \dots & \dots \\ \varphi_1(x_{N1}, \dots, x_{NG}) & \dots & \varphi_K(x_{N1}, \dots, x_{NG}) \end{bmatrix}$$

— матрица значений базисных функций модели (1.6), вычисленных в точках эксперимента (элементах матрицы X_W).

Ряд теоретических предположений регрессионного анализа может рассматриваться лишь как средства более или менее удачного приближения реальной ситуации, например предположение о том, что ошибки эксперимента $\xi_1, \dots, \xi_W$ аддитивны и представляют собой последовательность независимых одинаково распределенных случайных величин с нулевым математическим ожиданием и ограниченной дисперсией:

$$M\{\xi_i\} = 0; \quad M\{\xi_i^2\} = \sigma^2; \quad M\{\xi_i \xi_j\} = 0 \quad \forall i \neq j,$$

где $M\{\cdot\}$ — знак математического ожидания; $\sigma^2 < \infty$ — дисперсия. Хотя только при таких предположениях полученные оценки

несмещенные и эффективные [76], в явном виде эти предположения почти никогда не могут быть проверены на практике — можно лишь обнаружить их нарушение, оценив остатки $D_i = s_i - C_{\text{орт}}^T \varphi(x_i)$, которые должны иметь нормальное или близкое к нему распределение [31].

Один из способов улучшения оценки параметров — проведение активного эксперимента по оптимальному плану [5, 34]. Однако на практике конструктор ИТ часто лишен такой возможности и должен довольствоваться лишь имеющимися данными пассивных наблюдений.

В таких ситуациях классический регрессионный анализ, основанный на минимизации квадратичной функции потерь (1.7), может давать неустойчивые решения при плохой обусловленности матрицы F_W и резко выделяющимся наблюдениям (выбросам), которые ведут к несимметричности распределения остатков $D_i = s_i - C_{\text{орт}}^T \varphi(x_i)$, а значит, к нарушению условий независимости дисперсии и математического ожидания. Иногда обработка таких данных на основе МНК приводит к ошибкам настолько большим, что полученная модель не имеет практического смысла [309, 357]. Для разрешения этой проблемы применяют различные приемы регуляризации [191], алгоритмы робастного оценивания [152, 230] и другие методы, основанные на минимизации функций потерь, отличающихся от (1.7) [107, 332].

Мы коснулись лишь нескольких наиболее важных проблем, возникающих при решении задачи *параметрической* идентификации в предположении, что структура модели задана в виде (1.6) и известны базисные функции $\varphi_1(x), \dots, \varphi_K(x)$. Гораздо более серьезные проблемы возникают при решении задачи *структурной* идентификации.

Поскольку в общей постановке (при отсутствии априорной информации об истинной структуре модели $s = s(x_1, \dots, x_N)$) разрешить эту проблему невозможно, конструктору прикладной ИТ приходится выбирать один из двух вариантов:

- отказаться от решения задач, для которых истинная структура взаимосвязи s и $x_1, \dots, x_N$ не известна и не может быть получена на основе физических представлений или результатов имитационного моделирования;
- попытаться решить задачу, сведя общую проблему идентификации неизвестной модели (1.4) к упрощенной постановке,

предполагающей перебор различных структур-претендентов $f \in F$ некоторого множества F и оценку качества полученного решения по имеющимся экспериментальным данным с точки зрения некоторого критерия $R(f)$.

Рассмотрим кратко известные к настоящему времени подходы к реализации такой упрощенной постановки задачи структурно-параметрической идентификации.

Прикладной регрессионный анализ, оставляя за пределами рассмотрения проблему построения самого исходного набора базисных функций $\varphi_1(x), \dots, \varphi_K(x)$, предлагает методы формирования множества F на основе генераторов полного или направленного перебора моделей различной сложности [172]. Под сложностью понимается число $K_0 \leq K$ ненулевых компонент линейных моделей (1.6), построенных на основе заданного набора функций $\varphi_1(x), \dots, \varphi_K(x)$.

Для сравнения моделей-претендентов, генерируемых в процессе перебора, используют различные критерии, в частности критерии Фишера [172] и Мэллоуза [322], основанные на анализе величины

$$\sigma^2(K_0) = \frac{1}{W - K_0} \sum_{i=1}^W [s_i - \sum_{j=1}^{K_0} c_{\text{opt}}(j) \varphi_j(x)]^2,$$

представляющей собой остаточную дисперсию оценки s по модели сложности K_0 с оптимальными значениями параметров $c_{\text{opt}}(j)$, информационный критерий Акаике [245], критерий финальной ошибки прогноза [246] и ряд других критериев [185].

Поскольку при больших значениях K полный перебор $2^K - 1$ моделей-претендентов, реализуемый так называемым методом всех регрессий, связан с вычислительными проблемами, часто на практике используют упрощенные варианты перебора [172], в том числе методы:

- *включения* (добавление регрессоров по одному, максимально уменьшающему критерий);
- *исключения* (процедура начинается с полной модели, а затем число регрессоров уменьшают по одному, пока критерий убывает);
- *шаговой регрессии* (комбинация двух предыдущих);
- *ветвей и границ*.

Оригинальным методом структурно-параметрической идентификации моделей (в указанной выше упрощенной постановке) является метод группового учета аргументов (МГУА), основные принципы которого сформулировал А.Г. Ивахненко в 1968 г. [97]. Алгоритмы самоорганизации моделей на основе МГУА достаточно подробно описаны в монографиях [95, 96, 98–100] и в многочисленных научных статьях [301, 330, 360, 362, 378].

Главное отличие МГУА от классических методов прикладного регрессионного анализа — это разделение имеющихся наблюдений на две выборки: обучающую, по которой производится оценка параметров моделей-претендентов, и проверочную, обеспечивающую селекцию этих моделей по внешнему критерию. Тем самым при селекции моделей привлекается дополнительная информация (не участвующая в оценке параметров моделей-претендентов), что обеспечивает нужный компромисс между сложностью модели и ее точностью [95].

В качестве обобщенной структуры, аппроксимирующей функцию (1.4), алгоритмы МГУА чаще всего рассматривают конечный алгебраический или тригонометрический многочлен, на основе которого генерируются частные модели-претенденты различной степени сложности. Выбор наилучшей (с точки зрения внешнего критерия) структуры реализуется с помощью комбинаторных алгоритмов [184], осуществляющих полный перебор, или итерационных процедур [241]. Получили развитие также многоуровневые алгоритмы МГУА, в которых построение сложной аппроксимирующей функции, например полинома Колмогорова—Габора [289]:

$$\hat{s} = c_0 + \sum_{i=1}^m c_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m c_{ijk} x_i x_j x_k + \dots,$$

достигается последовательным использованием частных зависимостей

$$u_1 = u(x_1, x_2); \quad u_2 = u(x_1, x_3); \dots; \quad u_s = u(x_{m-1}, x_m)$$

между $M = C_m^2$ парами входных данных. В качестве функции $u(\cdot)$ обычно используются простейшие соотношения, например

$$u(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j$$

или

$$u(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_5 x_i^2 + a_6 x_j^2,$$

коэффициенты которых определяются по обучающей выборке.

Среди всех частных моделей выбирается $M^* < M$ моделей, которые с точки зрения внешнего критерия показали наилучшие результаты на проверочной выборке. Значения $u_1, \dots, u_{M^*}$, полученные на обучающей выборке, используются в качестве аргументов частных моделей второго уровня

$$v_1 = u(u_1, u_2); \quad v_2 = u(u_1, u_3); \dots; \quad v_s = u(u_{s-1}, u_s),$$

селекция которых производится аналогично, и т. д., пока не будет достигнут компромисс между сложностью и точностью синтезируемой модели [42].

Гипотеза о вероятностной природе ошибок эксперимента, на которой основан аппарат прикладного регрессионного анализа, как, впрочем, и алгоритмы МГУА, не всегда соответствует действительности. Этим объясняется повышенный интерес к методам идентификации, которые свободны от этого предположения и рассматривают неопределенность величин с позиций аппаратов интервально-множественного анализа и нечеткой логики [16, 77, 88, 114, 127, 128, 146, 159—161, 237].

Даже далеко неполный обзор существующих подходов показывает, что современная наука располагает достаточно широким набором методов, которые подтвердили свою эффективность при решении прикладных задач, связанных со структурно-параметрической идентификацией в условиях априорной неопределенности. Безусловно, эти методы могут и должны рассматриваться как важные компоненты синтеза ИТ обработки сигналов.

1.3. Методы построения диагностических правил

Диагностику состояния технических и биологических объектов довольно часто рассматривают с позиции теории распознавания образов [28, 105, 164], предполагая существование некоторой совокупности характеристик (свойств, признаков) $x^{(N)} = (x_1, \dots, x_N)$, измерение которых позволяет отличать одно состояние объекта от другого. В этом случае требуется построить решающее правило $s = D(x^{(N)})$, которое отображает множество $X^{(N)}$ возможных значений признаков $x_1, \dots, x_N$ на множество $\{1, \dots, M\}$ номеров возможных состояний (диагнозов) $V_1, \dots, V_M$. Иными словами, требуется построить алгоритм определения значений индикаторной переменной s в виде

$$s = \begin{cases} 1, & \text{если } x^{(N)} \in \Omega_1, \\ \dots & \\ M, & \text{если } x^{(N)} \in \Omega_M, \end{cases} \quad (1.8)$$

где $\Omega_m, m = 1, \dots, M$ — непересекающиеся области пространства X , в которых принимаются решения в пользу класса V_m .

Математические методы распознавания можно разделить на два класса: детерминированные, которые в явном виде не используют статистические свойства наблюдаемых образов, и статистические, основанные на методах теории вероятностей и математической статистики [194, с. 31]. Статистическая постановка задач распознавания [57, 227] возникает в тех практически важных случаях, когда при различных состояниях объекта $V_i \neq V_j$ могут совпадать значения каждого из имеющихся признаков x_n ($1 \leq n \leq N$), иными словами, допускается, что множества значений каждого из признаков, соответствующие различным классам, пересекаются.

При статистической постановке задачи состояния $V_1, \dots, V_M$ объекта рассматриваются как случайные события с априорными вероятностями $P(V_m)$, $\sum_{m=1}^M P(V_m) = 1$, признаки — как случайные величины, для которых объективно существуют условные распределения $p(x^{(N)} / V_m)$, а множества

$$X_m^{(N)} = \{x^{(N)} : p(x^{(N)} / V_m) \neq 0\} \quad X_1^{(N)} \cup \dots \cup X_M^{(N)} = X^{(N)}$$

именуются собственными областями классов в пространстве признаков [196, с. 31]. Интегрирование $p(x^{(N)} / V_m)$ по множествам возможных значений всех признаков $x_1, \dots, x_N$, кроме признака x_n ($1 \leq n \leq N$), позволяет получить представление об одномерных (маргинальных) распределениях этого признака $p(x_n / V_1), \dots, p(x_n / V_M)$, причем в рассматриваемой постановке задачи допускается, что для всех $n = 1, \dots, N$ носители маргинальных распределений различных классов заведомо пересекаются.

Часто в задачах диагностики рассматриваются только два возможных диагноза — «исправен—неисправен», «здоров—бо-

лен» [28, 40, 105, 164]. Поэтому для упрощения дальнейшего изложения ограничимся случаем дихотомии, когда $M = 2$, тем более что к дихотомии можно свести и общий случай $M \geq 2$, последовательно решая задачу распознавания одного класса и всех остальных [233, с. 94].

Если потерей $L(i, j)$, $i, j = 1, 2$, оценивать диагноз в пользу i -го класса, в то время как объект находится в j -м состоянии, то при известном отношении правдоподобия $\lambda(x^{(N)}) = p(x^{(N)} / V_1) / p(x^{(N)} / V_2)$ оптимальные решения, обеспечивающие минимум средних потерь (риска)

$$R = \int_{X^{(N)}} \sum_{j=1}^2 L(Z, j) P(V_j) p(x^{(N)} / V_j) dx^{(N)},$$

должны приниматься в соответствии с байесовым правилом [70]:

$$s = D(x^{(N)}) = \begin{cases} 1, & \text{если } \lambda(x^{(N)}) > \theta_{12}, \\ 2, & \text{если } \lambda(x^{(N)}) < \theta_{12}, \end{cases} \quad (1.9)$$

где

$$\theta_{12} = \frac{P(V_2)(L_{21} - L_{22})}{P(V_1)(L_{12} - L_{11})}.$$

Известно также [106], что в частном случае, когда потери от правильных решений равны нулю, а от ошибочных — единице, т. е. $L_{11} = L_{22} = 0$ и $L_{12} = L_{21} = 1$, то оптимальное правило (1.9) эквивалентно правилу максимума апостериорных вероятностей

$$s = D(x^{(N)}) = \begin{cases} 1, & \text{если } P(V_1 / x^{(N)}) > P(V_2 / x^{(N)}), \\ 2, & \text{если } P(V_1 / x^{(N)}) < P(V_2 / x^{(N)}), \end{cases} \quad (1.10)$$

где

$$P(V_m / x^{(N)}) = \frac{P(V_m) p(x_m / V_m)}{\sum_{m=1}^M P(V_m) p(x_m / V_m)}, \quad m = 1, 2,$$

причем это правило гарантирует минимум средней вероятности ошибочных решений на множестве $X^{(N)}$.

К сожалению, в большинстве практических случаев вероятностные характеристики, фигурирующие в оптимальных решаю-

щих правилах (1.9), (1.10), заранее не известны. Максимум, на что можно рассчитывать при решении прикладных задач, — получение оценок априорных вероятностей $\hat{P}(V_m)$, например знания распространенности (преваленса) исследуемого заболевания [53], одномерных эмпирических распределений (гистограмм) $\hat{p}(x_n/V_1)$, $\hat{p}(x_n/V_2)$, построенных для отдельных признаков [52], и в лучшем случае — коэффициентов корреляций между отдельными признаками.

Однако в общем случае такие ограниченные знания недостаточны для восстановления многомерных распределений $p(x^{(N)}/V_m)$. Следовательно, неизвестна и оптимальная разделяющая (дискриминантная) функция

$$f(x^{(N)}) = P(V_1)p(x^{(N)}/V_1) - P(V_2)p(x^{(N)}/V_2), \quad (1.11)$$

знак которой позволяет распознавать классы с минимальной средней вероятностью ошибочных решений.

Все разнообразие научных направлений в современной теории распознавания образов заключается в формализации методов построения решающего правила $s = D(x^{(N)})$ по выборке наблюдений $x_1^{(N)}, x_2^{(N)}, x_3^{(N)}, \dots$ с известной (обучение) или неизвестной (самообучение) принадлежностью к классам $V_1, \dots, V_M$ [239].

Рассмотрим кратко эти направления, останавливаясь лишь на деталях, которые нам понадобятся в дальнейшем.

Принципиально существуют два пути построения разделяющей функции на основе лишь той информации, которая содержится в обучающей выборке $x_1^{(N)}, x_2^{(N)}, x_3^{(N)}, \dots$ с известной принадлежностью к классам V_1 и V_2 . Первый (прямой) путь — восстановление многомерных распределений $p(x^{(N)}/V_m)$ по имеющейся выборке. Однако эффективные методы решения этой задачи разработаны лишь для случая, когда функции $p(x^{(N)}/V_m)$ заданы с точностью до небольшого числа неизвестных параметров [78, 227]. Как отмечается в [39, с. 49], реальные успехи на этом пути возможны лишь для вырожденных случаев, предполагая, например, что распределение $p(x^{(N)}, \alpha)$, зависящее от вектора α неизвестных параметров, удовлетворяет условию

$$p(x^{(N)}, a) = p(x_1, a_1) \times \dots \times p(x_N, a_N),$$

и, кроме того, каждый признак x_n может принимать лишь фиксированное число значений, либо плотность распределения вероятностей для каждого класса задана нормальным законом

$$p(x^{(N)}, \mu, \Delta) = \frac{1}{(2\pi)^{N/2} |\Delta|^{1/2}} \exp \left[-\frac{1}{2} (x^{(N)} - \mu)^T \Delta^{-1} (x^{(N)} - \mu) \right], \quad (1.12)$$

когда достаточно по выборке наблюдений оценить лишь вектор средних μ и ковариационную матрицу Δ для каждого класса.

Второй, более практичный путь решения этой задачи, не требующий восстановления упомянутых многомерных распределений, основан на аппроксимации (1.11) другой функцией, заданной с точностью до параметров [233, 234, 340, 341]. Тем более, что если уж ограничиться случаем нормальных распределений (1.12), то, как известно [129, 141], при произвольных, но одинаковых в обоих классах ковариационных матрицах $\Delta_1 = \Delta_2$ оптимальная дискриминантная функция линейная, а при $\Delta_1 \neq \Delta_2$ — квадратичная.

Если же гипотеза о нормальности распределений неправомерна, то построение разделяющей функции сводится к решению задачи, аналогичной той, которая рассматривалась в параграфе 1.2. Иными словами, приходится аппроксимировать (1.11) взвешенной суммой известных функций

$$\tilde{f}(x^{(N)}) = \alpha^T \varphi(x^{(N)}) \equiv \alpha_1 \varphi_1(x^{(N)}) + \dots + \alpha_K \varphi_K(x^{(N)}) \quad (1.13)$$

и оценивать параметры $\alpha_1, \dots, \alpha_K$ этого разложения по имеющейся обучающей выборке наблюдений $x_1^{(N)}, x_2^{(N)}, x_3^{(N)}, \dots$ [340].

Основная идея методов обучения, предложенных в [3, 4], основана на построении в каждой i -й точке обучающей последовательности $x_i^{(N)}$ потенциальных функций $U(x^{(N)}, x_i^{(N)})$, заданных на всем пространстве X и зависящих от $x_i^{(N)}$, как от параметра. Если собственные области классов $X_1^{(N)}$ и $X_2^{(N)}$ в пространстве X удовлетворяют так называемым условиям компактности [3], то суммирование функций $U(x^{(N)}, x_i^{(N)})$ отдельно для наблюдений первого и второго классов позволяет построить две функ-

ции

$$U_1(x^{(N)}) = \sum_{x_i^{(N)} \in V_1} U(x^{(N)}, x_i^{(N)}); \quad U_2(x^{(N)}) = \sum_{x_i^{(N)} \in V_2} U(x^{(N)}, x_i^{(N)})$$

и использовать их в решающем правиле вида

$$s = \tilde{D}(x^{(N)}) = \begin{cases} 1, & \text{если } U_1(x^{(N)}) > U_2(x^{(N)}), \\ 2, & \text{если } U_1(x^{(N)}) < U_2(x^{(N)}). \end{cases}$$

Разновидности алгоритмов, реализующих метод потенциальных функций, отличаются способом выбора параметров g_i, r_i базовой рекуррентной процедуры

$$\tilde{f}_{i+1}(x^{(N)}) = g_i \tilde{f}_i(x^{(N)}) + r_i U(x_i^{(N)}, x^{(N)}),$$

обеспечивающей последовательное приближение неизвестной разделяющей функции $f(\cdot)$.

В работе [233, с. 331] показано, что ряд алгоритмов метода потенциальных функций эквивалентны процедуре стохастической аппроксимации Роббинса-Монро [354], которую можно непосредственно использовать для приближения оптимальной разделяющей функции (1.11). Как следует из [234, с. 90], условие минимума квадратичного функционала

$$J(\alpha) = \int_{X^{(N)}} [f(x^{(N)}) - f(x^{(N)}, \alpha)]^2 dx^{(N)},$$

характеризующего степень уклонения функций (1.11) и (1.13), можно свести к условию

$$M\{\delta - \alpha^T \varphi(x^{(N)}) \varphi(x^{(N)})\} = 0, \quad (1.14)$$

где M — знак математического ожидания, а δ — случайная функция (указания учителя) вида ¹

$$\delta = \begin{cases} +1, & \text{если } x^{(N)} \in V_1, \\ -1, & \text{если } x^{(N)} \in V_2. \end{cases} \quad (1.15)$$

¹ Поскольку в общем случае допускается, что собственные области классов пересекаются, т. е. $X_1^{(N)} \cap X_2^{(N)} \neq \emptyset$, то при одном и том же значении $x^{(N)} \in X_1^{(N)} \cap X_2^{(N)}$ функция δ может принимать различные значения в соответствии с (1.15).

Согласно [235] оптимальное значение вектора параметров α_{opt} , обеспечивающего минимум функционала (1.14), может быть получено с помощью итеративной процедуры

$$\alpha_i = \alpha_{i-1} + \Gamma_i [\delta_i - \alpha_{i-1}^T \varphi(x_i^{(N)}) \varphi(x_i^{(N)})], \quad (1.16)$$

где Γ_i — диагональная матрица, элементы которой удовлетворяют известным условиям Дворецкого [274].

Вполне понятно, что при решении практических задач для построения решающего правила может быть использована лишь конечная обучающая выборка $x_1^{(N)}, \dots, x_W^{(N)}$ с указаниями учителя $\delta_1, \dots, \delta_W$. Поскольку процедура стохастической аппроксимации асимптотически оптимальна, т. е. позволяет найти оптимальное значение α_{opt} лишь при $i \rightarrow \infty$, совершенно очевидно, что значение $\alpha = \alpha_W$, найденное на W -м шаге с помощью процедуры (1.16), вообще говоря, может отличаться от оптимального значения $\alpha_{\text{opt}}(W)$, при котором оценка

$$\hat{J}(\alpha, W) = \frac{1}{W} \sum_{i=1}^W [\delta_i - \alpha^T \varphi(x_i^{(N)})]^2 \quad (1.17)$$

функционала (1.14) по конечному числу W наблюдений будет минимальной [122].

Для разрешения этой проблемы существует несколько возможностей. Одна из них состоит в циклическом повторении имеющихся в распоряжении данных [123], что требует запоминания сразу всей имеющейся последовательности и, как справедливо замечено в [39, с. 85], сразу же лишает итеративную процедуру (1.16) ее основного смысла.

Вторая возможность — применение вместо (1.16) так называемых локально-оптимальных процедур стохастической аппроксимации, которые, однако, требуют значительных вычислений, связанных с обращением на каждом i -м шаге итерации матрицы Калмана [234, с. 40]:

$$\Gamma_i^{\text{opt}} = \left[\sum_{j=1}^i \varphi(x_j^{(N)}) \varphi^T(x_j^{(N)}) \right]^{-1}. \quad (1.18)$$

Поэтому более практичным методом определения оптимального значения $\alpha_{\text{opt}}(W)$, доставляющего минимум функционалу (1.17),

является подход, рассмотренный нами в работе [220], который непосредственно опирается на метод наименьших квадратов и позволяет получить выражение для определения $\alpha_{\text{opt}}(W)$ в явном виде:

$$\alpha_{\text{opt}}(W) = \left[\sum_{i=1}^W \varphi(x_i^{(N)}) \varphi^T(x_i^{(N)}) \right]^{-1} \sum_{i=1}^W \delta_i \varphi(x_i^{(N)}). \quad (1.19)$$

Заметим, что разделяющая функция $\alpha_{\text{opt}}^T \varphi(x^{(N)})$, построенная из условия минимума (1.17), не всегда позволяет разделить множества $X_1^{(N)}, X_2^{(N)}$, даже если их можно полностью разделить функцией вида $\alpha^T \varphi(x^{(N)})$ [227, 233]. В связи с этим, несмотря на удобство квадратичного функционала (1.17) как универсального средства решения оптимизационной задачи приближения функций, на практике часто приходится использовать алгоритмы обучения, в основу которых положены иные критерии.

В последнее время также популярны методы распознавания на основе обучаемых нейронных сетей [351], нечетких правил [175, 258, 288] и даже так называемые генетические алгоритмы [253].

Разумеется, эффективность построения диагностического правила существенным образом зависит от объема W обучающей выборки. При этом до сих пор не существует методов, позволяющих в условиях ограниченной априорной информации обосновать выбор базисных функций $\varphi_1(x^{(N)}), \dots, \varphi_K(x^{(N)})$, фигурирующих в (1.13), и, как справедливо отмечено в [3, с. 57], на практике чаще всего приходится осуществлять такой выбор, опираясь лишь на интуицию, опыт и некоторые неформальные соображения.

Известно также, что расширение класса базисных функций неизбежно требует существенного увеличения объема W обучающей выборки [39, с. 95]. Поэтому при малом объеме данных часто не остается ничего иного, как ограничиться классом линейных функций $\tilde{f}(x^{(N)}) = \alpha_N x_N + \dots + \alpha_1 x_1 + \alpha_0$ [83, 170, 228, 342], если, конечно, нет веских оснований для выбора более «сложных» дискриминантных функций [26].

В последнее время внимание специалистов привлекают так называемые коллективные (комбинированные) классификаторы [18, 165, 335], суть которых состоит в интеграции частных реше-

ний, принимаемых отдельными классификаторами. Приобрели известность различные подходы к интеграции частных решений.

В одних случаях предлагается использовать метод голосования (*majority vote method*) [313, 350] или ранжирования (*label ranking method*) [251, 302], в других — схемы, основанные на усреднении или линейной комбинации апостериорных вероятностей, которые оцениваются отдельными классификаторами [299, 385], либо алгоритмы нечетких правил (*fuzzy rules*) [265]. Предлагается также проводить независимое обучение комбинированного классификатора на нейронных сетях, рассматривая частные решения как комплексные признаки [315, 383]. Развиваются также подходы, основанные на выделении в пространстве наблюдений локальных областей, в каждой из которых только один из частных классификаторов «компетентен» принимать решение [165, 384].

Все эти работы представляют несомненный теоретический интерес и позволяют обосновать выбор той или иной схемы интеграции, когда известны формальные алгоритмы, положенные в основу частных классификаторов.

Однако при построении ИТ оценки состояния объектов часто требуется сформировать коллективное решение, основанное на интеграции частных решений отдельных эвристических алгоритмов. Разумеется, и в этих практически важных случаях требуется обоснованный подход к интеграции частных решений. Например, какое окончательное решение должно быть принято, если в результате независимого обследования одна часть специалистов (экспертов) признала пациента здоровым, а другая — больным?

Оригинальные результаты исследования этих вопросов будут представлены в главе 5.

1.4. Анализ свойств сигналов как метод формирования диагностических признаков

Выбор адекватного множества диагностических признаков представляет собой одну из наиболее трудных проблем, с которой приходится сталкиваться конструктору ИТ при решении задач медицинской и технической диагностики [28, 52, 105]. Универсальных методов разрешения этой проблемы не существует, поскольку каждый объект, а значит, и порождаемый им сигнал уникальны.

В условиях ограниченной априорной информации, когда заранее не известен набор так называемых физических признаков

[194, с. 263], общая проблема формирования пространства признаков разрешается в два этапа. На первом этапе в исходное пространство признаков включается любое свойство сигнала, которое предположительно несет информацию о состоянии объекта. И только на втором этапе решается задача минимизации (селекции) признаков с использованием тех или иных формальных приемов.

В теории цифровой обработки сигналов разработано огромное количество универсальных методов анализа свойств сигналов [63, 87, 115, 144, 145, 154, 174]. При ограниченной априорной информации эти методы могут быть рассмотрены как инструмент формирования исходного пространства так называемых математических признаков [194, с. 263].

Поскольку в рамках ограниченного обзора невозможно перечислить известные подходы к анализу свойств сигналов различной физической природы, дадим лишь краткую характеристику некоторым из них, остановившись на деталях, которые нам понадобятся в дальнейших исследованиях.

1.4.1. Анализ гармонических компонент сигнала

Довольно часто ценная информация, которую сложно получить при амплитудно-временном представлении сигнала, может быть получена при его представлении в частотной области [21, 28, 103, 144, 260]. Для такого перехода используются методы спектрального анализа, позволяющие оценить частоты, амплитуды и начальные фазы гармонических составляющих сигнала.

В основе традиционного спектрального анализа лежит теория преобразования Фурье, история которой началась еще в 1822 г., когда французский математик Жан Батист Жозеф Фурье заметил, что практически любая периодическая функция $y(t)$ может быть разложена в тригонометрический ряд [167]:

$$y(t) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{2\pi t}{T} n + b_n \sin \frac{2\pi t}{T} n \right),$$

где $T = \frac{2\pi}{\omega} = \frac{1}{f}$ — период функции; ω — круговая частота; f — частота Гц. На основании известной формулы Эйлера ряд Фурье может быть представлен в комплексной форме [116]:

$$y(t) = \sum_{n=-\infty}^{\infty} Y_n e^{i(2\pi n f)t},$$

причем комплексные числа Y_n (комплексные амплитуды гармонических составляющих) вычисляются по формуле

$$Y_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} y(t) e^{-i(2\pi n f)t} dt.$$

Таким образом, частотный спектр периодического сигнала состоит из частот $f_n = nf$, $n = 0, 1, 2, \dots$, кратных его основной (базовой) частоте f , причем, если комплексную амплитуду Y_n

представить в экспоненциальной форме $Y_n = \frac{C_n}{2} e^{i\varphi_n}$, то величина C_n представляет собой амплитуду гармонической составляющей с частотой $f_n = nf$, а φ_n — начальную фазу этой гармоники, имеющей форму косинусоиды $C_n \cos(2\pi n f t + \varphi_n)$.

Много лет спустя сформулированное Ж. Фурье свойство периодических функций было обобщено на непериодические функции. Именно это обобщение и принято называть преобразованием Фурье [167], которое трактуется как взаимно однозначный переход от некоторой функции $y(t)$ действительного аргумента t (не обязательно времени) к другой функции $Y(f)$, аргумент которой $f = 1/t$. В частном случае, когда t — время, преобразование Фурье дает частотное представление сигнала $y(t)$. Формально такое преобразование определяется соотношениями

$$Y(f) = \Phi[y(t)] = \int_{-\infty}^{+\infty} y(t) e^{-i(2\pi f)t} dt, \quad y(t) = \Phi^{-1}[Y(f)] = \int_{-\infty}^{+\infty} Y(f) e^{i(2\pi f)t} df,$$

которые справедливы, если функция $y(t)$ ограничена, кусочно-непрерывна и имеет конечное число экстремумов (условия Дирихле) [352]. Эти соотношения принято называть соответственно прямым и обратным преобразованиями Фурье [116].

Дальнейшее развитие теории преобразований Фурье заключается в его обобщении на важный в практическом плане случай

дискретной функции [87] $y(t) \rightarrow y(t_k)$, представленной конечной последовательностью значений $y_k \equiv y(t_k)$ в равноотстоящие моменты времени $t_k \equiv k\Delta$, $k = 0, 1, \dots, K-1$, где Δ — шаг квантования по времени. Прямое и обратное дискретное преобразование Фурье (ДПФ) реализуется на основе соотношений [283]:

$$Y_n = \sum_{k=0}^{K-1} y_k e^{-2\pi i n k / K}, \quad (1.20)$$

$$y_k = \frac{1}{K} \sum_{n=0}^{K-1} Y_n e^{2\pi i n k / K}. \quad (1.21)$$

Как видно из (1.20), прямое ДПФ позволяет разложить (декомпозировать) сигнал y_k , зафиксированный с шагом квантования Δ в K точках ограниченного временного интервала $K\Delta$ (в секундах), на K гармонических компонент [116] с частотами (в герцах):

$$f_n = \frac{n}{K\Delta}, \quad n = -K/2, \dots, 0, \dots, K/2. \quad (1.22)$$

При этом выполняется условие симметрии комплексных амплитуд гармонических компонент, т. е. $Y_{-n} = Y_{K-n} \quad \forall n = 1, \dots, K/2$.

Переломным этапом в развитии преобразований Фурье стала работа [269], в которой впервые был предложен быстрый алгоритм вычисления ДПФ, впоследствии названный алгоритмом БПФ — быстрого преобразования Фурье (**FFT** — *Fast Fourier Transformation*).

Главная идея БПФ основана на применении рекурсивной процедуры вычисления дискретного преобразования Фурье K -точечного сигнала суммой двух ДПФ, вычисленных отдельно для четных и нечетных точек. Такая процедура позволяет уменьшить число вычислительных операций с $2K^2$ до $2K \log_2 K$, если K — целая степень 2. Поскольку процедура преобразования Фурье линейна, легко показать, что алгоритм БПФ можно использовать при любом числе точек, доопределив нулями недостающие члены временного ряда до ближайшего K , кратного двум.

В настоящее время известно множество модификаций алгоритмов БПФ [262, 263], которые применяют для обработки сигналов различного физического происхождения.

Следует, однако, заметить, что практическое использование алгоритмов **БПФ** связано с рядом известных ограничений. Как будет показано в главе 4, при обработке элементарного гармонического сигнала $y_k = A \sin 2\pi f_c k$, $k = 1, 2, \dots, K$, заданного конечным числом K дискретных значений, алгоритм позволяет определить амплитуду A и частоту f_c лишь в том случае, когда f_c в точности совпадает с одной из частот (1.22).

Поскольку на практике такой случай является скорее редким исключением, чем правилом, указанная особенность алгоритмов **БПФ**, вообще говоря, ограничивает область их непосредственного применения для задач выявления гармонических компонент в обрабатываемом сигнале, в частности гармонических помех.

Именно поэтому в ряде работ для задач гармонического анализа предлагается использовать иные подходы, например гармонический алгоритм **МГУА** [62, 240]. Однако эти алгоритмы обладают существенно меньшим быстродействием по сравнению с алгоритмами **БПФ**. К тому же, как показали исследования, проведенные в работе [363], гармонический алгоритм **МГУА** позволяет достаточно точно оценить частотные компоненты дискретного сигнала y_k , $k = 1, \dots, K$ лишь при весьма малых значениях K и невысоком уровне помех.

Второе известное ограничение, вытекающее из самой сути преобразования Фурье, заключается в том, что функция, полученная в результате этого преобразования, зависит от частоты, но не зависит от времени. Отсюда следует, что построенные на основании **БПФ** спектры стационарного полигармонического сигнала $y_1(t) = \sum_{i=1}^M A_i \cos(2\pi f_i t + \varphi_i)$ в виде суммы M различных гармоник, присутствующих на всем интервале наблюдения $0 < t < t_0$, и нестационарного гармонического сигнала

$$y_2(t) = \begin{cases} A_1 \cos(2\pi f_1 t + \varphi_1), & t \in [0, t_1], \\ A_2 \cos(2\pi f_2 t + \varphi_2), & t \in (t_1, t_2], \\ \dots & \\ A_M \cos(2\pi f_M t + \varphi_M), & t \in (t_{M-1}, t_M], \end{cases} \quad (1.23)$$

который представляет собой последовательное «переключение» элементарных гармонических колебаний, окажутся практически неразличимы.

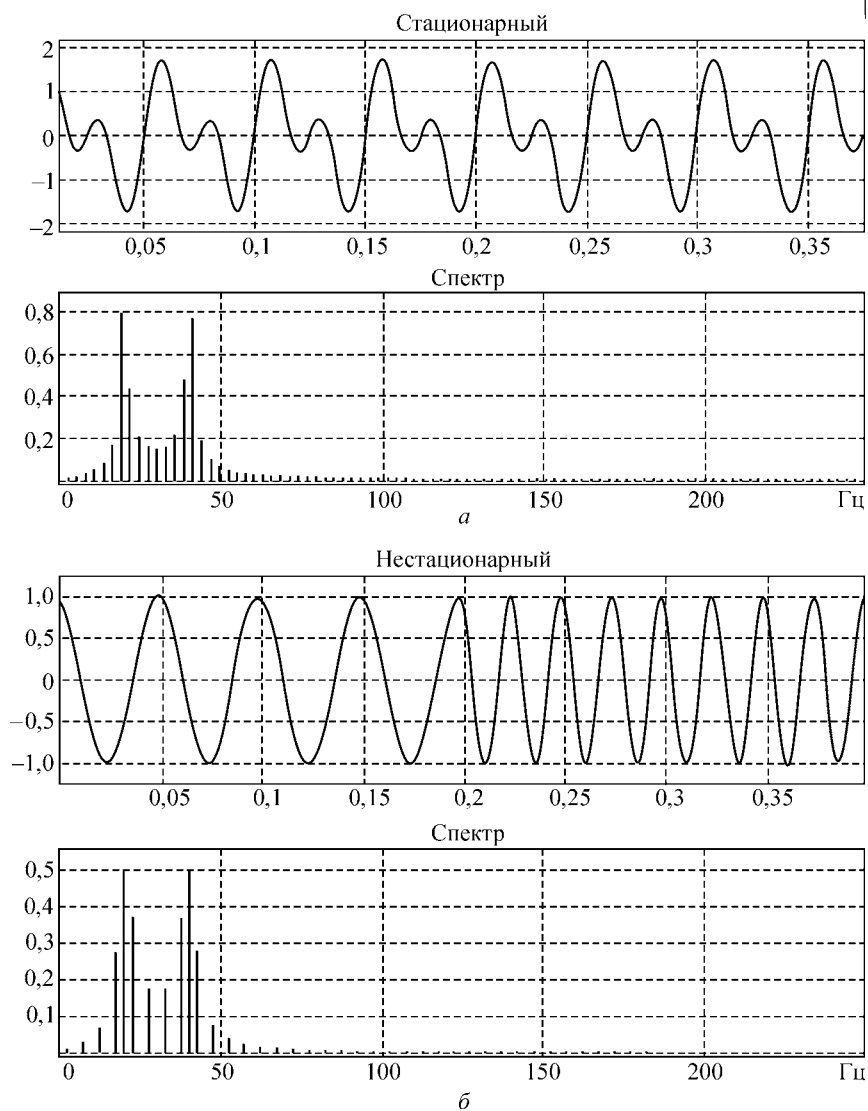


Рис. 1.3. Спектры стационарного (а) и нестационарного (б) сигналов с одинаковыми частотами и амплитудами

Для иллюстрации этого эффекта на рис. 1.3 представлены спектры стационарного гармонического сигнала, состоящего из суммы двух гармоник с частотами 20 и 40 Гц (а) и нестационар-

ного сигнала, в котором вначале присутствовала только гармоника с частотой 20 Гц, а затем — 40 Гц (б). Видно, что различия спектров этих сигналов незначительны.

Поскольку преобразование Фурье не позволяет локализовать во времени частотные компоненты, то это накладывает известное ограничение на область его практического применения только для анализа стационарных сигналов. В то же время многие реальные сигналы имеют сложные частотно-временные характеристики и, как правило, состоят из короткоживущих высокочастотных и долговременных низкочастотных компонент.

Для анализа таких сигналов необходимо обеспечить хорошее разрешение и по частоте, и по времени. Первое требуется для локализации низкочастотных составляющих, второе — для разрешения компонент высокой частоты.

Один из известных подходов к анализу гармонических компонент нестационарных сигналов, который был предложен в 1948 г. Денисом Габаром, получил название «локальное преобразование Фурье» (STFT — *Short-Time Fourier Transformation*) [248, 249, 271]. В соответствии с этим подходом нестационарный сигнал $y(t)$ предполагается стационарным для некоторых локальных областей времени (фреймов), в каждой из которых вычисляется преобразование Фурье обычным способом. Результатом преобразования является функция $\Phi_w(f, t)$, зависящая как от частоты, так и от времени:

$$\Phi_w(f, t) = \int [y(t)w(t - \tau)e^{-2\pi i f \tau}] d\tau, \quad (1.24)$$

где $w(\cdot)$ — так называемая функция окна. Чаще всего в качестве

$w(\cdot)$ используется гауссова функция $w(t) = e^{-\frac{at^2}{2}}$, в которой параметр a определяет ширину окна.

На первый взгляд может показаться, что на основе преобразования (1.24) можно получить представление о том, какие именно гармонические компоненты присутствуют в сигнале $y(t)$ в каждый момент времени. Однако в соответствии с расширенным толкованием фундаментального принципа неопределенности Гейзенберга [112] такое заключение неверно. Относительно рассматриваемой проблемы из этого принципа следует, что невозможно определить точное значение частотных компонент сигнала в фиксированный момент времени, а можно лишь получить представление об определенном диапазоне частоты f на некотором интервале времени t .

Отмеченный факт прямо следует из процедуры (1.24) для вычисления локального преобразования Фурье: при увеличении ширины окна улучшается разрешающая способность по частоте f , но теряется разрешение по времени t , а при уменьшении параметра a происходит обратный эффект.

Этот факт наглядно иллюстрирует рис. 1.4, на котором представлены результаты обработки одного и того же сигнала с различной шириной функции окна [348]. В качестве модельного использовался нестационарный гармонический сигнал (1.23) при $M = 4$ ($f_1 = 75$ Гц, $f_2 = 50$ Гц, $f_3 = 30$ Гц, $f_4 = 20$ Гц).

Видно, что при широкой функции окна ($a = 0,00001$) результат обработки имеет высокое разрешение по частоте, но низкое по времени (рис. 1.4, а), а при узкой функции окна ($a = 0,001$) — высокое разрешение по времени, но низкое по частоте (рис. 1.4, б).

Таким образом, при использовании **STFT** конструктор прикладной информационной технологии неизбежно сталкивается с проблемой выбора оптимального значения ширины окна, которая не может быть разрешена формальными методами.

Для преодоления этого недостатка в последние годы используется новый подход к анализу сигналов, названный вейвлет-преобразованием (**WT — Wavelet Transformation**) [305, 321, 361, 375, 377].

Этот подход предусматривает разложение нестационарного во времени сигнала $y(t)$ по базисным функциям, полученным сжатиями, растяжениями и сдвигами некоторого прототипа (так называемого материнского вейвлета). Согласно [377] адекватный решаемой задаче выбор материнского вейвлета дает хорошее разрешение и по времени, и по частоте.

В отличие от локального преобразования Фурье, **WT** позволяет проводить анализ сигнала со многими уровнями разрешения (**multiresolution analysis**), причем при больших значениях f метод обеспечивает высокую разрешающую способность по времени t , но низкую по частоте f , а при малых значениях f — высокую разрешающую способность по частоте, но низкую по времени.

Формально вейвлет-преобразование представляет собой функцию двух переменных

$$\Phi_{\Psi}(\tau, w) = \frac{1}{\sqrt{|w|}} \int_t y(t) \Psi\left(\frac{t - \tau}{w}\right) dt, \quad (1.25)$$

где $\Psi(\cdot)$ — материнский вейвлет; τ — сдвиг по времени; w — масштаб преобразования.

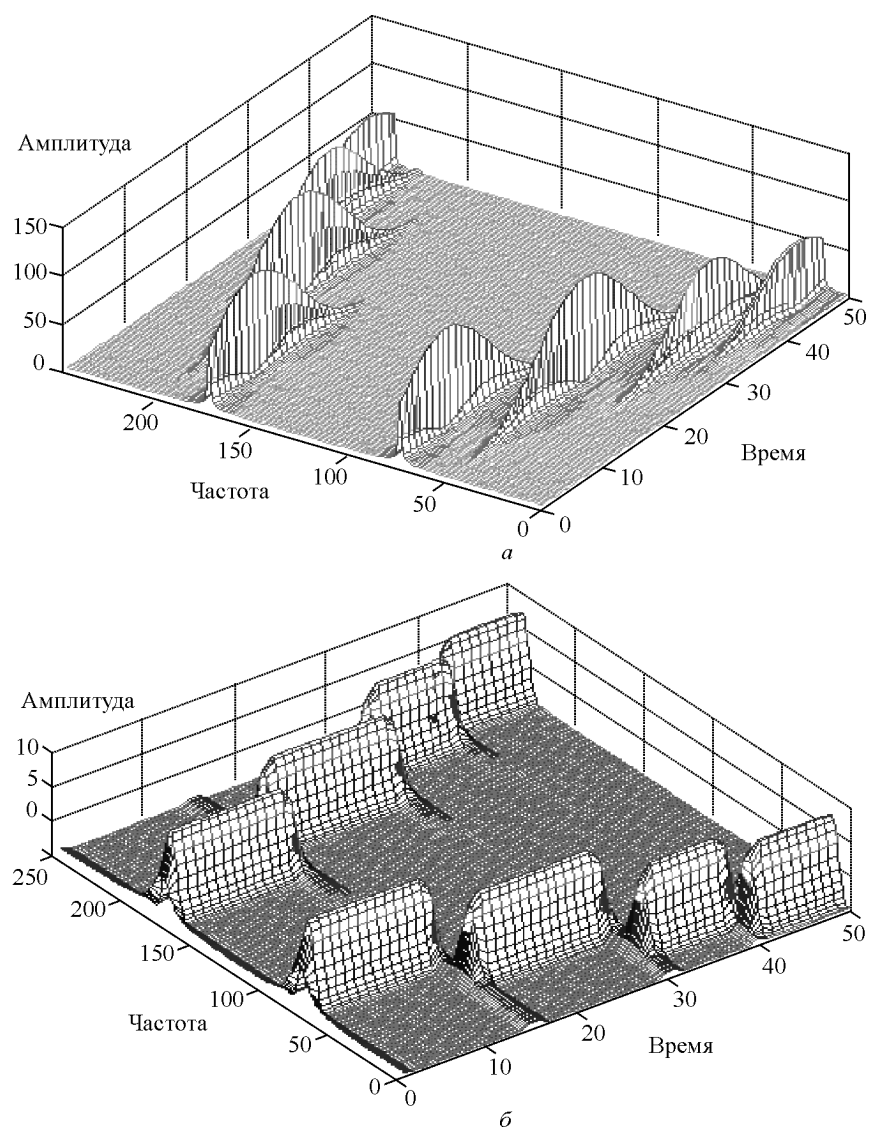


Рис. 1.4. Спектры нестационарного гармонического сигнала на основе STFT при широкой (а) и узкой (б) функции окна

К настоящему времени известны различные типы материнских вейвлетов $\Psi(\cdot)$ [344], однако на практике наибольшую популярность получили вейвлет-сомбреро (*Mexican hat*) в виде действительной функции

$$\Psi_1(t) = \frac{1}{\sqrt{2\pi w^3}} e^{\frac{-t^2}{2w^2}} \left(\frac{t^2}{w^2} - 1 \right),$$

представляющей собой вторую производную функции Гаусса, и вейвлет Морле (*Morlet wavelet*) в виде комплексной функции

$$\Psi_2(t) = e^{i\omega t} e^{\frac{-t^2}{2w}},$$

где ω — круговая частота; w — параметр масштаба.

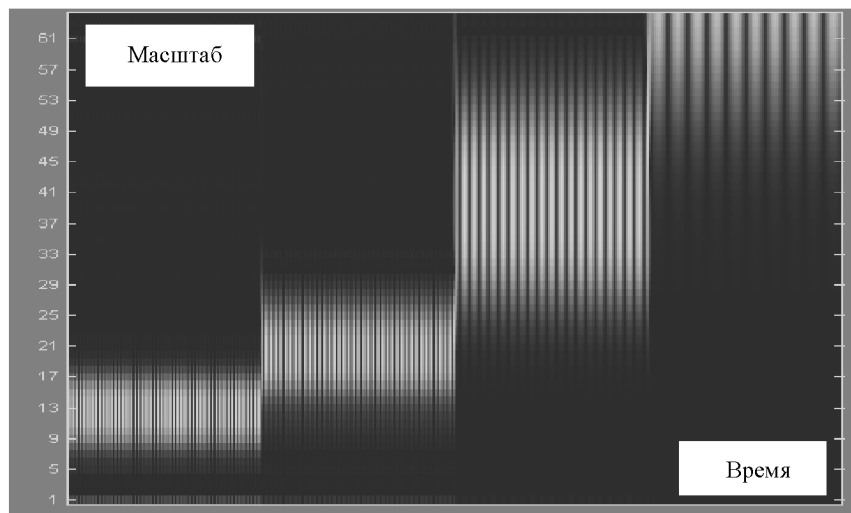
Параметр w имеет такой же смысл, как величина масштаба географической карты: большие значения w дают глобальное представление о низкочастотных компонентах $y(t)$, а малые значения детализируют высокочастотные компоненты $y(t)$.

Для дискретного сигнала y_k , представленного конечной последовательностью значений в равноотстоящие моменты времени $t_k \equiv k\Delta$, $k = 0, 1, \dots, K-1$, результат оценки непрерывного вейвлет-преобразования (CWT — *Continuous Wavelet Transform*) представляет собой матрицу коэффициентов $\|\Psi_{kj}\|$, вычисленных для различных моментов времени t_k при разных значениях масштаба w_j , $j = 1, \dots, W$, где W — число уровней детализации.

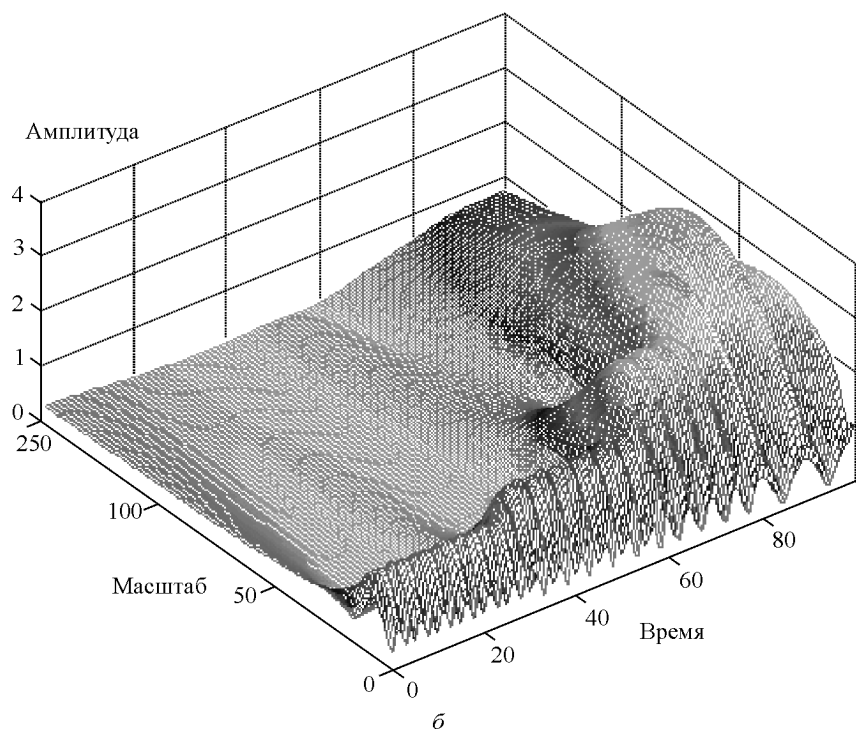
На рис. 1.5 показаны два способа визуализации результатов вычисления коэффициентов CWT для нестационарного гармонического сигнала типа (1.23) при $f_1 = 30$ Гц, $f_2 = 20$ Гц, $f_3 = 10$ Гц, $f_4 = 5$ Гц.

Практическое применение вейвлет-преобразования ограничено тем, что при больших значениях K и W вычисление матрицы $\|\Psi_{kj}\|$ занимает значительное время даже на современных компьютерах. Существенное ускорение процедуры обработки дает предложенный в работе [321] алгоритм, впоследствии названный «дискретное вейвлет-преобразование» (DWT — *Discrete Wavelet Transform*).

Главная идея DWT — декомпозиция обрабатываемого сигнала $y(t)$ на две составляющие:



a



б

Рис. 1.5. Графическое представление вейвлет-преобразования:
a — двумерное представление; *б* — трехмерное представление

$$y(\cdot) = a_1 + d_1,$$

где a_1 — приближение сигнала (*approximations*); d_1 — детали (*details*). При этом считается, что a_1 характеризует протяженные во времени низкочастотные компоненты, а d_1 — кратковременные высокочастотные компоненты.

Для декомпозиции применяются процедуры прореживания данных с последующей фильтрацией нижних и верхних частот. Их многократное продолжение к найденному приближению позволяет получить различные уровни детализации $y(t)$ в виде так называемого дерева **WT**.

Следует, однако, заметить, что при использовании вейвлет-преобразования конструктор прикладной ИТ сталкивается с непростыми проблемами, главными из которых являются:

а) выбор вида функции Ψ : хотя число стандартных вейвлет-прототипов достигает уже нескольких десятков, в ряде случаев приходится создавать свой собственный вейвлет, который обеспечит бы его адекватность обрабатываемому сигналу [142];

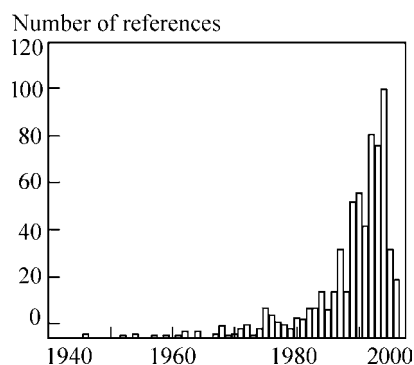
б) селекция оптимальной совокупности коэффициентов вейвлет-преобразования, которые могут быть использованы в качестве диагностических признаков: исходное число коэффициентов непрерывного вейвлет-преобразования равно числу K дискретных значений сигнала, что исключает возможность полного перебора при больших значениях K ;

в) определение оптимальной конфигурации дерева декомпозиции дискретного вейвлет-преобразования [266].

Поскольку разрешение этих проблем не может быть обеспечено формальными методами, на практике конструктор ИТ чаще всего полагается на интуицию [257] и методом проб и ошибок ищет приемлемое решение задачи, которое не всегда достижимо.

Возможно, поэтому после бурного «всплеска» практический интерес исследователей к вейвлет-преобразованию постепенно угасает, о чем свидетельствуют данные библиографических исследований (рис.1.6).

Рис. 1.6. Распределение публикаций по вейвлет-преобразованиям [304]



1.4.2. Структурный анализ сигналов сложной формы

Оценка состояния многих технических и биологических объектов может быть основана на анализе признаков, которые сосредоточены на малых участках (фрагментах) области определения наблюдаемого сигнала [33, с. 327]. При этом форма таких сигналов достаточно сложна в том смысле, что невозможно адекватно описать такой сигнал аналитической функцией, заданной с точностью до небольшого числа параметров. Будем в дальнейшем называть такие сигналы сигналами с локально сосредоточенными признаками.

Примером сигнала с локально сосредоточенными признаками является термограмма кристаллизации металла. На ней в результате термических эффектов фазовых превращений образца наблюдаются характерные фрагменты, амплитудно-временные параметры которых несут информацию о химическом составе и позволяют прогнозировать механические свойства металла [306, 307].

Другим характерным примером является электрокардиограмма (ЭКГ) — кривая изменения во времени суммарного электрического потенциала, возникающего в сердечной мышце за счет движения ионов через мышечную мембрану [137]. Амплитудно-временные параметры характерных фрагментов ЭКГ несут ценную информацию о состоянии сердца [79, 231].

К анализу значений локально сосредоточенных признаков сводятся задачи компьютерной обработки многих других физиологических сигналов, в частности, реограмм [113, 189], сфигмограмм [45], балистокардиограмм [13], сейсмокардиограмм [14], пульсограмм [124], магнитокардиограмм [158, 366], электроэнцефалограмм [171] и т. п.

Классические методы цифровой обработки сигналов, в частности, разложения по ортогональным базисам [145], которые дают «интегральное» представление о компонентах сигнала на всей области определения, неэффективны для анализа подобных сигналов, поскольку локальные вариации признаков на ограниченных фрагментах мало сказываются на коэффициентах разложения. Однако, как будет показано далее, даже незначительные изменения формы локальных фрагментов таких сигналов обладают диагностической ценностью.

Для анализа сигналов с локально сосредоточенными признаками, которые имеют сложную структуру [133], применяется из-

вестный подход, использующий понятия формальных языков [244]. В различных публикациях такой подход называют лингвистическим [85, 134, 138, 343], структурным [33, 155, 226, 298] или синтаксическим [292].

Применительно к анализу временных сигналов структурный подход предполагает сегментацию (расчленение) сигнала на последовательность отдельных фрагментов, отражающих чередование элементарных событий в ходе развития исследуемого физического процесса [136]. Тем самым осуществляется переход от i -й реализации y_i сигнала, наблюдаемого на ограниченном временном интервале $t \in [0, T]$, к предложению (слову), $S_i = b_1 b_2 \dots b_{K_S}$, представляющему собой конечную цепочку символов $b_j \in B$, $j = 1, \dots, K_S$ из алфавита B «названий» фрагментов. Множество $\{S_i\}$ всевозможных предложений (не обязательно конечное) образует формальный язык, для которого строится грамматика в виде четверки [292]:

$$G = \langle \Omega_N, \Omega_T, P_G, S_0 \rangle,$$

где Ω_N — множество нетерминальных символов (переменных); Ω_T — множество терминальных символов (констант), $\Omega_T \cup \Omega_N = B$, $\Omega_T \cap \Omega_N = \emptyset$; P_G — множество грамматических правил (правил подстановки); $S_0 \in \Omega_N$ — начальный (корневой) нетерминальный символ.

Существуют различные типы грамматик (обобщенная, НС-грамматика, контекстно-свободная, автоматная), которые различаются способом формирования множества P_G правил допустимых подстановок [226]. В свою очередь, язык $L(G)$, порождаемый грамматикой G , представляет собой множество цепочек $\{S_i\}$, каждая из которых составлена только из терминальных символов и может быть выведена из S_0 соответствующим применением правил подстановки из множества P_G [194, с. 339].

Тогда, если для каждого состояния объекта $V_1, \dots, V_M$ удастся построить грамматику G_m , ($m = 1, \dots, M$), которая порождает язык $L(G_m)$ описания сигналов, свойственных этому состоянию, то для косвенной оценки текущего состояния объекта достаточно провести грамматический разбор цепочки фрагментов $S_i =$

$= b_1 b_2 \dots b_{K_S}$, соответствующей наблюдаемому сигналу $y_i(\cdot)$, т. е. принять решение в пользу состояния V_m , если $S_i \in L(G_m)$.

Дальнейшим обобщением лингвистического подхода является введение представлений о стохастических грамматиках [287], которые, в отличие от детерминированных, используют случайные правила подстановок в соответствии с множеством вероятностных мер Q_G , заданных на множестве правил P_G :

$$G = \langle \Omega_N, \Omega_T, P_G, Q_G, S_0 \rangle.$$

В большинстве работ, посвященных структурному анализу сигналов, предполагается, что алфавит B эталонных фрагментов заранее известен [33, с. 329], а конструирование грамматик G_m , адекватных множеству наблюдаемых сигналов, осуществляется человеком на основе неформальных знаний эксперта в предметной области [40, с. 160].

Несмотря на то, что в ряде работ, например в [259, 295], даны основные теоретические положения общей постановки проблемы восстановления грамматик по обучающей выборке наблюдений, как отмечается в [194, с. 371], до сих не существует универсальных приемов, которые можно было бы считать приемлемыми для разрешения этой достаточно серьезной проблемы. Можно привести лишь отдельные удачные примеры применения структурного анализа для обработки конкретных сигналов.

Так, в работах [33, 133, 134, 138] предложены методы автоматического построения описания структурных сигналов, основанные на выделении фрагментов «простого» и «сложного поведения». Границы фрагментов определяются по локальным экстремумам так называемой функции сложности $\Phi(y(t), \omega_j)$, которая последовательно вычисляется на элементарных участках $\omega_j = (t_j, t_{j+1})$ области определения обрабатываемого $y(t)$.

В качестве функций сложности предлагается использовать скалярное произведение вектора $g_j = (\tilde{y}_{j1}, \dots, \tilde{y}_{jL_{\omega}})$ нормированных дискретных значений сигнала на участке ω_j и среднего между соседними векторами g_{j-1} и g_{j+1} :

$$\Phi(\omega_j) = (g_j, 0,5 (g_{j-1} + g_{j+1})),$$

либо норму остаточной невязки

$$\Phi(\omega_j) = \sqrt{\sum_{s=1}^{L_0} (y_{js} - \sum_{i=1}^K c_{ji} \varphi_{is})^2},$$

характеризующую качество локальной аппроксимации сигнала на участке ω_j по системе ортонормированных векторов $\varphi_1 = (\varphi_{11}, \dots, \varphi_{1L_0}), \dots, \varphi_K = (\varphi_{K1}, \dots, \varphi_{KL_0})$, и ряд других функций, которые могут использоваться как независимо, так и в качестве обобщенной функции сложности, построенной на их основе [33, с. 347].

После сегментации $y(t)$ на ряд чередующихся простых и сложных фрагментов проводится их кластеризация и приписывание одинаковых имен фрагментам, имеющим близкую форму. Тем самым формируется алфавит B символов (наименований фрагментов), цепочки которых используются для структурного представления сигнала.

Как отмечают авторы метода [33, с. 358], платой за простоту служит весьма высокий уровень требуемых априорных сведений о модели обрабатываемого сигнала. Поэтому эти методы мало пригодны для обработки ЭКГ и других физиологических сигналов, а попытка их использования на практике приводит к обнаружению множества ложных фрагментов [242].

Структурный анализ сигналов рассматривают также с позиций обнаружения моментов изменения (разладки) вероятностных свойств наблюдаемого стохастического процесса [22, 30, 135, 238, 254, 318, 333, 345]. Общая постановка такой задачи требует проверки статистической гипотезы H_0 о постоянстве условного распределения для всех точек сигнала

$$P(y_k / y_{k-1}, \dots, y_0) = P_0, \quad 0 \leq k \leq K-1,$$

против гипотезы H_1 о существовании одного или нескольких моментов времени t_r (точек разладки), при которых происходят изменения условного распределения, т. е.

$$P(y_k / y_{k-1}, \dots, y_0) = P_1, \quad 0 \leq k < r,$$

$$P(y_k / y_{k-1}, \dots, y_0) = P_2, \quad r \leq k \leq K-1, \quad P_2 \neq P_1.$$

Однако практическое использование этого подхода требует большого объема данных. Для упрощения задачи часто предполагается, что модель случайного процесса известна с точностью

до конечного числа параметров [22]. В этом случае для обнаружения t_r достаточно проверить гипотезу об изменении параметров распределений до и после момента t_r .

Если допустить, что в точке t_r происходит изменение математического ожидания $M[y(t_k)]$, то при нормальном распределении y_k для решения задачи можно воспользоваться известным критерием Стьюдента [28, с. 107]. Когда же предположение о нормальности не выполняется, предлагается использовать другие критерии или различные эвристические алгоритмы [22].

Получили также известность методы структурного анализа, основанные на модели механизма чередования фрагментов сигнала. Чаще всего такая модель задается в виде цепи Маркова с матрицей условных вероятностей переходов [37, 58—61, 134]. Однако на практике оценить такие вероятности по имеющимся экспериментальным данным сложно.

В работах Т.К. Винцюка [46—49] предложены оригинальные методы поэлементного (фонемного) распознавания речевых сигналов на основе процедур динамического программирования, в которых на этапе обучения, проводимого для конкретного диктора, используются алгоритмы восстановления структуры речевого сигнала в виде кусочно-постоянной модели (эталонной последовательности). Однако этот удачный опыт не может быть использован при обработке ЭКГ и других физиологических сигналов хотя бы потому, что невозможно проводить этап обучения для каждого пациента.

Кроме того, как справедливо отмечается в работе [37], применению метода эталонных последовательностей при обработке реальных ЭКГ препятствуют вычислительные трудности, связанные с большим разнообразием формы каждого из ее информативных фрагментов. Поэтому описание ЭКГ потребовало бы использования большого количества элементарных эталонов, к тому же зависящих от двух параметров — продолжительности и амплитуды, которыми пришлось бы варьировать при построении эталонной последовательности.

Следует также учесть, что на реальных ЭКГ и других сигналах с локально сосредоточенными признаками, как правило, нет четких границ между информативными фрагментами. Даже такая на первый взгляд простая задача, как автоматическое выделение *QRS* комплексов, требует привлечения достаточно слож-

ных вычислительных процедур [140, 252, 316, 320, 338, 347, 374, 386]. Но, тем не менее, существующие компьютерные алгоритмы все же не удовлетворяют требованиям кардиологической практики [320, 368].

Помимо «внутренних» возмущений, искажающих форму информативных фрагментов, в реальных ситуациях обрабатываемый сигнал подвержен влиянию различных «внешних» возмущений в виде гармонических, случайных и импульсных помех, а также других возмущений, не поддающихся контролю [151, 308, 327, 381].

Вообще говоря, восстановление полезного сигнала на фоне помех является классической и многократно исследованной задачей цифровой обработки данных [9, 50, 66, 131, 193, 296, 336]. Однако использование известных алгоритмов частотно-избирательной фильтрации [50, 66, 132] и сглаживания случайных шумов [346, 379] часто приводит к недопустимым искажениям формы информативных фрагментов [151].

Именно поэтому в компьютерных программах анализа и интерпретации ЭКГ часто рекомендуется вообще отказаться от использования традиционных алгоритмов фильтрации [327], а для повышения соотношения сигнал—шум использовать так называемый метод накопления, основанный на усреднении некоторой последовательности типичных циклов ЭКГ [151, 243, 327, 381].

Главной предпосылкой такого подхода является представление наблюдаемого сигнала $y(t_k)$ в виде аддитивной смеси

$$y(t_k) = y_0(t_k) + \xi(t_k) \quad (1.26)$$

полезного сигнала $y_0(t_k)$ и случайной помехи $\xi(t_k)$. Тогда, если допустить, что $\xi(t_k)$ — последовательность независимых нормально распределенных величин с нулевым математическим ожиданием $M_\xi = 0$ и дисперсией σ_ξ^2 , то после усреднения M_c циклов дисперсия отклонений $y(t_k) - y_0(t_k)$ уменьшается до величины σ_ξ^2 / M_c [151].

Однако при обработке реальных сигналов такой привлекательный прием не всегда эффективен. Дело в том, что, согласно современным представлениям физиологов [291], нерегулярность и непредсказуемость частоты сердечных сокращений (ЧСС) являются важными характеристиками здорового организма и от-

ражают способность его регуляторных систем адекватно реагировать на изменения окружающей среды. Чрезмерное снижение изменчивости сердечного ритма (ригидность сердечного ритма) как раз является предвестником многих опасных заболеваний. Наблюдения также показывают [64], что при изменении ЧСС на ЭКГ и других физиологических сигналах происходят неравномерные изменения продолжительностей однотипных фрагментов различных циклов.

Все это приводит к тому, что при использовании примитивных алгоритмов усреднения циклов во временной области происходит «размывание» информативных фрагментов [151] и, как следствие, возникновение ошибок в измерении значений диагностических признаков, сосредоточенных на этих фрагментах, что в конечном итоге снижает достоверность результатов диагностики.

Именно поэтому в гл. 3 будет исследован альтернативный подход к усреднению искаженных реализаций сигналов с локально сосредоточенными признаками.

1.4.3. Анализ свойств сигналов в фазовом пространстве

Изучение поведения динамической системы, описываемой конечным набором параметров состояния $x_1, \dots, x_N$, удобно проводить в некотором абстрактном пространстве, координатами которого являются эти параметры. Такое N -мерное пространство называется *фазовым пространством*, его координаты — *фазовыми координатами*, а семейство фазовых траекторий, изображающих изменения состояния системы, — *фазовым портретом* [119, с. 61].

В общем случае фазовые траектории с течением времени стягиваются к некоторой области фазового пространства, попав в которую остаются там навсегда. Такая область называется *аттрактором*. Аттракторы бывают статические (фиксированная точка), периодические (предельный цикл) и хаотические (странный аттрактор) [68] (рис.1.7).

Детерминированные системы, имеющие странный аттрактор, демонстрируют поведение, которое внешне напоминает случайный процесс: сплошной спектр мощности, экспоненциально спадающую автокорреляционную функцию и т. п. Однако их динамика детерминирована и воспроизводима при условии точно-

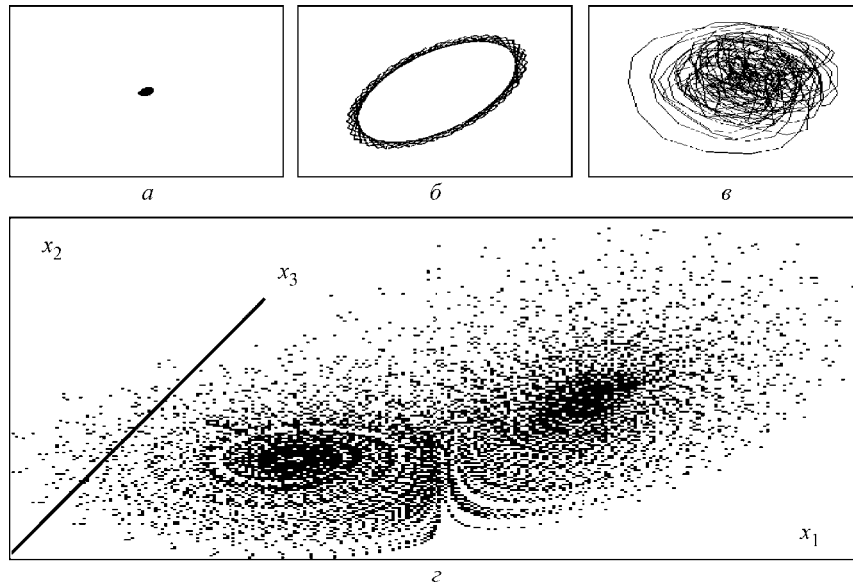


Рис. 1.7. Типы аттракторов:

a — статический; *б* — предельный цикл; *в* — странный аттрактор; *г* — странный аттрактор системы Лоренца

го повторения начальных условий. В 1963 г. Эдвард Лоренц опубликовал статью [319], в которой было показано, что фазовый портрет детерминированной системы, которая описывается тремя дифференциальными уравнениями

$$\begin{cases} \frac{dx_1}{dt} = -\sigma x_1 + \sigma x_2, \\ \frac{dx_2}{dt} = -x_1 x_3 + r x_1 - x_2, \\ \frac{dx_3}{dt} = x_1 x_2 - b x_3, \end{cases}$$

при определенных значениях параметров σ , r , b имеет вид странного аттрактора в пространстве фазовых координат x_1 , x_2 , x_3 (см. рис. 1.7, г).

Аттрактор Лоренца — первый классический пример детерминированной системы, демонстрирующей хаотическое поведение. Впоследствии изучением подобных процессов занимались многие исследователи. Появились многочисленные научные пу-

бликации, посвященные новому научному направлению — нелинейной динамике [8, 275, 285, 311, 358, 365] и тесно связанной с ней теории фракталов [371].

Согласно определению Бенуа Мандельброта, фракталом называется нерегулярная структура, состоящая из частей, которые в каком-то смысле подобны целому.

Важное место в нелинейной динамике отводится анализу бифуркаций — моментам перехода системы с одного аттрактора на другой, при которых происходит существенная перестройка динамики системы и даже возникает странный аттрактор [275, 285]. Для оценки степени хаотичности чаще всего вычисляют размерность аттрактора, используя различные численные методы, в частности алгоритм вычисления корреляционной размерности [293]. Интересно, что системы, демонстрирующие периодическую динамику, имеют целочисленную размерность аттрактора, в то время как размерность странного аттрактора — дробная.

Фазовый портрет изучаемой системы можно строить на основании экспериментальных данных. Традиционно для этого используется так называемый метод задержек [68, 225, 353], суть которого состоит в переходе от последовательности $y[1]$, $y[2]$, $y[3]$, ... дискретных значений сигнала $y(t)$ к последовательности D -мерных векторов:

$$z_k = (y_k, y_{k-\tau}, y_{k-2\tau}, \dots, y_{k-(D-1)\tau}),$$

где τ — величина задержки (лаг преобразования).

Теоретической основой такого метода являются теоремы, доказанные в работах [359, 370]. При этом считается, что длина временного ряда должна быть порядка 10^{R_A} , где R_A — оценка размерности аттрактора, а оптимальная величина задержки τ определяется по точке первого пересечения автокорреляционной функции с нулем.

Исследования физиологов показали, что многие сигналы, порождаемые живым организмом, демонстрируют хаотическое поведение и имеют фрактальную структуру [68, 291, 310]. Поэтому большинство прикладных работ, в которых исследуются методы анализа физиологических сигналов в фазовых координатах, посвящены изучению степени хаотичности таких сигналов на основе оценки размерности аттракторов.

Анализ физиологических сигналов в фазовом пространстве координат может быть полезен не только с точки зрения оценки

его хаотичности. Еще в 1972 г. Н.А. Амосов и его коллеги [10] говорили о возможности исследования сократительной функции миокарда в фазовом пространстве, координатами которого является амплитуда $y(t)$ и производная по времени dy/dt наблюдаемого процесса.

Диагностическая ценность таких исследований обусловлена тем, что при различных поражениях миокарда изменяется не только направление, но и скорость распространения волны деполяризации и реполяризации по миокарду [54, 104, 229]. На современном этапе развития компьютерной техники эти плодотворные идеи могут быть реализованы в медицинских информационных технологиях. Результаты практического применения таких технологий будут рассмотрены в гл. 6.

1.5. Задача оценки эффективности диагностических признаков

Главная цель селекции признаков состоит в переходе от исходного множества признаков $X = \{x_1, \dots, x_{N_0}\}$ к подмножеству

$X_0 \subset X$ из N_0 признаков ($N_0 < N$) без уменьшения достоверности диагностики. В отличие от задачи конструирования исходного множества X такая цель может быть достигнута при использовании формальных методов, исследованию которых посвящены многочисленные научные публикации.

Большинство доступных публикаций, например работы [41, 43, 67, 139, 195, 314, 331, 364, 369, 373], посвящены исследованию процедур селекции признаков на этапе обучения. Так, в работе [67, с. 62] рассмотрен метод селекции признаков, основанный на анализе коэффициентов $\alpha_1, \dots, \alpha_N$ уравнения линейной дискриминантной функции

$$D(\cdot) = \sum_{n=1}^N \alpha_n x_n + \alpha_0,$$

построенной по имеющейся обучающей выборке наблюдений. Бесполезными считаются признаки, которым соответствуют малые значения $|\alpha_n|$.

Следует отметить, что при неограниченной выборке добавление признака, даже не несущего никакой информации, не может ухудшить качество обучения. Однако при малом объеме вы-

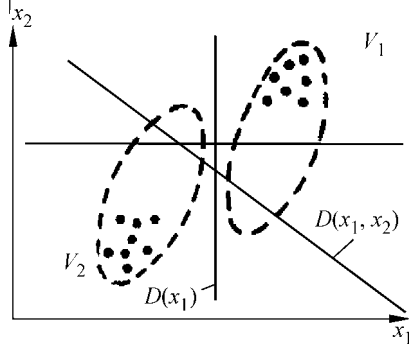


Рис. 1.8. Одномерная и двумерная решающие функции $D(\cdot)$

борки включение в описание бесполезных признаков может нанести вред. Для иллюстрации сказанного рассмотрим наглядный пример (рис. 1.8), приведенный в работе [86, с. 68].

Видно, что использование в линейной решающей функции $D(\cdot)$ только одного признака

x_1 гораздо лучше, чем совокупности $x = (x_1, x_2)$, которая разделяет выборку, но не разделяет генеральную совокупность объектов линейно разделимых классов V_1 и V_2 .

Поскольку конструктор прикладных ИТ довольно часто располагает лишь малой обучающей выборкой, важно иметь средства, позволяющие обоснованно подходить к выбору признаков еще до этапа обучения.

К сожалению, лишь немногие публикации [19, 20, 27, 69, 75, 84, 222, 317] посвящены исследованию свойств, влияющих на диагностическую ценность признаков.

В работе [317] диагностическую ценность признака x_n ,

$1 \leq n \leq N$, входящего в совокупность $x^{(N)} = (x_1, \dots, x_N)$, связывают с количеством информации $I(V, X_n)$, содержащейся во множестве $X_n = \{x_n\}$ возможных значений этого признака относительно множества $V = \{V_1, \dots, V_M\}$ распознаваемых классов. Иными словами, признак считается ценным, если

$$I(V, X_n) = H(V) - H(V / X_n) > 0, \quad (1.27)$$

где $H(V)$ и $H(V / X_n)$ — соответственно начальная и средняя условная энтропии по Шеннону.

Легко показать, что если признак x_n одинаково распределен в классах, т. е.

$$p(x_n / V_1) \equiv p(x_n / V_2), \quad (1.28)$$

то условие (1.27) не выполняется. По определению Бен-Бассата

[256], такой признак называется нерелевантным, поскольку при выполнении условия (1.28) $P(V_k / x_n) \equiv P(V_k)$, $k = 1, 2$.

Критерий различия условных распределений $p(x_n / V_1) \neq p(x_n / V_2)$ до сих пор активно применяют в задачах медицинской диагностики для оценки эффективности отдельных признаков (тестов) на основе **ROC**-анализа (*receiver operating characteristic curve*) [294, 326, 328, 339].

В то же время еще в работах [84, 277] было показано, что различие условных распределений признака в классах недостаточно для того, чтобы гарантировать уменьшение вероятности ошибочных решений. Там же предложен информационный критерий, гарантирующий полезность признака с этой точки зрения, который, в отличие от (1.27), сводится к проверке условия

$$I(V, X_n) > \log(1 + \lambda_0) - \frac{\lambda_0}{1 + \lambda_0} \log \lambda_0 - 2 \min \left\{ \frac{1}{1 + \lambda_0}, \frac{\lambda_0}{1 + \lambda_0} \right\}, \quad (1.29)$$

где $\lambda_0 = P(V_2) / P(V_1)$ — соотношения априорных вероятностей классов.

Между тем известно [106], что средняя вероятность ошибочных решений является лишь частным случаем среднего риска и не учитывает соотношения потерь от ошибок различного рода. Поскольку для задач медицинской и технической диагностики потери от ошибок пропуска цели и ложной тревоги чаще всего неравнозначны, актуальным является получение условий, гарантирующих полезность диагностических тестов и отдельных признаков с точки зрения уменьшения среднего риска.

Новые результаты таких исследований будут представлены в гл. 5.

1.6. Современные подходы к программной реализации информационных технологий

Анализ проблемы построения эффективных информационных технологий обработки сигналов будет неполным, если не рассмотреть современные тенденции к построению программных систем. Поскольку это направление представляет собой отдельную и достаточно обширную область знаний, которая подробно отражена в многочисленных монографиях [12, 24, 29, 36, 44, 101 и др.], дадим лишь краткую характеристику существующим подходам, остановившись на отдельных вопросах, которые нас будут

интересовать в дальнейшем при построении инструментальной системы для интерактивного синтеза информационных технологий обработки сигналов.

Безусловно, каждая конкретная ИТ имеет свои особенности, связанные с характером данных и методами их обработки, которые диктуют свою логику ее программной реализации. Но если следовать только этой логике, то опыт, накопленный в ходе создания одной системы, едва ли удастся использовать при построении другой. Это приведет к тому, что разработка каждой новой прикладной ИТ будет начинаться практически «с нуля».

Традиционные методы процедурного программирования не способны справиться с растущей сложностью программ, необходимостью их постоянной модернизации и повышения надежности работы. Поэтому возникла новая методология, получившая название «объектно-ориентированное программирование» (**ООП**) [12]. Фундаментальным понятием **ООП** является понятие программного класса, в котором инкапсулируются данные, свойства и методы их обработки. При этом класс рассматривается как абстрактная совокупность объектов (экземпляров класса), которые имеют набор общих свойств и одинаковы по поведению.

Важная особенность классов — это возможность их организации в виде некоторой иерархической структуры за счет механизма наследования, причем объекты производных классов (наследников) могут модифицировать свое внешнее поведение по отношению к одноименным методам родительского класса, используя специальный механизм полиморфизма.

Методология **ООП** поддерживается современными языками программирования высокого уровня (**C++**, **Java**, **Object Pascal** и др.) [24]. Заметно ее огромное влияние на процесс разработки программ, структура которых основывается на объектно-ориентированной декомпозиции [101]. При этом взаимодействие отдельных компонент программы друг с другом и внешним миром (пользователем) осуществляется с использованием механизма передачи сообщений (управляющих событий).

Отсюда следует, что главная проблема, возникающая при программной реализации прикладной ИТ на основе методологии **ООП**, — конструирование программных классов, достаточных для достижения поставленной цели, в том числе определение свойств и методов, необходимых для придания классам требуемого поведения и возможности их взаимодействия.

К сожалению, эта проблема не может быть разрешена формальными методами, а значит, успех «оптимальной» декомпози-

ции целиком зависит от опыта системного аналитика — архитектора проектируемой системы.

Все возрастающая сложность программных систем привела к созданию методологии объектно-ориентированного анализа и проектирования [36], а также инженерных средств разработки программного обеспечения, так называемых CASE-технологий (*Computer Aided Software Engineering*) [44].

В 1997 г. консорциумом **OMG** (*Object Management Group*) принята стандартная нотация визуального моделирования программных систем, на основе которой создан унифицированный язык моделирования **UML** (*Unified Modeling Language*) [29], который в настоящее время поддерживается многими объектно-ориентированными CASE-технологиями, в частности *Rational Rose*, *Model Maker* и др. Наконец, фирма *Rational Software Corp.* разработала методологию разработки программных систем **RUP** (*Rational Unified Process*).

Следует, однако, признать, что перечисленные средства главным образом ориентированы на построение организационных ИТ, в которых основными задачами являются проектирование сложных баз данных, организация специфических запросов, подготовка отчетов и т. п. [147]. Основные же проблемы построения ИТ обработки сигналов лежат в «алгоритмической» плоскости. Необходимо выбрать и настроить оптимальную в некотором смысле технологическую цепочку вычислительных процедур, обеспечивающих формирование информационного продукта на основе обработки сигнала, искаженного внутренними и внешними возмущениями.

Из проведенного выше анализа следует, что разрешить эту проблему только формальными методами невозможно: необходима интеграция опыта и знаний конструктора ИТ и эксперта предметной области. Поэтому для разработки математического и программного обеспечения современных компьютерных систем предлагаются новые подходы, в частности модели описания данных многоуровневыми алгоритмическими структурами (квантами знаний), которые допускают алгебраические и логические преобразования на основе векторно-матричных операторов [176].

При построении прикладных ИТ в условиях недостаточной априорной информации часто требуется проводить дополнительные экспериментальные исследования по изучению сигналов, порождаемых объектом исследований. Традиционной инструментальной поддержкой таких исследований служат современные средства компьютерной математики [63, 82, 174, 312,

324 и др.], в разработке которых специализируются крупные фирмы *MathWork Inc.*, *MathSoft Inc.*, *Wolfram Research Inc.*, *MathWare Inc.*, *StatSoft Inc.*, *Management Intelligenter Technologien GmbH*, *Hewlett-Packard Laboratories* и многие другие. На рынке есть огромное количество программных продуктов (системы *Matlab*, *Derive*, *MuPad*, *Mathcad*, *Mathematica*, *Data Engine*, *Statistica*, *Khoros*, *DesignLab* и др.), в которые включены средства анализа сигналов — фильтрации, аппроксимации, преобразования и т. п.

Безусловно, при разработке прототипа проектируемой прикладной ИТ использование таких систем позволяет существенно повысить качество и сократить сроки исследований, направленных на поиск подходящих алгоритмов обработки сигнала. В то же время эти системы не могут быть в полной мере отнесены к инструментальным системам разработки прикладных ИТ, реализация которых требует использования языков программирования высокого уровня (C++, **Delphi** и др.). Дело в том, что известные системы компьютерной математики (например, **MATLAB** [81]), являясь открытыми по расширению их функций, остаются закрытыми для пользователя в том смысле, что он, по существу, лишен возможности «извлечения» из системы отдельной вычислительной процедуры для ее использования в собственном программном продукте. Опыт показывает, что даже если на этапе экспериментальных исследований в среде **MATLAB** удастся выбрать подходящую цепочку вычислительных процедур, то на конечном этапе построения прикладной ИТ все равно приходится писать собственные программы для реализации вычислительных процедур на языке программирования высокого уровня.

Но при написании собственной программы на языке программирования конструктора ИТ часто подстерегают дополнительные проблемы, связанные с недостаточным пониманием тонкостей того или иного вычислительного алгоритма, который он с успехом применял в стандартном пакете типа **MATLAB**. К тому же, как уже отмечалось, эффективная обработка сигналов с локально сосредоточенными признаками не может быть обеспечена применением только традиционных (стандартных) алгоритмов.

Сказанное выше подтверждает актуальность разработки специализированной инструментальной системы, которая содержала бы готовый к практическому использованию набор базовых вычислительных процедур, необходимых при построении прикладных ИТ обработки сигналов с локально сосредоточенными признаками. Предлагаемый нами подход к построению такой системы будет описан в гл. 4.

**2.1. Особенности компьютерной обработки
временных сигналов
с признаками, сосредоточенными
на локальных фрагментах**

Вернемся к абстрактной модели ИТ (см. рис. 1.2), предполагая, что сигнал $y(t)$ несет информацию о состоянии s объекта, однако «физическая» модель в виде функциональной зависимости $s = s[y(t)]$ не известна.

Наиболее естественное представление скалярного сигнала $y(t)$, с которым имеет дело специалист при визуальном анализе, — представление $y(t)$ на отрезке наблюдения $t \in [t_1, t_2]$ в виде графика функции. Компьютерный же анализ сигнала предполагает дискретную форму представления $y(t)$ в виде вектора $y^{(K)} = (y[1], \dots, y[K])$, компонентами которого служат ординаты, взятые в точках отсчета с некоторым шагом Δ вдоль оси t , чаще всего равномерным. Поскольку реальные сигналы имеют ограниченный спектр, то, в соответствии с теоремой В.А. Котельникова, всегда может быть выбран такой шаг квантования Δ , при котором непрерывный сигнал $y(t)$ полностью определяется дискретной последовательностью $y^{(K)}$ [91, с. 51].

В этом случае массив экспериментальных данных, имеющихся в распоряжении конструктора ИТ, можно представить в виде матрицы, строки которой соответствуют наблюдаемым реализациям, а столбцы — их ординатам. Если при этом для каждой строки такой таблицы известно состояние s обь-

ТАБЛИЦА 2.1

Гипотетическая выборка наблюдений

№	Регрессоры (аргументы модели)					Отклик (выход)
1	T_{11}	T_{21}	T_3	...	T_{K1}	C_1
2	T_{12}	T_{22}	T_{32}	...	T_{K2}	C_2
...	...	...	...	...	...	...
W	T_{1W}	T_{2W}	T_{3W}	...	T_{KW}	C_W

екта, то, казалось бы, общая проблема синтеза ИТ может быть сведена к восстановлению неизвестной зависимости по таблице экспериментальных данных (табл. 2.1) с использованием классических методов, о которых шла речь в гл. 1. Однако при обработке сигналов с локально сосредоточенными признаками использования этих методов не достаточно для разрешения общей проблемы [207].

Для того чтобы убедиться в этом, рассмотрим конкретный пример — задачу косвенной оценки химического состава железоуглеродистых расплавов по сигналу $y = T(t)$, несущему информацию о температуре расплава в процессе его кристаллизации. Вначале предельно упростим задачу, предполагая, что конечного пользователя (металлурга) интересует всего лишь один технологический параметр — процентное содержание углерода C .

Рассмотрим рис. 2.1, в левой части которого в координатах T (температура) — C (содержание углерода) показан фрагмент известной диаграммы состояния Fe-C сплава [74]. На диаграмме показаны две линии — ликвидуса и солидуса. Выше линии ликвидуса металл находится в жидком состоянии (область L), ниже линии солидуса — в твердом состоянии (область S), а между ними — в переходном состоянии (область $L + S$).

Как видно из рисунка, линия ликвидуса не параллельна оси температур, а значит, при изменении содержания углерода изменяется температура начала кристаллизации металла (температура ликвидуса T_L): чем выше содержание углерода, тем при более низкой температуре начинается процесс кристаллизации расплава. Поэтому неизвестное содержание углерода C можно определить по температуре ликвидуса T_L .

Однако сталь является сложным по химическому составу сплавом и, помимо углерода, в ней есть еще и другие примеси (марганец, кремний, фосфор, сера и др.), которые так же, как

углерод, влияют на температуру ликвидуса, хотя и в меньшей степени. Поэтому зависимость между C и T_L отличается от зависимости, вытекающей из диаграммы состояния Fe-C сплава.

В то же время, если воспользоваться простейшей моделью в виде линейного уравнения регрессии

$$\hat{C} = \alpha_0 + \alpha_1 T_L, \quad (2.1)$$

и определить параметры α_0, α_1 для условий конкретного предприятия, то, как показывает практика, содержание углерода можно оценивать по температуре T_L с достаточно высокой точностью (порядка 0,02 абс. % C).

Разумеется, такая сравнительно высокая точность может быть достигнута лишь в том случае, когда будет правильно определена температура T_L . Однако возникает вопрос: каким образом определить температуру ликвидуса.

Природа пошла навстречу в разрешении этой проблемы. Дело в том, что в момент t_L начала кристаллизации происходит экзотермическая реакция — выделение скрытой теплоты кристаллизации. Поэтому, если зарегистрировать процесс охлаждения расплава (см. правую часть рис. 2.1), то на графике сигнала $T(t)$ при достижении температуры ликвидуса появляется характерный фрагмент в виде температурной остановки (площадки). Обнаружив такой фрагмент (визуально или на основе компьютерной обработки сигнала), можно определить температуру T_L , а затем и процентное содержание углерода по модели (2.1).

Предположим теперь, что нам известна лишь некоторая часть описанных выше закономерностей, т. е. только гипотеза о том, что наблюдаемый сигнал $y = T(t)$ несет информацию об интересующем нас параметре C , но не известно, каким образом.

Можно ли в этих условиях воспользоваться алгоритмом самоорганизации МГУА для восстановления зависимости (2.1) по таблице экспериментальных данных?

Пусть в нашем распоряжении есть конечная выборка термограмм, зарегистрированных для W проб металла, и для каждой пробы из результатов химического анализа известны

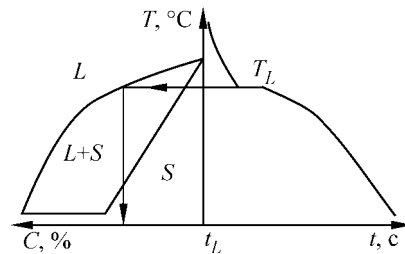


Рис. 2.1. Иллюстрация идеи косвенной оценки C по сигналу $T(t)$

значения содержания углерода $C_1, \dots, C_W$. В этом случае дискретные значения температуры $T_1, \dots, T_K$ можно считать потенциальными регрессорами, для которых известны точные значения отклика C (табл. 2.1).

Казалось бы, имея такую выборку, можно легко восстановить зависимость (2.1), задав класс базисных функций в виде всевозможных линейных структур

$$\begin{aligned} g_{(1)} &= \alpha_i T_i + \alpha_0, \quad i = 1, \dots, K, \\ g_{(2)} &= \alpha_i T_i + \alpha_j T_j + \alpha_0, \quad i, j = 1, \dots, K; j \neq i, \\ &\dots \\ g_{(K)} &= \alpha_1 T_1 + \alpha_2 T_2 + \dots + \alpha_K T_K + \alpha_0. \end{aligned} \quad (2.2)$$

На самом же деле это не так! Дело в том, что момент t_L начала кристаллизации определяется множеством случайных факторов (начальной температурой расплава, условиями теплообмена и т. п.). Поэтому на разных реализациях информативный фрагмент (площадка ликвидуса) будет появляться в случайные моменты времени, а значит, для разных строк табл. 2.1 температуре ликвидуса T_L (истинному аргументу искомой модели) будут соответствовать разные номера дискретных значений $T_1, \dots, T_K$.

Иными словами, несмотря на взаимосвязь C и T_L , может отсутствовать корреляция между C и отдельными значениями сигнала $T_i, i = 1, \dots, K$, и при сколь угодно большом объеме W выборки невозможно восстановить истинную зависимость (2.1) в классе структур (2.2) без привлечения дополнительной информации. К тому же при больших значениях K возникают известные сложности в вычислении при использовании алгоритмов самоорганизации, но это уже другие проблемы.

Таким образом, попытка трактовать отдельные дискретные значения сигнала $y[1], \dots, y[K]$ как потенциально полезные признаки в общем случае не правомерна, поскольку она игнорирует две существенные особенности данных — их линейную упорядоченность по оси аргумента и многообразие возможностей представления информации, содержащейся в них.

Следуя [33, с. 328], можно считать, что выбор тех или иных методов обработки сигнала, прежде всего, связан с содержательной гипотезой о том, как *распределена* диагностически ценная информация на области определения наблюдаемого сигнала. В тех

случаях, когда информация *равномерно распределена* на всем интервале наблюдения $[0, T]$ сигнала, его правомерно характеризуют интегральными характеристиками, вычисленными на всей области определения. Примерами таких характеристик служат методы получения обобщенных коэффициентов Фурье при разложении сигнала $y(t)$ по различным системам базисных функций [145].

Совершенно иные методы извлечения информации необходимы в том случае, когда она сосредоточена на определенных относительно небольших фрагментах сигнала, как это было показано в рассмотренном выше примере. Такие сигналы мы и будем рассматривать в дальнейшем, называя их сигналами с локально сосредоточенными признаками.

Прежде чем конкретизировать общую схему формирования информационного продукта в ИТ обработки сигналов с локально сосредоточенными признаками, рассмотрим некоторые характерные примеры таких сигналов, которые порождают объекты различной физической природы.

2.2. Примеры сигналов с локально сосредоточенными признаками

2.2.1. Электрокардиограмма, несущая информацию о состоянии сердечно-сосудистой системы живого организма

Метод электрокардиографии по справедливости относят к одному из наиболее распространенных инструментальных методов практической медицины. Это объясняется не только тем, что заболевания сердечно-сосудистой системы лидируют среди прочих, в том числе в Украине, но и тем, что этот метод достаточно долго и хорошо себя проявил при диагностике различных патологий.

Дадим краткую характеристику этому методу [137, 231], останавливаясь лишь на тех деталях, которые нас будут интересовать в дальнейших исследованиях.

Сердце обладает рядом функций, определяющих особенности его работы. К ним относятся функции:

- *автоматизма* — способность сердца самостоятельно вырабатывать электрические импульсы (при отсутствии внешних раздражителей);

- *возбудимости* — способность сердца возбуждаться под влиянием электрических импульсов;
- *проводимости* — способность к проведению возбуждения, возникшего в каком-либо участке сердца, к другим отделам сердечной мышцы (миокарда);
- *сократимости* — способность миокарда сокращаться в ответ на возбуждение.

Именно в результате последовательного сокращения различных отделов сердца осуществляется его основная насосная функция.

В основе возникновения электрических явлений в сердце лежит механизм проникновения ионов через мышечную мембрану, в основном ионов калия K^+ и натрия Na^+ . При этом в объеме сердца одновременно функционирует большое количества элементарных источников тока — сердечных диполей, каждый из которых обусловлен возбуждением отдельных мышечных волокон и отличается от других таких же диполей как по величине, так и по направлению.

Возбуждение мышцы сердца разделяется на две фазы — *деполяризацию*, т. е. уменьшение разности потенциалов, существующей между внутренней и наружной стороной мембран клеток сердца, и *реполяризацию*, т. е. восстановление этой разности потенциалов. Благодаря механизму электромеханического сопряжения процесс электрического возбуждения миокарда вызывает его механическое сокращение — систолу сердца.

Следует отметить, что процессы деполяризации и реполяризации отличаются друг от друга. Деполяризация представляет собой автоволновой процесс, при котором каждый возбужденный участок миокарда как бы «поджигает» последующий участок и в каждый момент времени можно установить, где находится волна возбуждения. Скорость распространения волны возбуждения выше в начале деполяризации и ниже в конце.

В отличие от деполяризации, реполяризация — процесс диффузный, т. е. сразу большие зоны миокарда охватываются этим процессом. Считается, что у здорового человека скорость реполяризации ниже в начале процесса и выше в конце.

Весьма сложно получить точное математическое описание реальных электродинамических закономерностей, происходящих в объеме сердца. Однако согласно дипольной концепции, принятой в современной электрокардиографии, при определенных допущениях сердце можно условно рассматривать как один то-

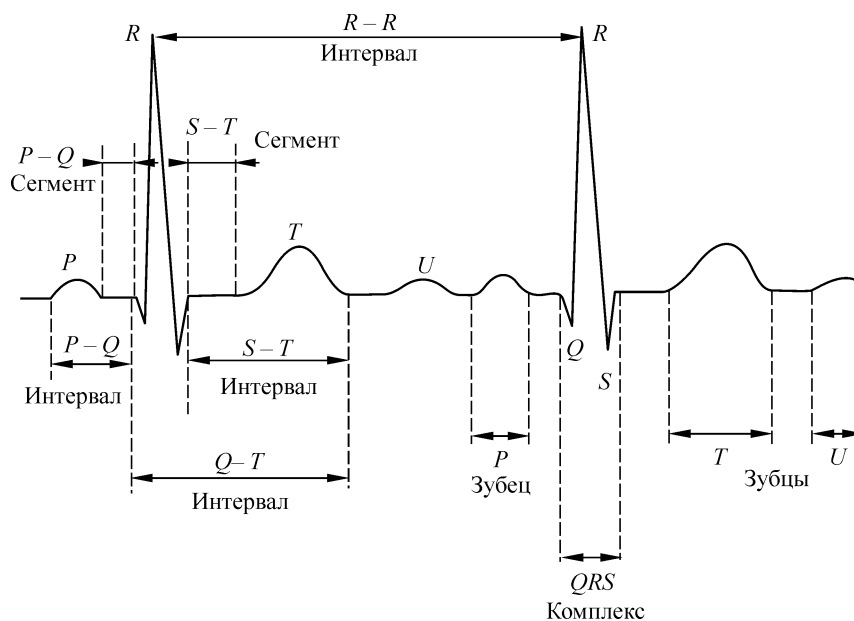


Рис. 2.2. Локальные фрагменты ЭКГ здорового человека

чечный источник тока (сердечный диполь), создающий в окружающем его объемном проводнике (теле) электрическое поле, которое может быть зарегистрировано с помощью электродов, расположенных в определенных точках поверхности тела.

Электрокардиограмма (ЭКГ) представляет собой запись сигнала $E(t)$, несущего информацию об изменении во времени суммарного электрического потенциала, возникающего в сердечной мышце за счет движения ионов через мышечную мембрану.

Как известно, сердце анатомически состоит из двух предсердий (правого и левого) и двух желудочков (правого и левого). Соответственно на сигнале $E(t)$ наблюдаются характерные фрагменты, которые отображают стадии возбуждения отдельных участков сердца (рис. 2.2).

Зубец P образуется вследствие возбуждения предсердий. Сегмент $P-Q$ обычно располагается на нулевой линии (изолинии) и соответствует времени прохождения возбуждения по так называемому атриовентрикулярному (предсердно-желудочковому) соединению. Комплекс QRS отражает сложный процесс возбуждения (деполяризации) желудочков. Начальный зубец Q регистри-

руется во время возбуждения левой части межжелудочковой перегородки.

Зубец R (чаще всего наиболее выраженный зубец ЭКГ) обусловлен возбуждением основной массы миокарда левого и правого желудочков. Зубец S в основном обусловлен возбуждением основания левого желудочка. Интервал $S—T$ называется конечной частью желудочкового комплекса и отражает реполяризацию желудочков. Он разделяется на сегмент $S—T$, отражающий период угасания возбуждения желудочков, и зубец T , форма которого отражает процесс быстрой реполяризации желудочков.

Изредка за зубцом T следует небольшой зубец U , происхождение которого до сих пор не установлено [137].

Интервал $T—P$ от окончания зубца T и до начала зубца P следующего цикла ЭКГ соответствует диастоле (расслаблению) сердца.

Традиционными диагностическими признаками ЭКГ являются амплитудно-временные параметры, сосредоточенные на указанных фрагментах. Например, основные электрокардиографические признаки ишемической болезни сердца (ИБС) сосредоточены на $S—T$ интервале и зубце T [79, 381, 389].

В кардиологической практике ЭКГ чаще всего измеряется в нескольких отведениях [137]. Каждое такое отведение несет информацию о разности потенциалов между двумя определенными точками электрического поля сердца, в которых установлены электроды. Обычно используют 12 отведений ЭКГ: три стандартные отведения, которые принято обозначать римскими цифрами I, II, III, три усиленных отведения aVR , aVL , aVF и шесть грудных отведений V_1 , ..., V_6 .

Установлено, что каждое из отведений преимущественно отражает состояние участков сердечной мышцы определенной локализации. Поэтому по изменениям ЭКГ в тех или иных отведениях можно судить о локализации патологических изменений в миокарде. Анализ диагностических признаков ЭКГ позволяет выявить те или иные патологические изменения в состоянии миокарда (инфаркт, ишемия, гипертрофия, гипоксия, миокардиодистрофия, кардиосклероз и др.).

Понятно, что результаты визуальной оценки ЭКГ существенно зависят от опыта и квалификации врача, осуществляющего «расшифровку» ЭКГ. При этом не всегда одна и та же ЭКГ однозначно интерпретируется разными специалистами, поскольку

ку, учитывая формальные критерии, каждый из врачей в большинстве случаев действует неформально, опираясь на свой предшествующий опыт и интуицию.

Именно поэтому все большее внимание специалистов привлекают компьютерные системы обработки ЭКГ. Однако существующие компьютерные системы еще не в полной мере удовлетворяют требованиям врачей-кардиологов. Довольно часто такие системы допускают ошибки еще на стадии морфологического анализа информативных фрагментов ЭКГ и измерения значений признаков, сосредоточенных на этих фрагментах.

Согласно [37, 109], даже при одновременной обработке трех синхронно записанных отведений совпадение результатов компьютерного и визуального способов измерений с требуемой точностью получено только в 90 % случаев для продолжительности интервала $P-R$ и комплекса QRS и в 80 % случаев для продолжительности зубца P и интервала $Q-T$.

Такие проблемы объясняются тем, что на практике встречается не менее 10 различных типов зубца P , 13 типов сегмента $S-T$, 12 типов зубца T и 55 типов комплекса QRS , которые могут существенно различаться по форме [137, с. 58] (рис. 2.3).

Дополнительные проблемы компьютерной обработки ЭКГ создают внутренние и внешние возмущения, которые искажают форму ее информативных фрагментов. Поэтому при решении да-

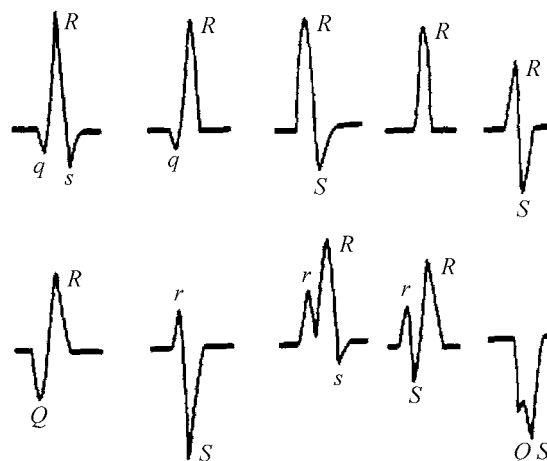


Рис. 2.3. Некоторые типичные формы комплекса QRS

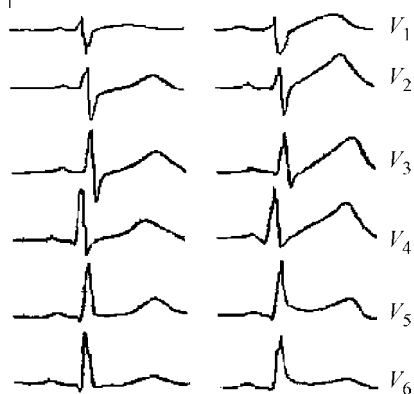


Рис. 2.4. ЭКГ в грудных отведениях при остром мелкоочаговом инфаркте: *a* — до инфаркта; *б* — в 1-й день инфаркта

же такой на первый взгляд простой задачи, как автоматическое разделение записи ЭКГ на отдельные циклы (R — R -интервалы), до сих пор возникают известные проблемы [320, 368].

Еще сложнее проводить компьютерный анализ амплитудно-временных параметров ин-

тервала S — T и зубца T , которые, как уже отмечалось, являются основными электрокардиографическими признаками ИБС. Даже незначительные изменения этих параметров могут свидетельствовать о серьезном заболевании [137, с. 235] (рис.2.4).

Вместе с тем из-за внутренних и внешних возмущений указанные признаки варьируют от цикла к циклу. В существующих компьютерных электрокардиографах реализованы алгоритмы, которые предусматривают накопления и усреднения во временной области последовательности сердечных циклов с последующим анализом диагностических признаков по усредненному циклу (эталону) [381].

Однако при изменении частоты сердечных сокращений (ЧСС) не только изменяется общая продолжительность R — R -интервалов, но и соотношения продолжительностей отдельных фрагментов этих циклов. Например, продолжительность комплекса QRS меньше связана с изменением ЧСС, чем продолжительности зубцов P и T [64].

Понятно, что для корректной оценки эталона требуется совместить во времени однотипные фрагменты усредняемых циклов. Однако на реальных ЭКГ обычно отсутствуют четкие границы между указанными фрагментами, что затрудняет их надежное распознавание во временной области [37].

Все это приводит к тому, что при использовании алгоритмов усреднения циклов ЭКГ во временной области происходит «размывание» формы информативных фрагментов эталона [151] и, как следствие, ошибки в измерении диагностических признаков, что в конечном итоге снижает достоверность результатов компьютерной диагностики.

2.2.2. Электрофизиологические сигналы, также используемые в кардиологической практике

Для получения дополнительной информации о состоянии сердечно-сосудистой системы в клинической практике, помимо электрокардиографии, используются и другие инструментальные методы, которые также основаны на анализе временных сигналов с локально сосредоточенными признаками, в частности, реограммы [113, 168, 190], сфигмограммы [45], баллистокардиограммы [13], сейсмокардиограммы [14], пульсограммы [124], магнитокардиограммы [156, 157, 183].

Для полноты картины дадим краткую характеристику некоторым из этих сигналов.

Реография — метод исследования кровоснабжения органов, в основе которого лежит принцип регистрации изменений электрического сопротивления тканей в связи с меняющимся кровенаполнением в изучаемом участке тела, например в конечностях. Чем больше приток крови к тканям, тем меньше их сопротивление.

Для получения реограммы (РГ) через тело пациента пропускают переменный ток частотой 50—100 кГц, малой величины (не более 10 мкА), создаваемый специальным генератором, и изучаются пульсовые колебания электрического сопротивления. При увеличении кровенаполнения амплитуда РГ возрастает и наоборот; другими словами, регистрируется динамика импеданса в обратной полярности.

Чаще всего реограмму регистрируют параллельно с ЭКГ, *QRS*-комплекс которой «сигнализирует» о моменте сокращения миокарда (рис. 2.5).

Анализ РГ сводится к измерению амплитудно-временных параметров, сосредоточенных на систолическом и диастолическом фрагментах сигнала, обусловленных соответственно артериальным притоком и венозным оттоком крови. В частности, диагностически ценными признаками являются: амплитуда A_1 сигнала, соответствующая восходящему колену систолической волны; максимальная A_2 и минимальная A_3 амплитуды систолической волны; максимальная амплитуда A_4 диастолической волны; время Q_x распространения реографической волны от источника (сердца) до исследуемого органа; время α максимального систолического наполнения и продолжительность β нисходящей волны реограммы (см. рис. 2.5).

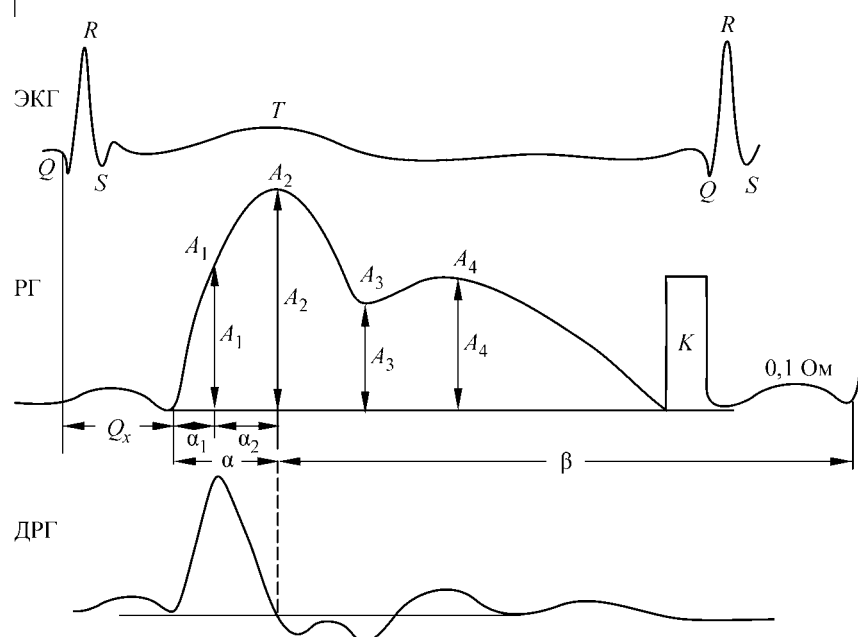


Рис. 2.5. Параллельная запись реограммы (РГ), дифференциальной реограммы (ДРГ) и электрокардиограммы (ЭКГ)

Разумеется, реограмму, показанную на рис. 2.5, можно встретить лишь в учебниках по реографии. В реальных же условиях наблюдаемый сигнал подвержен различным возмущениям (рис. 2.6.), что затрудняет компьютерный анализ указанных локально сосредоточенных признаков¹.

Сфигмография — метод, направленный на изучение колебаний артериальной стенки, обусловленных выбросом ударного объема крови в артериальное русло. С каждым сокращением сердца увеличивается давление в артериях и имеет место прирост их поперечного сечения, затем происходит восстановление исходного состояния. Весь этот цикл превращений получил название «артериальный пульс», а запись его в динамике — «сфигмограмма».

¹ Регистрация реограмм проводилась оптоэлектронным датчиком [189]. Исходные данные предоставлены автору аспирантом Винницкого политехнического института Аль-Кхури Талал.

Различают сфигмограммы центрального пульса, когда запись производится на крупных артериях, близко расположенных к сердцу (подключичной, сонной), и сфигмограммы периферического пульса более мелких артериальных сосудов. Для регистрации сфигмограммы используют пьезоэлектрические датчики.

Синхронная запись ЭКГ и сфигмограмм сонной, бедренной и лучевой артерий (рис. 2.7.) позволяет определить скорость распространения пульсовой волны. Для этого по так называемым опознавательным точкам определяют продолжительность характерных фрагментов сфигмограммы. Кроме того, анализ сфигмограммы позволяет *раздельно* анализировать фазы сердечного цикла правого и левого желудочков, что дополняет результаты ЭКГ-диагностики.

Центральный венный пульс определяется методом *флебографии*. Флебограмма чаще всего регистрируется синхронно с ЭКГ, сфигмограммой сонной артерии и фонокардиограммой (ФКГ) — записью звукового сигнала, порождаемого работой сердца.

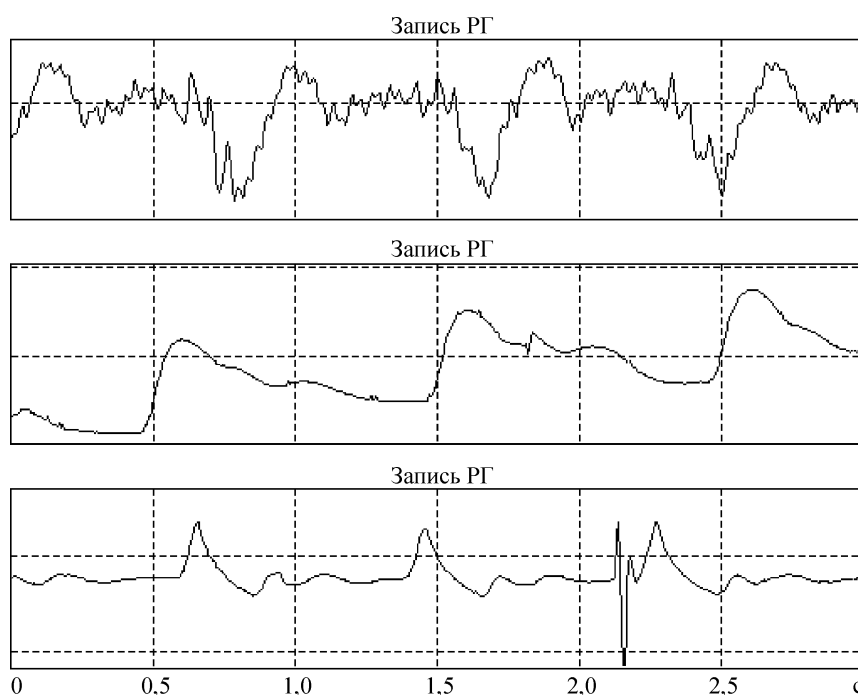


Рис. 2.6. Записи реограмм в реальных клинических условиях



Рис. 2.7. ЭКГ (1) и сфигмограммы сонной (2), лучевой (3), бедренной артерий (4)

Для регистрации флебограммы используют различные бесконтактные датчики (емкостные, пьезоэлектрические, фотоэлектрические и др.), которые накладываются в правой надключичной области пациента.

Флебограмма здорового взрослого человека (рис. 2.8) подобно сфигмограмме состоит из ряда характерных фрагментов (волн), амплитудно-временные параметры которых несут ценную информацию о состоянии правого предсердия.

Магнитокардиография (МКГ) — один из новых методов диагностики в кардиологии, который основан на измерении пара-

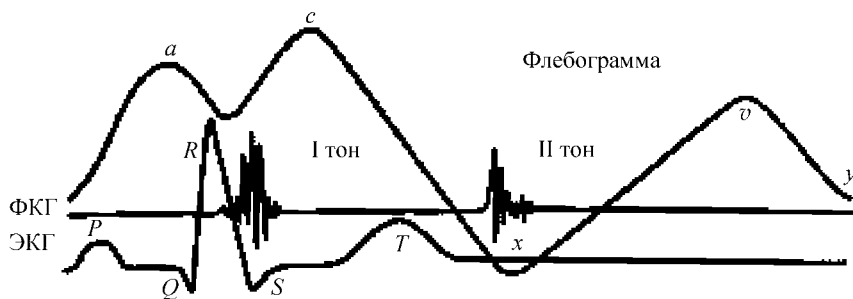


Рис. 2.8. Флебограмма здорового человека

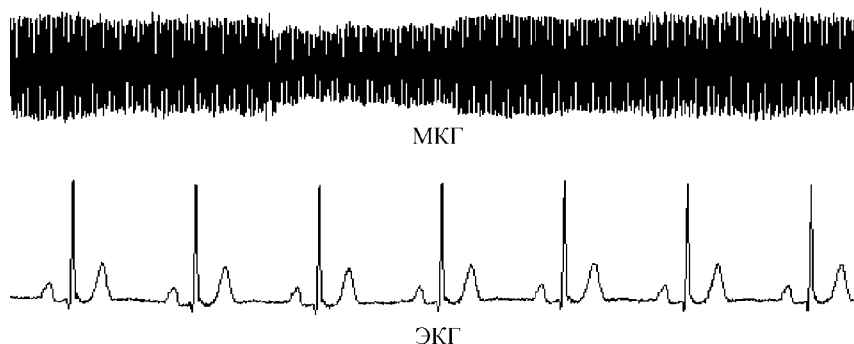


Рис.2.9. Синхронные записи МКГ и ЭКГ в клинических условиях

метров магнитного поля сердца, зарегистрированных над грудной клеткой пациента. Преимущество этого метода состоит в том, что он позволяет диагностировать среди прочих и заболевания, при которых обычно используемые методы, такие как электрокардиография и эхокардиография, не выявляют патологии [156—158, 270].

Источником первичной информации являются сигналы, которые регистрируются сверхчувствительной магнитометрической аппаратурой [183, 366, 367], построенной на основе сверхпроводящих квантовых интерферометрических датчиков (СКВИД). Поскольку магнитокардиограмма (МКГ) и электрокардиограмма (ЭКГ) порождаются одними и теми же источниками, форма МКГ и ЭКГ должна совпадать при их синхронной записи.

Однако непосредственно интерпретировать первичную информацию, полученную от магнитометрических датчиков, невозможно. Во-первых, это связано с наличием большого уровня шума, который существенно искажает полезный сигнал, регистрируемый в реальных клинических условиях (рис. 2.9). Во-вторых, с точки зрения медицинской диагностики более ценную информацию несет не магнитное поле сердца, а характер распределения тока источников, породивших это магнитное поле, а также оценка неоднородной проводимости ткани.

Для получения такой информации необходимо решить так называемую обратную задачу теории стационарного поля. Однако, как известно, такая задача не имеет однозначного решения. В то же время, если сделать определенные предположения о модели источников, то, как показано в работах С.С. Романовича [355, 356], М.А. Прими́на [157] и других исследователей, приемлемую оценку

распределения плотностей тока в плоскости сердца можно получить на основе использования численных методов решения обратной задачи. Но для этого, прежде всего, необходимо построить эффективные алгоритмы, обеспечивающие выделение полезного МКГ-сигнала из существенно искаженной реализации (см. рис. 2.9).

2.2.3. Термическая кривая, несущая информацию о свойствах многокомпонентного материала

Термический анализ широко используют в областях науки и техники для решения самых разнообразных задач — от построения диаграмм плавкости различных систем до анализа грунта на поверхности Марса.

Главные преимущества метода, которые привлекают исследователей — оперативность, высокая информативность отдельного эксперимента при относительной дешевизне аппаратуры. Термический анализ успешно применяют при изучении физико-химических закономерностей образования минералов, горных пород, цветных металлов, полупроводников и других многокомпонентных материалов [110, 150].

Одно из перспективных направлений использования термического анализа в промышленном производстве — комплексный контроль показателей жидкого металла в процессе выплавки. Оказалось, что термограмма кристаллизации пробы серого чугуна несет информацию не только о содержании углерода, но и о целом ряде других химических элементов и позволяет прогнозировать механические свойства металла, в частности прочность на растяжение и твердость [306, 307].

Учитывая актуальность этой задачи, рассмотрим ее более подробно.

Термическая кривая (термограмма) представляет собой график изменения во времени температуры $T(t)$ в некоторой точке температурного поля, образуемого внутри пробы жидкого чугуна. Для получения такой кривой используют высокоточные контактные или бесконтактные измерители температуры.

В отличие от термограммы кристаллизации стали, показанной в правой части рис. 21, термограмму кристаллизации чугуна регистрируют до полного завершения процесса кристаллизации.

Диагностически ценная информация сосредоточена на фрагментах термограммы, отображающих отдельные стадии процесса кристаллизации (рис. 2.10).

Границы информативных фрагментов (точки) соответствуют характерным тепловым эффектам фазовых превращений металла, а именно:

- A — максимальному тепловому эффекту дендритной кристаллизации;
- B — максимальной скорости охлаждения после периода дендритной кристаллизации;
- C — температуре начала кристаллизации эвтектики;
- D — температуре метастабильного превращения расплава;
- E — максимальному тепловому эффекту процесса кристаллизации эвтектики;
- F — температуре солидуса расплава;
- G — максимальному изменению скорости охлаждения в период кристаллизации эвтектики;
- H — температуре окончания кристаллизации эвтектики;
- J — моменту времени, при котором температура T расплава становится равной 1050 °С.

Согласно результатам экспериментальных исследований, приведенным в работах [306, 307], амплитудно-временные параметры указанных фрагментов могут использоваться как аргументы регрессионных моделей, обеспечивающих косвенный контроль состояния металла. Однако практическая реализация этой заманчивой идеи сталкивается с рядом проблем.

Во-первых, информативные фрагменты $T(t)$ не имеют четких границ на реальных сигналах (см. рис. 2.10). Во-вторых, значения амплитудно-временных параметров указанных фрагментов подвержены влиянию случайных возмущений, сопутствующих процессу кристаллизации, в частности, условий теплообмена, толщины изложницы, в которой происходит кристаллизация, и т. п. (рис. 2.11).

Легко заметить, что информативные фрагменты искаженных реализаций имеют неодинаковую продолжительность, хотя приведенные термограммы зарегистрированы при фиксированном состоянии металла. Понятно, что механизм искажения полезного сигнала не может быть сведен лишь к аддитивной помехе.

Поэтому, прежде чем использовать термограмму для оценки состояния расплава, необходимо восстановить полезный сигнал по искаженным реализациям.

Мы рассмотрели только малую часть сигналов с локально сосредоточенными признаками. Несмотря на различие формы информативных фрагментов этих сигналов и физической природы

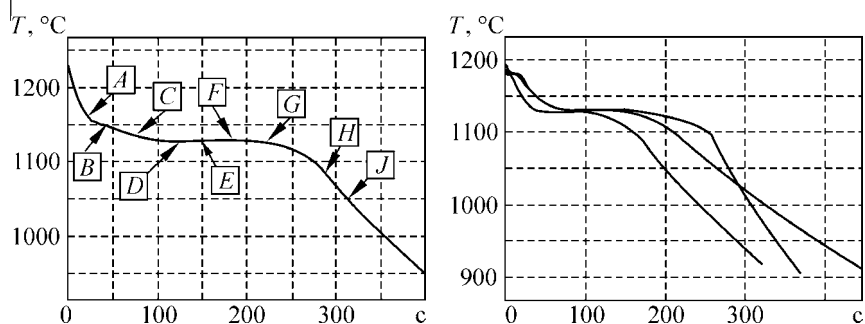


Рис. 2.10. Термограмма кристаллизации пробы серого чугуна

Рис. 2.11. Термограммы кристаллизации расплава из одной пробы ¹

источника их порождения, можно сделать общий вывод: построение эффективных ИТ обработки таких сигналов в первую очередь связано с необходимостью разрешения проблемы восстановления полезного сигнала по искаженным реализациям. При этом следует принимать во внимание тот факт, что области определения (продолжительности) однотипных фрагментов могут не совпадать на разных реализациях даже в тех случаях, когда они порождены при одном и том же состоянии объекта.

Для того чтобы попытаться разрешить эту проблему, нужно построить обобщенную модель порождения сигналов в условиях действия внутренних и внешних возмущений и на основе ее исследования предложить конструктивные вычислительные алгоритмы, обеспечивающие оценку полезного сигнала.

2.3. Модели порождения сигналов с локально сосредоточенными признаками

Рассмотрим один из возможных механизмов порождения реального сигнала $y(t)$ с локально сосредоточенными признаками в условиях действия двух видов возмущений: внутреннего возмущения $\zeta(t)$, которое создает сам объект исследования, и возмущения $h(t)$, создаваемого внешней средой. Такой механизм описывают рассматриваемые далее стохастические модели, предложенные нами в работах [202, 204, 281].

¹ Эксперименты проводились в 1988 г. при участии автора совместно с сотрудниками НПО НИИТракторосельхозмаш [177].

2.3.1. Базовая модель

Предположим, что полезный сигнал $y_0(t)$, соответствующий некоторому текущему состоянию $s \in S$ объекта, представляет собой функцию, кусочно-заданную на области определения $t \in [0, T_0]$ последовательностью $K_F \geq 2$ фрагментов:

$$y_0(t) = \begin{cases} y_0^{(1)}(t), & 0 \leq t < t_0^{(1)}, \\ \dots & \\ y_0^{(K_F)}(t), & t_0^{(K_F-1)} \leq t \leq t_0^{(K_F)} = T_0, \end{cases} \quad (2.3)$$

которые отображают отдельные стадии процесса, происходящего в объекте.

Понятно, что форма и количество указанных фрагментов зависят от природы обрабатываемого сигнала. Например, для ЭКГ фрагменты $y_0^{(i)}(t)$ соответствуют стадиям деполяризации и реполяризации отдельных участков сердца (см. рис. 2.2), а для термограммы — стадиям дендритной кристаллизации, кристаллизации эвтектики, метастабильному превращению и т. п. (см. рис. 2.10).

Будем считать, что наблюдаемый сигнал

$$y(t) = \Phi[y_0(t), \zeta(t)] + h(t) \quad (2.4)$$

является результатом искажения полезного сигнала $y_0(t)$ внутренним $\zeta(t)$ и внешним возмущением $h(t)$ (рис. 2.12).

Предположим, что внутреннее возмущение $\zeta(t)$ приводит к независимым изменениям области определения отдельных фрагментов $y_0^{(i)}(t)$, $i = 1, \dots, K_F$ сигнала (2.3), а сама функция $y_0(t)$ линейно растягивается (сжимается) по амплитуде. Иными словами, допускается, что процесс искажения i -го фрагмента m -й реализации осуществляется на основе операторного преобразования

$$y_m^{(i)}(t) = a_m y_0^{(i)}\left(\frac{t - \tau_m^{(i)}}{b_m^{(i)}}\right), \quad (2.5)$$

где a_m , $b_m^{(i)}$ — соответственно случайные параметры линейного растяжения (сжатия) по амплитуде и времени; $\tau_m^{(i)}$ — сдвиг по времени.

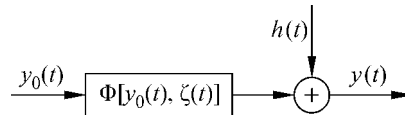


Рис. 2.12. Модель порождения сигнала

Допустим, что для m -й реализации параметр a_m принимает фиксированное значение

$$a_m = 1 + \xi_m, \quad m = 1, 2, \dots, \quad (2.6)$$

где ξ_m — случайная величина, которая с нулевым математическим ожиданием $M\{\xi\} = 0$ распределена на интервале $[-\varepsilon_0, \varepsilon_0]$, ограниченном фиксированным числом $\varepsilon_0 \in [0, 1)$, а параметры $b_m^{(i)}$ принимают фиксированные значения в процессе порождения каждого i -го фрагмента m -й реализации:

$$b_m^{(i)} = 1 + \delta_m^{(i)}, \quad m = 1, 2, \dots, \quad i = 1, \dots, K_F, \quad (2.7)$$

где $\delta_m^{(i)}$ — последовательность независимых случайных величин, которые с нулевым математическим ожиданием $M\{\delta\} = 0$ распределены на интервалах $[-\Delta_0^{(i)}, \Delta_0^{(i)}]$, ограниченных фиксированными числами $\Delta_0^{(i)} \in [0, 1)$.

При таких предположениях продолжительность i -го фрагмента m -й реализации связана с продолжительностью $\Delta T_0^{(i)} = t_0^{(i)} - t_0^{(i-1)}$ соответствующего фрагмента сигнала (2.3) соотношением $\Delta T_m^{(i)} = (t_0^{(i)} - t_0^{(i-1)})(1 + \delta_m^{(i)})$, а началу i -го фрагмента m -й реализации будет соответствовать момент времени

$$t_m^{(i-1)} = \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(l-1)})(1 + \delta_m^{(l)}). \quad (2.8)$$

Применим к i -му фрагменту $y_0^{(i)}(t)$ операторное преобразование (2.5), положив параметр сдвига $\tau_m^{(i)} = t_m^{(i-1)} - (1 + \delta_m^{(i)}) t_0^{(i-1)}$. Тогда из (2.5), с учетом соотношений (2.4) и (2.6)—(2.8), следует, что модель порождения i -го фрагмента m -й реализации можно представить в виде функции

$$y_m^{(i)}(t) = (1 + \xi_m) y_0^{(i)}(\theta) + h(t), \quad m = 1, 2, \dots, \quad i = 1, \dots, K_F, \quad (2.9)$$

аргумент которой θ нелинейно зависит от времени и определяется выражением

$$\theta = \frac{t + (1 + \delta_m^{(i)}) t_0^{(i-1)} - \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(l-1)})(1 + \delta_m^{(l)})}{1 + \delta_m^{(i)}}. \quad (2.10)$$

2.3.2. Модель порождения циклических сигналов

Повторяющиеся во времени процессы часто протекают в технических и биологических системах. Такие процессы порождают специфические сигналы, которые в научной литературе принято называть циклическими [310] или квазипериодическими [380].

Мы будем пользоваться первым названием, поскольку термин «квазипериодический сигнал» легко отождествить с принятым в математике более узким определением почти периодической функции времени [117, с. 275], однако реальные процессы, о которых пойдет речь далее, не всегда удовлетворяют этому определению.

Типичными примерами циклических сигналов являются электрокардиограмма (ЭКГ) и другие физиологические сигналы, отражающие циклический характер работы системы кровообращения живого организма.

Покажем, что рассмотренный механизм действия внутренних возмущений, которые приводят к неравномерным изменениям области определения (продолжительностей) отдельных фрагментов $y_0^{(1)}, \dots, y_0^{(K_F)}$ искаженных реализаций, может быть естественным образом обобщен для описания циклических процессов.

Будем считать, что функция (2.3) — часть ненаблюдаемой периодической функции на любом из периодов, а наблюдаемый сигнал $y(t)$ — последовательность функций $y_1(t), y_2(t), \dots, y_m(t), \dots$, связанных с эталоном $y_0(t)$ операторным соотношением (2.5), в котором, как и прежде, параметры a_m и $b_m^{(i)}$ определяются соотношениями (2.6) и (2.7).

Для обеспечения непрерывности порождаемого сигнала потребуем дополнительно, чтобы $y_0(0) = 0$. Последнее требование всегда можно обеспечить, выполнив предварительное центрирование эталона $\tilde{y}_0(t) = y_0(t) - y_0(0)$.

Видно, что момент времени $t_m^{(i-1)}$, соответствующий началу i -го фрагмента m -го цикла, определяется соотношением (2.8). Следовательно, общая продолжительность m -го цикла порождаемого сигнала $y(t)$ составляет

$$T_m = T_0 + \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_m^{(i)},$$

началу m -го цикла соответствует момент времени

$$t_m = (m-1) T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_j^{(i)},$$

а началу i -го фрагмента m -го цикла — момент времени

$$t_m^{(i-1)} = (m-1)T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_j^{(i)} + \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(l-1)})(1 + \delta_m^{(l)}). \quad (2.11)$$

Если теперь применить к i -му фрагменту эталона $y_0(t)$ операторное преобразование (2.5), положив параметр сдвига $\tau_m^{(i)} = t_m^{(i-1)} - (1 + \delta_m^{(i)}) t_0^{(i-1)}$, то в силу (2.3) с учетом (2.5)—(2.8) следует, что процесс порождения i -го фрагмента на m -м цикле можно представить функцией (2.9), аргумент которой θ по-прежнему нелинейно зависит от времени и уже определяется более сложным выражением

$$\theta = \frac{t - (m-1) T_0 + (1 + \delta_m^{(i)}) t_0^{(i-1)} - \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_j^{(i)} - \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(l-1)})(1 + \delta_m^{(l)})}{1 + \delta_m^{(i)}}. \quad (2.12)$$

Предложенная модель, которая описывает неравномерные по времени искажения области определения отдельных фрагментов эталона $y_0(t)$, более пригодна для описания реальных циклических сигналов, в частности ЭКГ, нежели ее упрощенный вариант

$$y_m^{(i)} = (1 + \xi_m) y_0^{(i)} [(t - (m-1)T_0 - T_0 \sum_{j=1}^{m-1} \delta_j)(1 + \delta_m)^{-1}], \quad m = 1, 2, \dots, \quad (2.13)$$

полученный в предположении, что случайный параметр δ , фигурирующий в (2.12), зависит только от номера m цикла, но не зависит от номера i фрагмента.

Несложно показать, что стохастическая модель (2.9), (2.12) является прямым обобщением известных моделей строго периодического и почти периодического процессов. Действительно, если положить в (2.12)

$$\delta_j^{(i)} = 0 \quad \forall j = 1, \dots, m, i = 1, \dots, K_F,$$

то модель (2.9) можно представить в виде соотношения

$$|y_m[t] - y_0[t - (m-1)T_0]| \leq \varepsilon_0, \quad m = 1, 2, \dots,$$

которое описывает почти периодический процесс, а при дополнительном условии $\varepsilon_0 = 0$ модель сводится к строго периодической функции

$$y[t - (m-1)T_0] = y_0[t].$$

С другой стороны, предложенная модель легко может быть обобщена для описания процесса порождения более сложных циклических сигналов, в частности ЭКГ с изменяющейся морфологией отдельных циклов (экстрасистолами). Для этого достаточно рассмотреть не один, а $G > 1$ эталонов $y_{01}(t), \dots, y_{0G}(t)$ и предположить, что каждый m -й цикл порождается аналогичными искажениями одного из этих эталонов, выбираемых случайным образом в соответствии с вероятностями P_g , $\sum_{g=1}^G P_g = 1$.

2.3.3. Генератор циклических последовательностей

Рассмотрим алгоритм генерации дискретных циклических последовательностей по заданным эталонам.

Пусть есть $G > 1$ эталонов $y_{0g}(t)$, ($g = 1, \dots, G$), каждый из которых представлен конечным числом N_g дискретных значений $y_{0g}[1], \dots, y_{0g}[N_g]$, зафиксированных с постоянным шагом квантования по времени. Зададим также общее число $K_F^{(g)}$ фрагментов каждого эталона и номера точек $N_g^{(i)} \in [1, N_g]$, $i = 1, \dots, K_F^{(g)} - 1$, которые определяют границы i -го и $i + 1$ -го фрагмента g -го эталона.

При таких исходных данных процедура генерации циклической последовательности сводится к следующим шагам.

Шаг 1. Задаем число M циклов генерируемой последовательности.

Шаг 2. Определяем число M_g циклов, порождаемых g -м эталоном, по формуле $M_g = R[P_g M]$, где здесь и далее $R[\eta]$ — операция округления до целого числа η .

Шаг 3. Выбираем номер g_m эталона, порождающего m -й цикл ($m = 1, \dots, M$), по значению реализации λ_m целочисленной случайной величины λ , распределенной на интервале $[1, G]$, т. е. $g_m = \lambda_m$.

Шаг 4. Если $M_g = 0$, то повторяем шаг 3.

Шаг 5. Определяем число точек i -го фрагмента m -го цикла по формуле

$$N_m^{(i)} = R[(N_g^{(i)} - N_g^{(i-1)})(1 + \delta_m^{(i)}),$$

где $\delta_m^{(i)}$ — реализация случайной величины δ , которая с нулевым математическим ожиданием распределена на интервале $[-\Delta_0, \Delta_0]$, $\Delta_0 \in [0, 1)$.

Шаг 6. По дискретным значениям i -го фрагмента g -го эталона в $N_g^{(i)} - N_g^{(i-1)}$ узлах любым из методов интерполяции вычисляем значения генерируемой последовательности в $N_m^{(i)}$ точках.

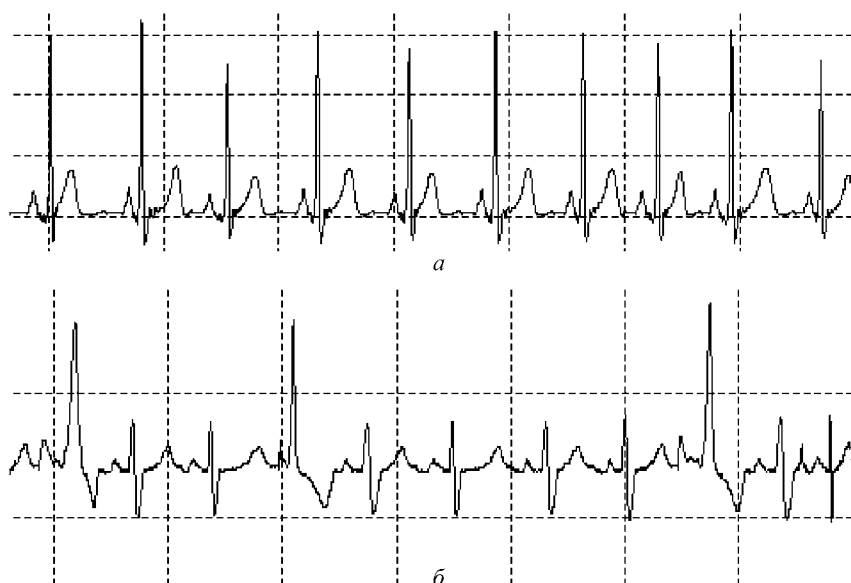


Рис. 2.13. Искусственные ЭКГ, порожденные моделью (2.9): по одному эталону (а); по двум эталонам (б)

Шаг 7. Модифицируем каждое вычисленное значение y на основе мультипликативной процедуры $\tilde{y} = y(1 + \xi_m)$, где ξ_m — реализация случайной величины ξ , которая с нулевым математическим ожиданием распределена на интервале $[-\varepsilon_0, \varepsilon_0]$, $\varepsilon_0 \in [0, 1)$.

Шаг 8. Если $i < K_F^{(g)}$, то возвращаемся к шагу 5.

Шаг 9. Присваиваем $M_g := M_g - 1$.

Шаг 10. Если $m < M$, то возвращаемся к шагу 3.

На рис. 2.13 представлены примеры искусственных ЭКГ, порожденных по одному и двум эталонам. Опытные врачи-кардиологи, которым наряду с реальными записями ЭКГ предъявляли искусственные ЭКГ, сгенерированные по различным эталонам, признавали их реальными. Это дало нам основание использовать предложенную модель для проведения дальнейших исследований.

2.4. Декомпозиция общей задачи компьютерной обработки сигналов с локально сосредоточенными признаками

Предложенные генеративные модели порождения сигналов с локально сосредоточенными признаками позволяют конкретизировать проблему построения ИТ обработки сигналов с локально сосредоточенными признаками, проведя последовательную декомпозицию общей задачи таких технологий [214].

Предположим, что в нашем распоряжении есть конечная последовательность искаженных реализаций $y_1(t), \dots, y_M(t)$, порожденных моделью (2.9) при фиксированном состоянии $s \in S$ объекта.

При реализации ИТ обработки термограмм кристаллизации металла, наблюдаемых в реальных условиях (см. рис. 2.11), исходным сырьем служит совокупность искаженных сигналов $T_1(t), \dots, T_M(t)$, которые зарегистрированы при фиксированном состоянии s объекта (металла), а при построении компьютерных систем обработки физиологических сигналов (ЭКГ, РГ, МКГ и т. п.) — последовательность искаженных циклов, наблюдаемых при текущем состоянии s пациента.

Тогда обобщенная схема формирования информационного продукта (оценки s) по искаженным реализациям $y_1(t), \dots, y_M(t)$

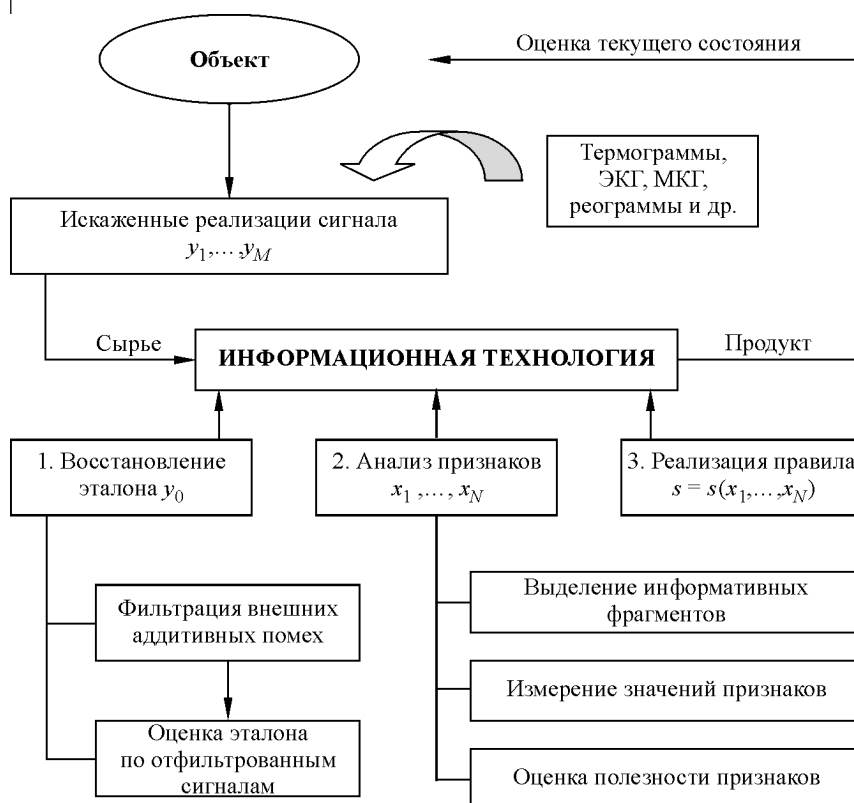


Рис. 2.14. Обобщенная схема формирования информационного продукта

может быть сведена к последовательному решению трех отдельных задач:

- восстановление полезного сигнала (эталона) $y_0(t)$ по искаженным реализациям;
- выделение информативных фрагментов эталона и анализ значений признаков $x_1, \dots, x_N$, сосредоточенных на информативных фрагментах.
- оценка текущего значения s по модели $\hat{s} = s(x_1, \dots, x_N)$ (рис. 2.14).

Проведенный в главе 1 анализ существующих подходов к построению моделей сложных систем по экспериментальным данным показывает, что в распоряжении конструктора прикладной

ИТ есть широкий арсенал универсальных средств, позволяющих по выборке наблюдений восстановить зависимость $\hat{s} = s(x_1, \dots, x_N)$, если, конечно, абстрагироваться от упомянутой проблемы выбора системы базисных функций $\varphi_1(x^{(N)}), \dots, \varphi_K(x^{(N)})$. Поэтому основное внимание дальнейших исследований будет направлено на решение первой и второй задач.

В соответствии с (2.4) задачу восстановления $y_0(t)$ по искаженным реализациям $y_1(t), \dots, y_M(t)$ можно подвергнуть дальнейшей декомпозиции, рассматривая ее решение в два этапа: подавление внешних аддитивных помех $h(t)$ и оценка $y_0(t)$ по реализациям $y_1(t), \dots, y_M(t)$, наблюдаемым в условиях действия внутреннего возмущения $\zeta(t)$, предполагая, что $h(t) \equiv 0$.

Решение второй, не менее важной, задачи анализа признаков $x_1, \dots, x_N$, сосредоточенных на фрагментах восстановленного эталона $y_0(t)$, вообще говоря, зависит от объема априорных знаний в соответствующей предметной области. Задача существенно упрощается, когда специалистам-прикладникам известны свойства $y_0(t)$, обладающие диагностической ценностью. Например, в кардиологической практике уже давно сформировался традиционный набор диагностических признаков ЭКГ, связанных с тем или иным заболеванием.

Но даже в тех случаях, когда набор диагностических признаков заранее известен, остается проблема построения эффективных вычислительных процедур, обеспечивающих компьютерный анализ признаков. Чаще всего на практике приходится искать приемлемые алгоритмы, действуя методом «проб и ошибок», многократно совершенствуя и проверяя эффективность вычислительных процедур на имеющихся экспериментальных данных.

Еще большие проблемы возникают, когда знаний в предметной области не достаточно для априорного выбора словаря диагностических признаков. Такая ситуация характерна для новых областей знаний, например для магнитокардиографии, с пока еще достаточно скромным клиническим материалом.

Да и в таких уже «устоявшихся» областях знаний, как электрокардиография, клиницисты непрерывно ведут поиски новых диагностических признаков, обеспечивающих повышение чувствительности и специфичности результатов диагностики.

Решение такой уже более сложной задачи предполагает:

- формирование гипотез о свойствах $y_0(t)$, которые можно считать потенциально полезными признаками;
- разработка вычислительных алгоритмов, пригодных для определения этих свойств по реальным сигналам;
- оценка диагностической эффективности каждого такого свойства в отдельности и в совокупности друг с другом с тем, чтобы расширение пространства признаков было обоснованным.

Таким образом, на основании проведенного анализа задач, связанных с компьютерной обработкой сигналов с локально сосредоточенными признаками, можно сделать следующие общие выводы.

1. Неправомерно сводить проблему построения ИТ обработки временных сигналов с признаками, сосредоточенными на локальных фрагментах, лишь к решению задачи идентификации неизвестной модели по таблице наблюдений, в которой потенциальными регрессорами являются дискретные значения сигнала.

2. Построение эффективных ИТ обработки таких сигналов в первую очередь связано с необходимостью разработки эффективных методов восстановления полезного сигнала по реализациям, искаженным внутренними и внешними возмущениями.

3. Предложенная стохастическая модель порождения сигнала с локально сосредоточенными признаками в условиях действия внутренних и внешних возмущений позволяет обосновать обобщенную схему построения ИТ обработки таких сигналов, предусматривающую формирование информационного продукта в три этапа (см. рис. 2.14).

Детальному изложению методов, положенных в основу такой обобщенной схемы, а также практическим результатам применения этой схемы для построения эффективных ИТ обработки сигналов разной физической природы посвящены последующие главы монографии.

**3.1. Метод усреднения реализаций
во временной области**

Предположим, что в нашем распоряжении есть $M > 2$ искаженных реализаций $y_1(t), \dots, y_M(t)$, порожденных в соответствии с моделью (2.9) по эталону $y_0(t)$, который представляет собой последовательность K_F фрагментов $y_0^{(1)}(t), \dots, y_0^{K_F}(t)$. Требуется получить такую оценку $\hat{y}_0(t)$ эталона $y_0(t)$ по наблюдениям $y_1(t), \dots, y_M(t)$, чтобы амплитудно-временные параметры $x_1, \dots, x_N$ (диагностические признаки), характеризующие форму отдельных фрагментов $\hat{y}_0(t)$, были максимально близкими к соответствующим параметрам, характеризующим форму фрагментов эталона.

Понятно, что такая цель может быть достигнута лишь в том случае, когда процедура оценки эталона позволит в максимальной степени сохранить форму фрагментов $y_0^{(1)}(t), \dots, y_0^{K_F}(t)$.

Будем считать, что наблюдения представлены дискретными значениями $y_m[k] \equiv y_m(t_k)$, $m = 1, \dots, M$, зафиксированными в равноотстоящие моменты времени $t_k \equiv k\Delta$, $k = 1, \dots, K_m$, где Δ — шаг квантования по времени.

Вначале рассмотрим частный случай, предполагая, что внешнее возмущение $h(\cdot)$ отсутствует, а внутреннее возмущение приводит лишь к изменению амплитуд $y_m[k] = (1 + \xi_m)y_0[k]$, в то время как продолжительности отдельных фрагментов $y_0^{(1)}(\cdot)$, ..., $y_0^{K_F}(\cdot)$ сохраняются неизменными. Для этого достаточно в общей модели (2.9) положить $h(\cdot) \equiv 0$ и $\delta_j^{(i)} = 0$, $j = 1, \dots, m$, $i = 1, \dots, K_F$.

Поскольку модель (2.9) предполагает, что мультипликативная помеха ξ_m имеет нулевое математическое ожидание $M\{\xi\} = 0$, причем в рассматриваемом случае области определения реализаций $y_1(\cdot)$, ..., $y_M(\cdot)$ совпадают, то для получения несмещенной оценки $\hat{y}_0(k)$ достаточно просто усреднить соответствующие дискретные значения $y_1(k)$, ..., $y_M(k)$ наблюдаемых реализаций:

$$\hat{y}_0(k) = \frac{1}{M} \sum_{m=1}^M y_m(k). \quad (3.1)$$

Понятно, что оценка $\hat{y}_0(k)$ тем лучше, чем больше M .

На рис. 3.1, а показан результат оценки эталона во временной области по последовательности $M = 20$ искаженных реализаций, полученных по модели (2.9) при $h(\cdot) \equiv 0$, $\delta_j^{(i)} = 0$, $j = 1, \dots, m$, $i = 1, \dots, K_F$ и ограничении на уровень мультипликативной помехи $\varepsilon = 0,5$.

Усреднение наблюдений во временной области позволяет получить приемлемую оценку $y_0(t)$ и в том случае, когда, помимо мультипликативной помехи ξ_m , на эталон воздействует аддитивная помеха $h(\cdot)$ с нулевым математическим ожиданием $M\{h\} = 0$ и ограниченной дисперсией $D\{h\}$. При статистической независимости h и ξ оценка (3.1) будет также эффективной и позволит в M раз уменьшить дисперсию помехи.

Рис. 3.1, б иллюстрирует результат усреднения реализаций в условиях действия случайной помехи $h(t)$, уровень которой составлял 10 % диапазона изменения $y_0(t)$.

Покажем, что процедура (3.1) может быть использована и для оценки полезного сигнала по наблюдениям, порождаемым

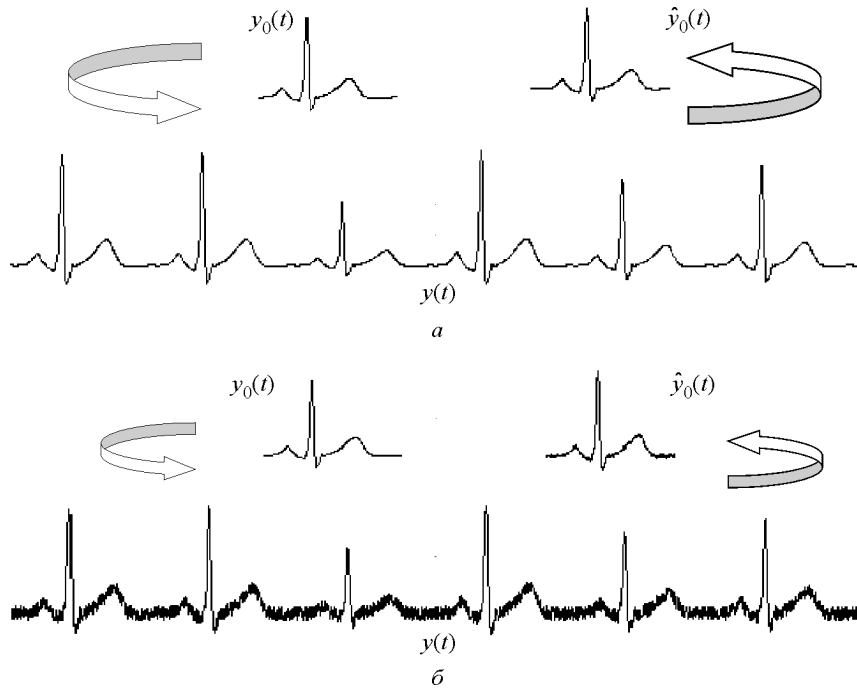


Рис. 3.1. Восстановление эталона y_0 по искаженной последовательности циклов ЭКГ, порожденных моделью (2.9):

$a - \delta_j^{(i)} = 0, h(t) \equiv 0$; $b - \delta_j^{(i)} = 0, h_0 = 10\%$

моделью (2.13), которая, в отличие от общей модели (2.9), предполагает, что случайный параметр δ искажения продолжительностей отдельных фрагментов эталона зависит только от номера m реализации, но не зависит от номера i фрагмента.

Понятно, что в этом случае продолжительности $T_1, \dots, T_M$ усредняемых реализаций будут различны. Однако, используя методы линейной интерполяции, можно модифицировать имеющиеся данные таким образом, чтобы и в этом случае процедура (3.1) была эффективной. Для этого достаточно определить среднюю продолжительность $T_0 = 1/M \sum_{m=1}^M T_m$ усредняемых реализаций и по дискретным значениям $y_m(k_m)$ на интервалах T_m вычислить мо-

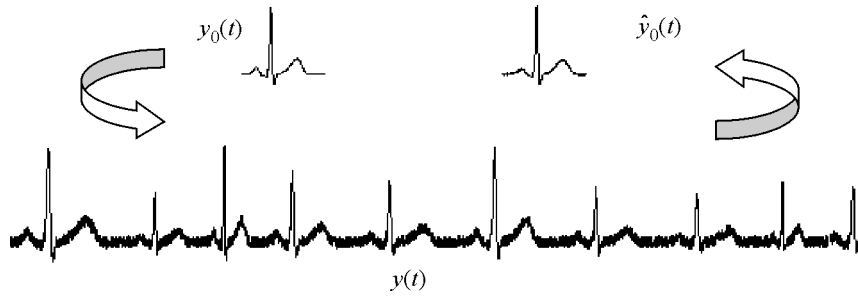


Рис. 3.2. Восстановление эталона y_0 по сигналу, порожденному моделью (2.13) с параметрами $\Delta_m = 0,9$, $\varepsilon = 0,5$, $h_0 = 10\%$

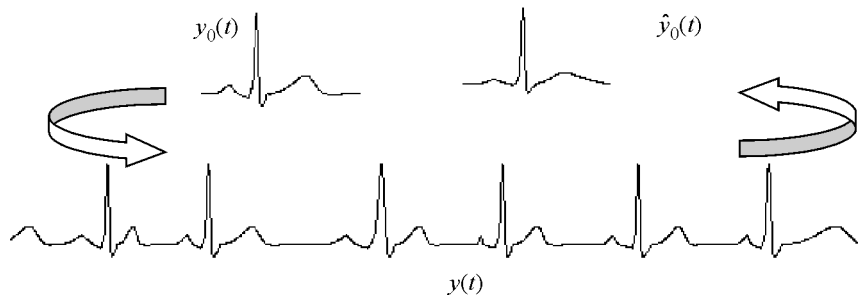


Рис. 3.3. Восстановление эталона y_0 по сигналу, порожденному моделью (2.9) с параметрами $\Delta_m^{(i)} = 0,9$, $\varepsilon = 0$, $h_0 = 0$

дифицированные значения $\tilde{y}_m(k_m)$ на интервале T_0 , а далее использовать в процедуре (3.1) $\tilde{y}_m(k_m)$ вместо $y_m(k_m)$.

Результат модельных экспериментов, реализующий такой подход, приведен на рис. 3.2.

Мы рассмотрели лишь частные случаи модели (2.9), которые, однако, не адекватны многим реальным сигналам. Как уже отмечалось, при регистрации реальных ЭКГ наблюдаются неравномерные изменения области определения фрагментов внутри каждого из циклов.

Понятно, что в этом общем случае использование простейшего алгоритма линейной интерполяции не позволит корректно совместить дискретные значения $y_m(k_m)$ различных реализаций,

соответствующие k -му значению эталона $y_0(k)$, и усреднение таких точек по формуле (3.1) неизбежно приведет к недопустимым искажениям (сглаживанию) формы информативных фрагментов эталона (рис. 3.3).

Даже визуальная оценка полученного результата показывает, что при неравномерных изменениях продолжительности фрагментов их форма на восстановленном сигнале $\hat{y}_0(t)$ отличается от формы соответствующих фрагментов эталона $y_0(t)$. В частности, существенно искажена (сглажена) форма сегмента $S-T$, амплитудно-временные параметры которого, как известно [137, с. 207], являются основными диагностическими признаками ишемической болезни сердца (ИБС).

Совершенно очевидно, что в общем случае, когда наблюдаются неравномерные искажения продолжительностей отдельных фрагментов, для корректного усреднения реализаций $y_1(\cdot), \dots, y_M(\cdot)$ необходимо совместить во времени однотипные фрагменты. Для этого требуется, прежде всего, провести морфологический анализ каждой из усредняемых реализаций и разделить ее на отдельные фрагменты.

Однако реализовать такой алгоритм не так уж просто, поскольку информативные фрагменты искаженных реализаций, в частности циклы реальных ЭКГ, чаще всего не имеют четких границ [37].

Поэтому рассмотрим альтернативный метод, предложенный нами в работе [204], который основан на усреднении искаженных реализаций в фазовом пространстве координат.

В отличие от рассмотренных выше методов, предлагаемый метод не требует предварительного выделения фрагментов усредняемых реализаций $y_1(t), \dots, y_M(t)$. В то же время, как будет показано дальше, он позволяет получить приемлемую оценку $\hat{y}_0(t)$ эталона $y_0(t)$, пригодную для последующего морфологического анализа и определения значений диагностических признаков, сосредоточенных на информативных фрагментах $\hat{y}_0(t)$.

3.2. Метод усреднения реализаций в фазовом пространстве

3.2.1. Основная идея метода

Вернемся к базовой модели порождения сигнала (2.9), полагая, что $h(t) \equiv 0$, т. е. внешнее возмущение (аддитивная помеха) отсутствует или предварительно отфильтровано.¹

Как и прежде, будем считать, что в нашем распоряжении есть выборка $M \geq 2$ искаженных реализаций $y_1(t), \dots, y_M(t)$, каждая из которых представлена дискретными значениями $y_m[k] \equiv y_m(t_k)$, зафиксированными в равноотстоящие моменты времени $t_k \equiv k\Delta$, $k = 1, \dots, K_m$, где Δ — шаг квантования по времени.

Оценим (используя алгоритмы численного дифференцирования) производные $\dot{y}_m[k]$ в каждой k -й точке m -й реализации. Далее выполним нормировку:

$$y_m^*[k] = \frac{y_m[k] - \min_{1 \leq k \leq K_m} y_m[k]}{\max_{1 \leq k \leq K_m} y_m[k] - \min_{1 \leq k \leq K_m} y_m[k]}, \quad k = 1, \dots, K_m,$$

$$\dot{y}_m^*[k] = \frac{\dot{y}_m[k] - \min_{1 \leq k \leq K_m} \dot{y}_m[k]}{\max_{1 \leq k \leq K_m} \dot{y}_m[k] - \min_{1 \leq k \leq K_m} \dot{y}_m[k]}, \quad k = 1, \dots, K_m.$$

В результате вместо исходных последовательностей скалярных величин $y_m[1], \dots, y_m[K_m]$, соответствующих амплитудам наблюдаемых реализаций в дискретные моменты времени $k = 1, \dots, \dots, K_m$, получим M последовательностей

$$Q_1 = \{z_1[k], k = 1, \dots, K_1\}, \dots, Q_M = \{z_M[k], k = 1, \dots, K_M\} \quad (3.2)$$

нормированных векторов $z_m[k] = (y_m^*[k], \dot{y}_m^*[k])$, $m = 1, \dots, M$, причем $y_m^*[k] \in [0, 1]$ и $\dot{y}_m^*[k] \in [0, 1]$. Последовательности $Q_1, \dots, Q_M$

¹ Методы фильтрации внешних помех будут рассмотрены в гл. 4. Там же будет описана процедура численного дифференцирования, используемая для практической реализации предлагаемого метода.

принадлежат траекториям искаженных сигналов в двумерном фазовом пространстве с координатами $y^* - \dot{y}^*$.

Понятно, что каждой m -й последовательности будет соответствовать своя траектория в фазовом пространстве, которая за счет действия внутреннего возмущения ξ отклоняется от ненаблюдаемой траектории полезного сигнала $y_0(\cdot)$. Однако поскольку случайные величины $\delta_m^{(i)}$ и ξ_m , фигурирующие в (2.9), (2.10), ограничены, имеют нулевые математические ожидания и независимы, то для любой фиксированной точки $\theta = \theta_x$ области определения $y_0(\cdot)$ координаты $y_m(\theta_{mx})$, $\dot{y}_m(\theta_{mx})$ соответствующих точек $\theta_{1x}, \dots, \theta_{Mx}$ наблюдаемых сигналов будут группироваться в локальной области нормированного фазового пространства (рис. 3.4) с центром в точке $X = y_0(\theta_x), \dot{y}_0(\theta_x)$ и радиусом $\rho_x = \rho(\theta_x)$, удовлетворяющим соотношению

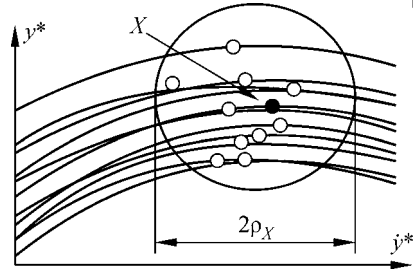


Рис. 3.4. Иллюстрация идеи метода

$$\rho^2(\theta_x) = y_0^2(\theta_x) \xi_m^2 + \dot{y}_0^2(\theta_x) \left[\frac{\delta_m^{(i)} - \xi_m}{1 + \delta_m^{(i)}} \right]^2,$$

причем математическое ожидание отклонений $y_m(\theta_{mx}), \dot{y}_m(\theta_{mx})$ от истинных значений $y_0(\theta_x), \dot{y}_0(\theta_x)$ равно нулю.

Этот факт дает ключ к построению метода оценки полезного сигнала $y_0(\cdot)$, основанного на усреднении наблюдаемых фазовых траекторий.

Рассмотрим более подробно алгоритмы, направленные на реализацию этой идеи.

3.2.2. Базовый алгоритм усреднения траекторий

Пусть $Q = \{Q_1, \dots, Q_M\}$ — множество последовательностей (3.2). Вычислим $M \times M$ матрицу $D = \|R_H(Q_i, Q_j)\|$ расстояний между всеми парами $Q_i \in Q$ и $Q_j \in Q$, $i, j = 1, \dots, M$.

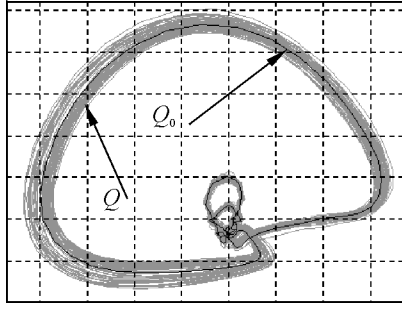


Рис. 3.5. Опорная траектория Q_0 в фазовом пространстве

Заметим, что в общем случае количество элементов i -й и j -й последовательностей неодинаково. Для построения матрицы $D = \|R_H(Q_i, Q_j)\|$ воспользуемся хаусдорфовой метрикой [178]:

$$R_H(Q_i, Q_j) = \max \left\{ \max_{z_j \in Q_j} \min_{z_i \in Q_i} \rho(z_i, z_j), \max_{z_i \in Q_i} \min_{z_j \in Q_j} \rho(z_i, z_j) \right\}, \quad (3.3)$$

где $\rho(z_i, z_j) = \|z_i - z_j\|$ — евклидово расстояние между точками $z_i \in Q_i$ и $z_j \in Q_j$.

При этом номер строки матрицы $D = \|R_H(Q_i, Q_j)\|$, сумма элементов которой минимальна, определит последовательность $Q_0 \in Q$ нормированных векторов $z[k] = (y^*[k], \dot{y}^*[k])$, принадлежащих фазовой траектории одного из циклов. Будем называть эту последовательность

$$Q_0 = \arg \min_{1 \leq j \leq M} \sum_{i=1}^M R_H(Q_i, Q_j) \quad (3.4)$$

опорной траекторией в нормированном фазовом пространстве.

Поскольку в соответствии с (3.4) среднее расстояние от опорной траектории до остальных траекторий минимально, траектория Q_0 расположена «внутри» множества $Q = \{Q_1, \dots, Q_M\}$ (рис. 3.5).

Будем использовать траекторию Q_0 в качестве первого приближения ненаблюдаемой траектории эталонного сигнала, а для получения более точной оценки $\hat{y}_0(\cdot)$ усредним точки наблюдаемых траекторий в окрестности соответствующих точек опорной траектории Q_0 .

С этой целью вычислим последовательность векторов $\hat{z}[k]$, $k = 1, \dots, K_0$, по формуле

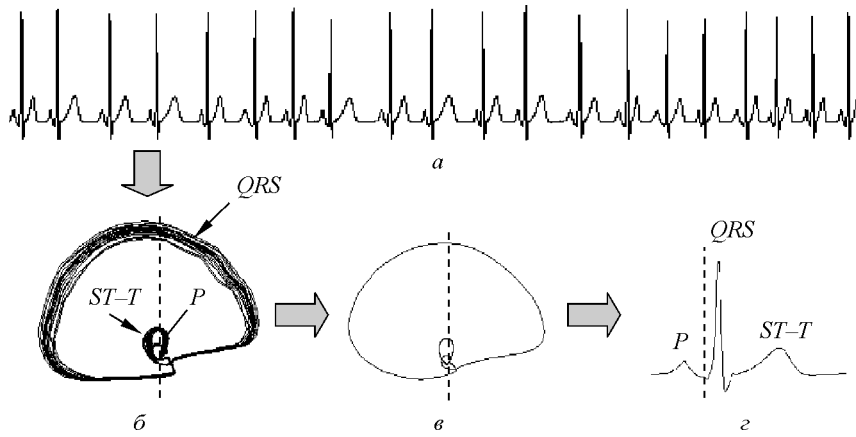


Рис. 3.6. Последовательность этапов обработки ЭКГ:

a — исходная ЭКГ; b — ее фазовая траектория; c — усредненная фазовая траектория; d — эталонный цикл ЭКГ во временной области

$$\hat{z}[k] = \frac{z_0[k] + \sum_{m=1}^M z_{mk}^{\min}}{M+1}, \quad (3.5)$$

где $z_{mk}^{\min}$ — вектор, принадлежащий m -й траектории (не являющейся опорной) и находящийся на минимальном евклидовом расстоянии

$$\rho = \sqrt{(y_0^*[\cdot] - y_m^*[\cdot])^2 + (\dot{y}_0^*[\cdot] - \dot{y}_m^*[\cdot])^2} \quad (3.6)$$

от вектора $z_0[k] \in Q_0$; K_0 — число точек на опорной траектории.

Последовательность векторов $\hat{z}[k] = (\hat{y}^*[k], \dot{\hat{y}}^*[k])$, $k = 1, \dots, K_0$, вычисленная согласно (3.5), принадлежит средней траектории в нормированном фазовом пространстве, а соответствующая последовательность значений $\hat{y}[1], \dots, \hat{y}[K_0]$ дает оценку эталона $y_0(t)$ во временной области (рис. 3.6).

3.2.3. Модифицированный алгоритм усреднения

При построении опорного цикла мы использовали классическую хаусдорфову метрику (3.3), которая, вообще говоря, предназначена для определения расстояния между двумя множествами

точек. В то же время нас интересует расстояние между двумя *последовательностями* точек. К тому же в общем случае фазовые координаты $y^*, \dot{y}^*$ любой из наблюдаемых реализаций могут совпадать в разные моменты времени, а значит, ближайшие по расстоянию (3.6) точки могут принадлежать разным фрагментам.

Поэтому для улучшения оценки эталона предлагается представить опорную траекторию Q_0 и каждую из остальных $M - 1$ траекторий $Q_m \in Q$ последовательностью расширенных векторов $u[k] = (z[k], \tau[k])$, которые, помимо нормированных фазовых координат $z[k] = (y^*[k], \dot{y}^*[k])$, содержат дополнительную компоненту $\tau[k]$, имеющую смысл относительного времени появления k -й точки на m -й траектории.

Величину $\tau[k]$ определим по формуле $\tau[k] = (k - 1) / (K_m - 1)$, где K_m — число точек m -й траектории. Заметим, что величина $\tau[k]$ нормирована в тех же пределах, что и компоненты $y^*[k]$, $\dot{y}^*[k]$ векторов $z[k]$, т. е. $\tau[k] \in [0, 1]$.

Введение дополнительной компоненты $\tau[k] \in [0, 1]$ позволяет оценивать близость усредняемых траекторий не только с точки зрения значений фазовых координат $y^*[k], \dot{y}^*[k]$ (которые, вообще говоря, могут совпадать для различных моментов времени), но и их синхронности во времени.

Для определения последовательности $\hat{z}[k]$, $k = 1, \dots, K_0$, будем по-прежнему использовать соотношение (3.5), но вместо (3.6) при поиске «ближайших» векторов $z_{mk}^{\min}$ определять вектора, которые находятся уже на минимальном евклидовом расстоянии от расширенного вектора $u_0[k] = (z_0[k], \tau[k])$ опорной траектории Q_0 :

$$z_{mk}^{\min} = \arg \min_l \rho_\tau(u_0[k], u_m[l]), \quad l = 1, \dots, K_m, \quad (3.7)$$

где

$$\rho_\tau(u_0[\cdot], u_m[\cdot]) = \sqrt{(y_0^*[\cdot] - y_m^*[\cdot])^2 + (\dot{y}_0^*[\cdot] - \dot{y}_m^*[\cdot])^2 + (\tau_0[\cdot] - \tau_m[\cdot])^2}. \quad (3.8)$$

Рассмотренная схема усреднения реализаций в фазовом пространстве может быть естественным образом обобщена на общий

случай, когда производится переход от исходных последовательностей скалярных величин $y_m[k]$, $k = 1, \dots, K_m$, к последовательностям нормированных векторов $z^* = (y^*[k], \dot{y}[k], \ddot{y}[k], \dots)$, компоненты которых представляют собой оценки производных наблюдаемых сигналов более высоких порядков. Единственная проблема в данном случае — это устойчивость процедур численного дифференцирования реальных сигналов. Однако, как будет показано далее, при использовании соответствующих процедур регуляризации удастся получить приемлемые оценки производных дискретно заданных реализаций реальных сигналов.

3.3. Алгоритмы обработки циклических сигналов

3.3.1. Процедура разбиения на циклы

Рассмотренный выше метод получения оценки $\hat{y}_0$ эталона $y_0(t)$ может быть применен для обработки циклических сигналов, описываемых моделью (2.9), в которой аргумент θ определяется выражением (2.12). В данном случае используется одна реализация наблюдаемого сигнала, а усреднению подлежат ее отдельные циклы, искаженные внутренним возмущением.

Понятно, что для реализации такого подхода необходимо предварительно разбить наблюдаемый циклический сигнал $y(t)$ на отдельные циклы $y_1(t), \dots, y_M(t)$.

Применительно к задаче обработки ЭКГ это можно сделать, используя известные процедуры обработки сигнала во временной области (*QRS*-детекторы) [286, 338]. Однако, как показывает практика, даже весьма сложные алгоритмы не всегда эффективно справляются с этой задачей из-за многообразия формы *QRS*-комплексов реальных ЭКГ (см. рис. 2.3).

Коль скоро мы отказались от обработки сигнала во временной области, привлекательно было бы и эту задачу решить в фазовом пространстве. Покажем, что такая возможность действительно существует. Основная идея предлагаемого метода состоит в следующем.

Пусть сигнал $y(t)$ представлен последовательностью $y[k]$, $k = 1, \dots, K$, дискретных значений, наблюдаемых в течение $M > 2$ циклов. Следуя базовому алгоритму, рассмотренному в пункте 3.2.1, перейдем от исходной последовательности $y[k]$, $k = 1, \dots, K$, ска-

лярных величин к последовательности нормированных векторов $z^*[k] = (y^*[k], \dot{y}^*[k])$, $k = 1, \dots, K$.

Вычислим квадраты расстояний

$$r_0^2[k] = (y^*[k] - y_0^*)^2 + (\dot{y}^*[k] - \dot{y}_0^*)^2, \quad k = 1, \dots, K, \quad (3.9)$$

между нормированными фазовыми координатами каждой k -й точки и некоторой фиксированной (базовой) точкой $k_0 \in [1, K]$.

Поскольку высокие значения амплитуды y и производной $\dot{y}$ отличают комплекс QRS от других фрагментов ЭКГ, то в моменты появления QRS -комплексов на профиле расстояний $r_0^2[k]$ будут появляться выраженные экстремумы (максимумы) даже в тех случаях, когда наблюдаемый сигнал имеет сложную форму, в частности высокий зубец T (рис. 3.7).

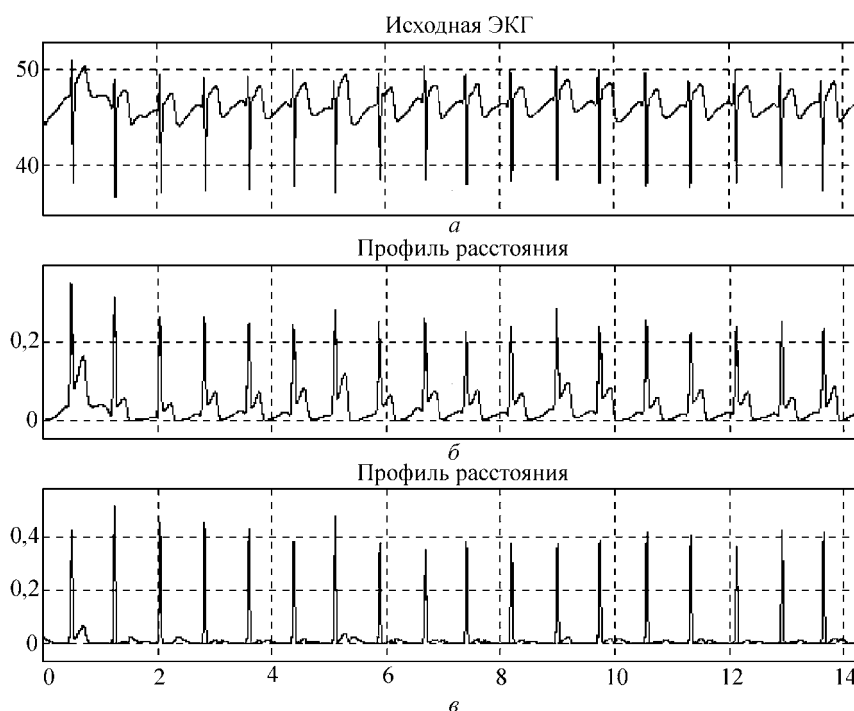


Рис. 3.7. Профили расстояний, построенные для одной и той же ЭКГ:

a — исходный сигнал; $б$ — профиль расстояний (3.9); $в$ — профиль расстояний (3.14)

Тогда, если выделить подмножества Ω_m точек $k \in [1, K]$, таких, что

$$\rho_0^2[k] \geq \rho_{QRS} \quad \forall k \in \Omega_m, \quad (3.10)$$

где ρ_{QRS} — некоторый порог, и определить номера точек k_m , $m = 1, \dots, M$, удовлетворяющие условию

$$k_m = \arg \max_{k \in \Omega_m} \rho_0^2[k], \quad (3.11)$$

то полученная последовательность номеров k_m , $m = 1, \dots, M$, определит границы циклов обрабатываемого сигнала. Тем самым исходная траектория $Q = \{z[k] = (y^*[k], \dot{y}^*[k])\}$ будет разделена на M траекторий $Q_1, \dots, Q_M$, которые могут быть усреднены по описанной выше схеме.

К сожалению, нельзя формальными методами определить оптимальное значение порога ρ_{QRS} , фигурирующего в (3.10), которое обеспечило бы необходимый компромисс между возможными ошибками пропуска цели и ложной тревоги для всего многообразия реальных ЭКГ. Впрочем, такая же ситуация характерна и для традиционных QRS -детекторов [338].

Однако можно «адаптировать» этот параметр к конкретной реализации, используя дополнительный критерий.

Эксперименты показали, что при обработке на современных компьютерах 1–2-минутной записи ЭКГ с частотой дискретизации $F_D = 500$ Гц (в этом случае $K < 60\,000$) оптимальное значение порога можно получить практически мгновенно методом последовательного перебора значений ρ_{QRS} из упорядоченного набора допустимых значений, остановив процедуру перебора на значении, после которого разбиение на циклы сохраняется неизменным.

Заметим, что для фиксированной последовательности $y[1], \dots, y[K]$ профиль квадратов расстояний (3.9) зависит от выбора номера k_0 базовой точки. При указанных значениях K «оптимальный» профиль для обрабатываемой реализации можно также построить методом полного перебора, например, выбрав параметр k_0 функции $r_0^2[k]$ по условию

$$k_0 = \arg \max_{1 \leq k_j \leq K} \left\{ \frac{\max_{1 \leq k \leq K} r[k, k_j]}{\frac{1}{K} \sum_{k=1}^K r[k, k_j]} \right\}. \quad (3.12)$$

В этом случае фрагменты функции $r_0^2[k]$, соответствующие *QRS*-комплексам, становятся более выраженными по сравнению с остальными участками области определения (рис. 3.7, б).

Процедуру выделения *QRS*-комплексов можно сделать еще более эффективной, если предварительно перейти от исходной последовательности наблюдений $y[k]$, $k = 1, \dots, K$, к последовательности величин

$$\tilde{y}[k] = y[k] - \frac{1}{2W} \sum_{i=-W}^W y[k-i], \quad k = W, \dots, K-1-W, \quad (3.13)$$

где W — окно скользящего сглаживания. Поскольку продолжительность *QRS*-комплекса существенно меньше продолжительности зубцов P и T , такой прием позволяет еще больше усилить «контраст» искомым фрагментов (рис. 3.7, в) на профиле квадратов расстояний:

$$\tilde{r}_0^2[k] = (\tilde{y}^*[k] - \tilde{y}_0^*)^2 + (\tilde{y}^*[k] - \tilde{y}_0^*)^2, \quad k = W, \dots, K-1-W. \quad (3.14)$$

3.3.2. Селекция нетипичных циклов

Метод усреднения фазовых траекторий, описанный в параграфе 3.2, строго говоря, можно применять лишь в том случае, когда сигнал $y(t)$ порожден одним эталоном (см. рис. 2.13, а). Если же на сигнале появляются «нетипичные» циклы, иными словами, наблюдаемый сигнал порожден не одним, а несколькими эталонами (см. рис. 2.13, б), то точки траекторий таких циклов не должны использоваться в формуле усреднения (3.5).

Рассмотрим один из возможных методов селекции нетипичных циклов [203], предполагая лишь, что число таких циклов $M_N \ll M$ (в противном случае определение «нетипичный» цикл теряет смысл).

При таком предположении опорный цикл Q_0 , найденный согласно (3.4), правомерно считать «типичным», что позволяет обнаружить «нетипичные» циклы по расстоянию до Q_0 .

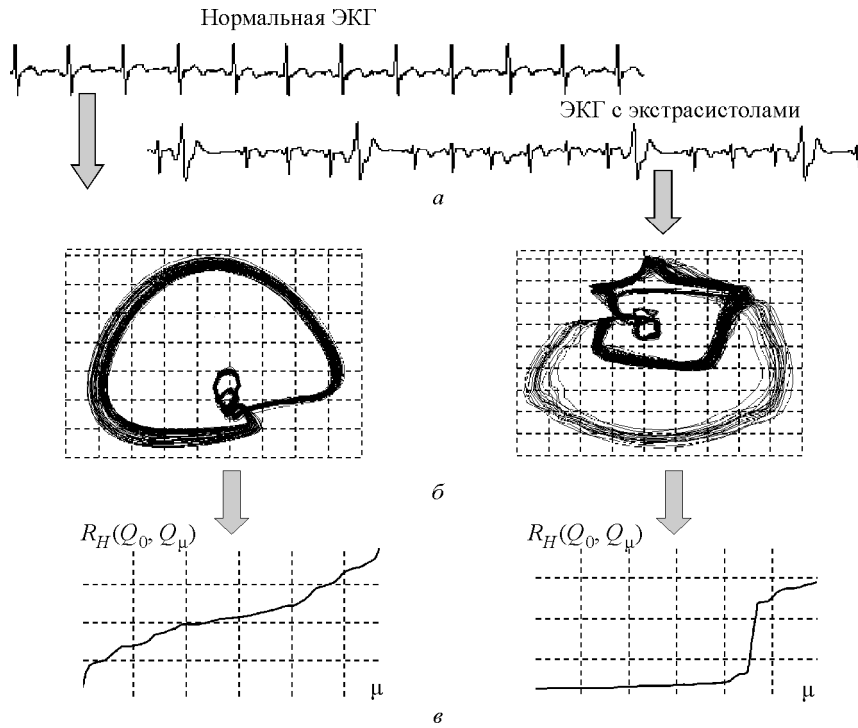


Рис. 3.8. Иллюстрация метода селекции нетипичных циклов:

a — исходные ЭКГ; *б* — их фазовые траектории; *в* — соответствующие кривые упорядоченных расстояний $\mathfrak{R}$

Упорядочим по возрастанию элементы строки матрицы $D = \|R_H(Q_i, Q_j)\|$, соответствующей опорному циклу Q_0 . Полученная последовательность значений $\mathfrak{R} = R_H(Q_0, Q_\mu)$, $\mu \in [1, M]$, дает представление о вариабельности формы траекторий наблюдаемых сигналов по отношению к Q_0 (рис. 3.8).

Упорядоченные значения $R_H(Q_0, Q_\mu)$ возрастают более-менее равномерно с ростом μ , когда наблюдаемый сигнал порожден одним эталоном. Если же на сигнале появляются нетипичные циклы, форма которых существенно отличается от эталона (артефакты, экстрасистолы ЭКГ и т. п.), то последовательность $\mathfrak{R}$ содержит один или несколько выраженных скачков.

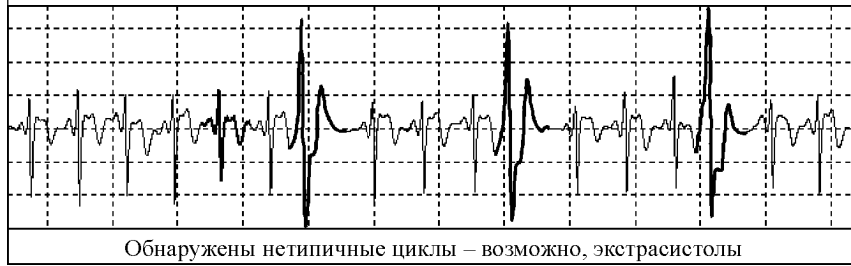


Рис. 3.9. Результаты обнаружения нетипичных циклов реальной ЭКГ

Определив положение первого из таких скачков, можно найти пороговое значение $\mathfrak{R}_0$ и тем самым выделить подмножество траекторий, подлежащих усреднению: $\Omega_0 = \{Q_\mu : R_H(Q_0, Q_\mu) < \mathfrak{R}_0\}$.

Эксперименты, проведенные на реальных ЭКГ с артефактами и экстрасистолами (более 1000 наблюдений), подтвердили эффективность предложенного подхода: процедура безошибочно обнаруживала все нетипичные циклы.

Рис. 3.9 иллюстрирует результат одного из этих экспериментов.

3.3.3. Модельный пример

Для того чтобы продемонстрировать эффективность предложенного подхода к оценке эталона, рассмотрим вначале модельный пример.

Пусть эталон $y_0(t)$ имеет форму равнобедренного треугольника (рис. 3.10, а), заданного двумя фрагментами в виде линейных функций

$$y_0(t) = \begin{cases} t, & 0 \leq t < 3, \\ 6 - t, & 3 \leq t < 6. \end{cases} \quad (3.15)$$

Предположим, что мы наблюдаем два цикла сигнала, порожденного в соответствии с моделью (2.9) по эталону (3.15), причем на 1-м цикле параметры растяжения по времени приняли значения $b_1^{(1)} = 2/3$ и $b_1^{(2)} = 2$, а на 2-м цикле — $b_2^{(1)} = 2$ и $b_2^{(2)} = 2/3$.

В результате наблюдаемый сигнал будет описывать функция

$$y(t) = \begin{cases} 3/2t, & 0 \leq t < 2, \\ 4 - t/2, & 2 \leq t < 8, \\ t/2 - 4, & 8 \leq t < 14, \\ 24 - 3/2t, & 14 \leq t < 16, \end{cases} \quad (3.16)$$

график которой показан на рис. 3.10, б.

Поскольку в данном случае наблюдаемые циклы имеют одинаковую продолжительность, их можно совместить во времени на интервале $[0, 8]$ (рис. 3.10, в) и усреднить во временной об-

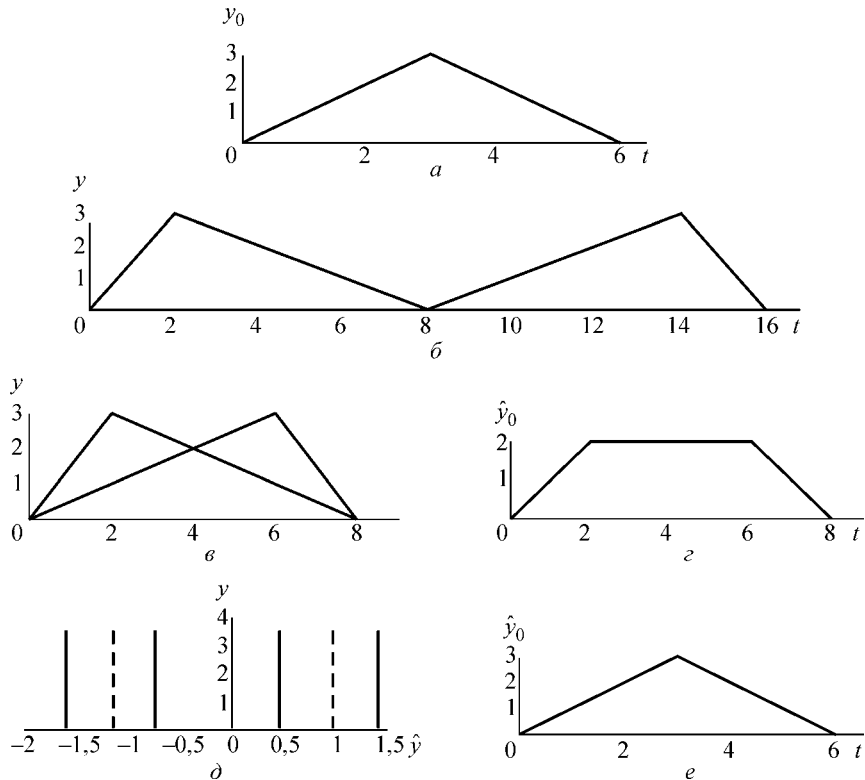


Рис. 3.10. Иллюстрация к модельному примеру:

a — эталон; $б$ — наблюдаемый сигнал; $в$ — совмещенные во времени циклы; $г$ — оценка эталона при усреднении во временной области; $д$ — фазовая траектория; $е$ — оценка эталона при усреднении в фазовом пространстве

ласти. Однако при этом будет получена оценка (рис. 3.10, *г*)

$$\hat{y}_0(t) = \begin{cases} t, & 0 \leq t < 2, \\ 2, & 2 \leq t < 6, \\ 8 - t, & 8 \leq t < 8, \end{cases}$$

которая по форме не соответствует эталону.

В то же время усреднение этих же циклов в фазовом пространстве координат (рис. 3.10, *д*) с последующим переходом во временную область (рис. 3.10, *е*) позволяет точно восстановить эталон (3.10, *а*).

3.4. Результаты численного моделирования

Для количественной оценки точности восстановления полезного сигнала по искаженным реализациям проводились модельные эксперименты. Их схема состояла в следующем (рис. 3.11).

По эталону $y_0[k]$ с известным набором локально сосредоточенных признаков $x_1, \dots, x_N$ генерировалась последовательность искаженных реализаций $y_1[k], \dots, y_M[k]$. Генерация осуществлялась согласно модели (2.9) с различными значениями порогов Δ_0, ε_0 , ограничивающих уровень внутреннего искажения $\xi(\cdot)$.

По каждой сгенерированной последовательности $y_1[k], \dots, y_M[k]$ вычислялась оценка $\hat{y}_0[k]$ эталона на основе усреднения реализаций во временной области и на основе усреднения реализаций

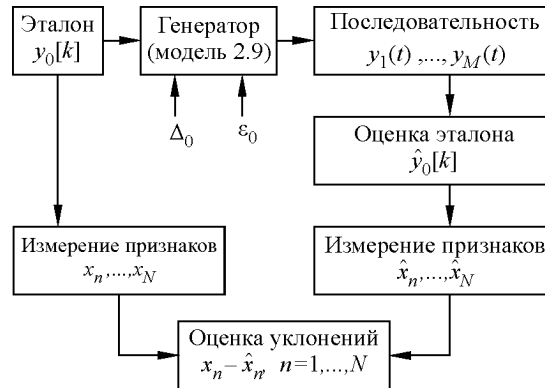


Рис. 3.11. Общая схема организации экспериментов

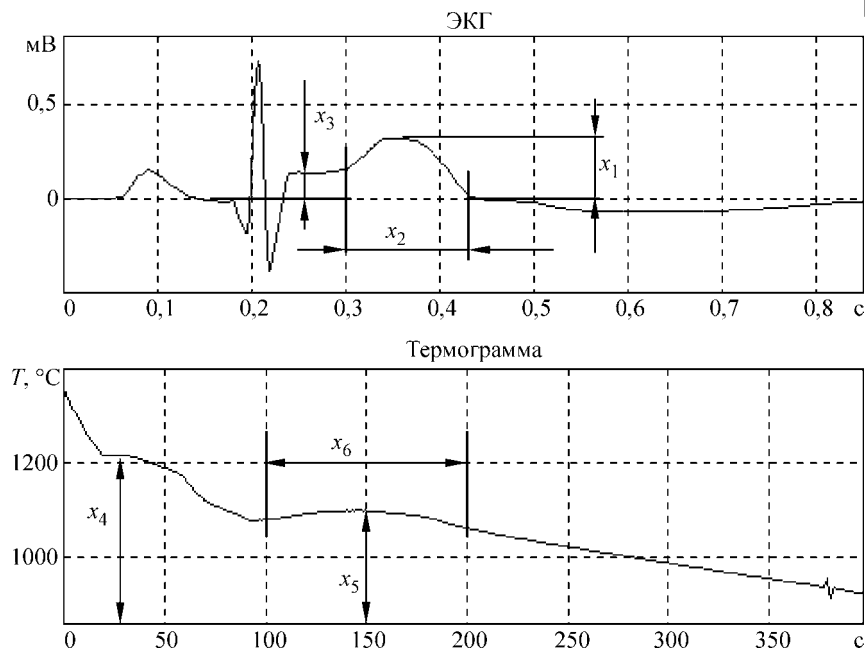


Рис. 3.12. Локально сосредоточенные признаки ЭКГ (а) и термограммы (б)

в фазовом пространстве. Усреднение проводилось по 20 искаженным реализациям. Эффективность процедур оценивалась точностью оценки значений признаков $x_1, \dots, x_N$, сосредоточенных на информативных фрагментах эталона.

В качестве эталонов использовались два типа сигналов — цикл ЭКГ и термограмма кристаллизации пробы жидкого чугуна. Для первого типа эталонов изучались точность оценки основных электрокардиографических признаков ишемии миокарда (рис. 3.12, а):

- амплитуда x_1 и продолжительность x_2 зубца T ;
- смещение x_3 сегмента $S - T$ относительно изолинии, которое принято называть депрессией и элевацией сегмента.

Для второго типа эталонов изучалась точность оценки основных диагностических параметров [306, 307], сосредоточенных на фрагментах дендритной кристаллизации и кристаллизации эвтектики (рис. 3.12, б), а именно:

- уровень температуры ликвидуса x_4 ;

ТАБЛИЦА 3.1

Ошибки в измерении значений признаков

Признак	Эталон	Усреднение во временной области		Усреднение в фазовом пространстве	
		абс. величина	ошибка	абс. величина	ошибка
x_1	0,324 МВ	0,284 МВ	12,3 %	0,321 МВ	0,9 %
x_2	0,134 с	0,158 сек	17,9 %	0,134 с	0 %
x_3	0,141 МВ	0,161 МВ	14,2 %	0,148 МВ	5 %
x_4	1216 °C	1213 °C	0,2 %	1215 °C	0,08 %
x_5	1099 °C	1088 °C	1 %	1093 °C	0,7 %
x_6	100 с	73 с	27 %	100,7 с	4 %

- уровень x_3 и продолжительность x_6 «площадки» солидуса.

Эффективность процедур оценивалась относительной погрешностью $\delta_n = |\hat{x}_n - x_n| / x_n$, $n = 1, \dots, N$, определения значений признаков по отношению к эталону. Численные результаты одного из экспериментов приведены в табл. 3.1.

Из табл. 3.1 видно, что использование предложенного подхода к восстановлению полезного сигнала позволяет существенно уменьшить ошибки в определении значений локально сосредоточенных признаков. Результаты численного моделирования показывают, что средняя погрешность в оценке значений диагнос-

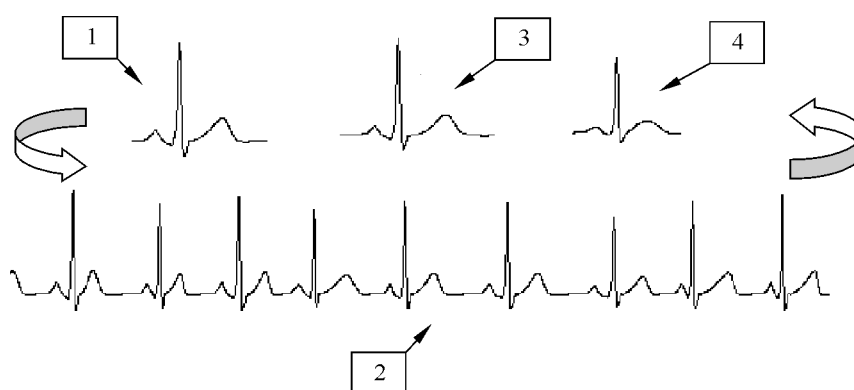


Рис. 3.13. Иллюстрация результатов восстановления цикла ЭКГ:

эталон (1); наблюдаемый сигнал (2); оценка эталона, полученная при усреднении сигнала в фазовом пространстве (3) и во временной области (4)

тических признаков, сосредоточенных на фрагментах ЭКГ и термограмм, была в пределах 1,5—2,5 % при оценке эталонов в фазовом пространстве, в то время как при использовании алгоритма усреднения во временной области пределы средних ошибок составляли 15—20 %.

И это не случайно, поскольку практически во всех экспериментах при усреднении во временной области наблюдались заметные искажения («размывание») формы информативных фрагментов (рис. 3.13).

Даже при визуальной оценке результатов видно, что предлагаемый метод позволяет более точно оценить форму информативных фрагментов эталона (зубца P , комплекса QRS , сегмента $S-T$ и зубца T), чем алгоритм усреднения во временной области.

4.1. Анализ методов подавления аддитивных помех

При обосновании алгоритма усреднения реализаций в фазовом пространстве мы опирались на базовую модель (2.9) порождения сигнала, предполагая, что внешнее возмущение (аддитивная помеха) отсутствует, т. е. $h(t) \equiv 0$. Поэтому рассмотрим общий случай, когда $h(t) \neq 0$, который требует применения эффективных процедур подавления внешних возмущений различных типов.

4.1.1. Общие замечания

При решении прикладных задач часто приходится сталкиваться с ситуацией, когда объект исследования порождает сигнал на низких уровнях амплитуды. Для регистрации таких сигналов используются высокочувствительные датчики, которые, помимо полезного сигнала, фиксируют электрические и магнитные сигналы от посторонних источников. Такие помехи, неизбежно возникающие в реальных условиях, чаще всего рассматриваются как аддитивный шумовой компонент, искажающий полезный сигнал.

Поэтому задача предварительной обработки в первую очередь связана с выделением полезного сигнала $y_0(t)$ по реализации $y(t) = y_0(t) + h(t)$, искаженной аддитивным возмущением $h(t)$. Заметим, что в отличие от

мультипликативного возмущения $\zeta(t)$, которое фигурирует в обобщенной модели (2.4), мешающее действие аддитивной помехи $h(t)$ не зависит от уровня сигнала.

Известны различные подходы к восстановлению полезного сигнала $y_0(t)$ на фоне аддитивной помехи $h(t)$.

Если при регистрации сигнала $y(t)$ можно синхронно измерять и помеху $h(t)$, то задача восстановления $y_0(t)$ существенно упрощается: достаточно использовать компенсационную схему, на один вход которой подавать искаженный сигнал, а на другой — сигнал помехи. Однако на практике чаще всего приходится иметь дело с неконтролируемой помехой.

В таких ситуациях не остается ничего иного, как выделять полезный сигнал на фоне помехи, используя различные алгоритмы фильтрации и сглаживания. Очевидно, что отделить полезный сигнал от помехи возможно лишь в том случае, когда есть хоть какие-то различия в характеристиках $y_0(t)$ и $h(t)$, например спектральных или статистических.

Прежде чем показать особенности решения задачи разделения $y_0(t)$ и $h(t)$ для сигналов с локально сосредоточенными признаками, проведем краткий анализ известных подходов к задаче фильтрации.

4.1.2. О критериях оптимальности процедур фильтрации

Цель фильтрации — наилучшее восстановление полезного сигнала на фоне помехи. Выбор того или иного критерия оптимальности, а значит, и метода фильтрации существенно зависит от круга задач, которые предстоит решать.

Согласно общей теории обнаружения сигналов [388] оптимальным фильтром для обнаружения сигнала известной формы на фоне белого гауссовского шума является параллельный набор согласованных фильтров. Например, при противофазной передаче двух сигналов $y_{01}(t)$ и $y_{02}(t) = -y_{01}(t)$ достаточно иметь только один согласованный фильтр.

Если же шумы гауссовы, но не «белые», то оптимальный фильтр становится значительно более сложным и для точного определения его структуры требуется решение соответствующего интегрального уравнения [388].

В случае бесконечного времени наблюдения оптимальный приемник представляет собой каскадное соединение предварительно «отбеливающего» и согласованного фильтров, причем частотная передаточная функция «отбеливающего» фильтра инверсна функции спектральной плотности шумов, а согласованный фильтр должен быть предварительно настроен на принимаемый полезный сигнал [132].

В тех случаях, когда шум не является гауссовым, ситуация усложняется: оптимальный (по критерию максимума правдоподобия) фильтр становится нелинейным и его структуру сложно реализовать [334].

Таким образом, даже если форма полезного сигнала $y_0(t)$ известна, построение оптимального фильтра для выделения $y_0(t)$ на фоне аддитивной помехи $h(t)$ представляет собой далеко не простую задачу, которая имеет формальное решение лишь при определенных предпосылках о модели $h(t)$, но они не всегда проверяемы на практике.

Если же функция $y_0(t)$ не известна, но может быть описана некоторой параметрической моделью $\hat{y}_0(t, \theta)$, то задачу фильтрации часто сводят к оценке вектора θ параметров по наблюдаемой реализации. Для дискретного сигнала $y[k]$, $k = 1, 2, 3, \dots$, получение такой оценки возможно на основе рекурсивных процедур, базирующихся на теории оптимальной фильтрации Калмана [17]. При определенных предпосылках о сигнале и помехе эти процедуры являются оптимальными в смысле минимума среднеквадратичного отклонения $\hat{y}_0(t)$ от $y_0(t)$.

Однако такие процедуры вовсе не гарантируют сохранение формы полезного сигнала на локальных фрагментах области определения и поэтому непригодны для обработки сигналов с локально сосредоточенными признаками.

Совершенно очевидно, что в основу построения эффективных вычислительных процедур фильтрации сигналов с локально сосредоточенными признаками должен быть положен иной критерий оптимальности, нежели среднеквадратическое отклонение сигнала на всей области определения. Например, минимум среднеквадратического отклонения сигнала на информативных фрагментах или, еще лучше, минимум среднеквадратического отклонения значений диагностических признаков, сосредоточенных на таких фрагментах.

Однако универсальной теории, позволяющей формализовать синтез фильтров, основанных на таких критериях, не существует. Это можно объяснить тем, что реальные сигналы с локально сосредоточенными признаками достаточно сложны для их описания моделью, заданной с точностью до конечного числа параметров.

Поэтому не остается ничего иного, как ввести некоторые предположения о модели помехи и искать приемлемые процедуры, обеспечивающие максимально возможное подавление аддитивной помехи при минимальных искажениях значений диагностических признаков, сосредоточенных на локальных фрагментах сигнала.

Естественно, что подтверждением эффективности (прагматической «оптимальности») таких процедур будут уже не формальные приемы, а результаты численного моделирования на тестовых и реальных данных, что вполне допустимо при построении прикладных ИТ.

4.1.3. Характерные типы внешних возмущений

Понятно, что при регистрации сигналов с локально сосредоточенными признаками могут наблюдаться различные типы внешних возмущений. Однако анализ практических задач позволяет выделить наиболее характерные типы помех, с которыми чаще всего приходится сталкиваться при разработке прикладных ИТ (рис. 4.1).

По соотношению ширины спектров помехи и полезного сигнала различают узкополосные и широкополосные частотные помехи. Естественно, что одна и та же помеха по отношению к одному сигналу может быть узкополосной, а по отношению к другому — широкополосной.

Особый случай составляют так называемые сосредоточенные частотные помехи, которые могут наблюдаться в очень узкой по-



Рис. 4.1. Характерные типы помех

лосе частот, например, сетевые помехи на частоте 50 Гц и кратных частотах. Значительно сложнее подавить помехи на основе узкополосных заградительных (режекторных) фильтров, если гармоническая помеха сосредоточена в диапазоне «полезных» частот.

К нестационарной помехе можно отнести низкочастотный дрейф (смещение) базового уровня сигнала, например смещение изолинии ЭКГ.

При решении прикладных задач случайные помехи часто рассматривают как дискретно заданный белый шум — последовательность независимых одинаково распределенных величин. Обычно предполагается, что математическое ожидание такой помехи равно нулю, а дисперсия ограничена некоторой величиной.

Импульсной называется помеха малой длительности, которая в общем случае представляет собой одиночные выбросы, случайно распределенные по времени и амплитуде. К подобному типу помех можно отнести и резкие изменения базового уровня сигнала («скачки»), происходящие в случайные моменты времени.

В общем случае можно предположить, что перечисленные типы помех присутствуют одновременно и независимо искажают полезный сигнал с локально сосредоточенными признаками. Поэтому будем рассматривать модель помехи как сумму частных моделей

$$h(t) = h_1(t) + h_2(t) + \dots + h_F(t) \quad (4.1)$$

и строить процедуру предварительной обработки сигнала в виде последовательности F фильтров, каждый из которых ориентирован на подавление помех определенного типа.

Поскольку подтверждением эффективности процедур, реализующих такие фильтры, могут быть результаты численного моделирования, прежде чем исследовать возможные подходы к реализации фильтров, рассмотрим основные принципы, которые должны быть положены в основу таких экспериментов.

4.2. Принципы построения инструментальной системы

4.2.1. Концептуальная идея

Понятно, что невозможно построить универсальную инструментальную систему, пригодную для синтеза любой прикладной ИТ, поскольку каждый объект, а значит, и порождаемый им сигнал, вообще говоря, уникален. Поэтому ограничимся построением

инструментальных средств, которые главным образом ориентированы на предварительную обработку сигналов с локально сосредоточенными признаками, предполагая, что механизм их порождения описывается моделью (2.9).

Для экспериментальной оценки эффективных вычислительных процедур обработки сигналов необходима интеграция опыта, знаний и интуиции конструктора ИТ и эксперта предметной области.

Однако профессиональные языки этих специалистов разные. Конечного пользователя (эксперта) интересуют лишь потребительские свойства разрабатываемой ИТ, в то время как конструктора интересует то, как в условиях ограниченной априорной информации получить нужную для пользователя информацию на основе обработки доступных данных.

Возникает необходимость разработки специализированной инструментальной системы, которая позволяла бы конструктору прикладной ИТ совместно с экспертом проводить экспериментальные исследования и в конечном итоге реализовать технологическую цепочку процедур из готовых и доступных ему компонент самой инструментальной среды [179].

Сложность разработки такой системы состоит в том, что она должна объединять формальные методы построения базовых вычислительных компонент с возможностью формальных и неформальных приемов их выбора и настройки для конкретной прикладной задачи [214].

Заметим, что в соответствии с расширенным толкованием известной теоремы Курта Геделя [35, 130] невозможно в рамках компьютерных наук построить дедуктивную систему доказательства полноты базового набора вычислительных компонент для реализации прикладных ИТ. Поэтому нужно обеспечить открытость инструментальной системы и, если потребуется, дополнить имеющийся набор необходимыми компонентами (рис. 4.2).

Таким образом, основная концептуальная идея, которую предлагается положить в основу построения инструментальной системы [179, 214, 280], состоит в том, что с ее помощью конструктор получает возможность

- *проводить* необходимые эксперименты для выбора и оптимальной настройки вычислительных процедур, реализующих отдельные элементарные стадии обработки, опираясь на результаты численного моделирования и средства образного анализа данных [72] на основе методов когнитивной компьютерной графики [89, 90, 163];

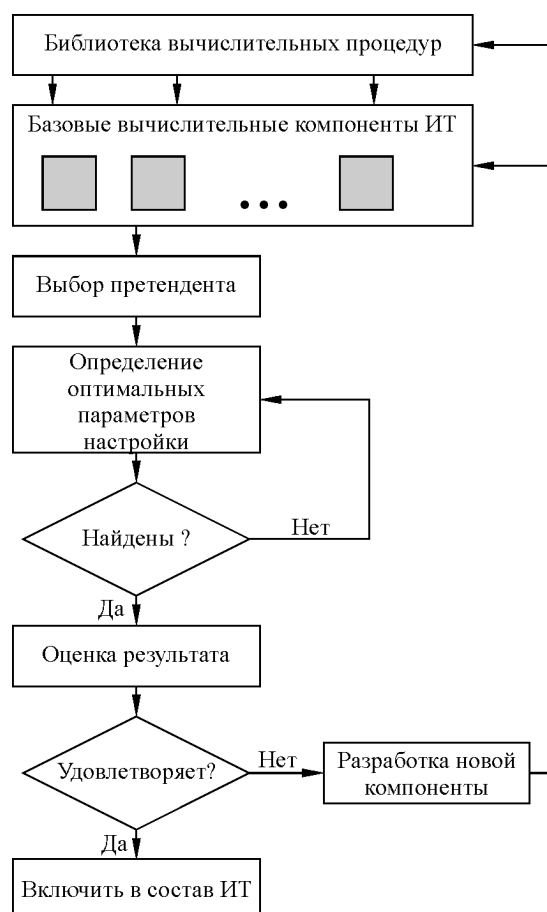


Рис. 4.2. Концептуальная идея интерактивного синтеза ИТ

- *реализовывать* технологическую цепочку алгоритмов обработки сигнала, по возможности используя готовые вычислительные компоненты самой инструментальной среды;

- *расширять* при необходимости состав компонент инструментальной системы (базовых вычислительных процедур и средств отображения результатов), если имеющийся набор этих средств недостаточен для решения конкретной прикладной задачи.

Для практической реализации этой идеи в рамках объектно-ориентированного подхода необходимо, прежде всего, провести

исследования, направленные на конструирование главных программных классов, на основе которых может быть реализовано ядро инструментальной системы.

4.2.2. Архитектура инструментальной системы

В основе ИТ обработки сигналов лежит технологическая цепочка, состоящая из конечного числа N_Q вычислительных процедур (обработок), посредством которых вектор $Y = (y[1], \dots, y[K])$ дискретных значений наблюдаемого сигнала последовательно преобразуется в выходной продукт — оценку $\hat{s}$ текущего состояния объекта:

$$\begin{aligned} D_R^{(1)} = A_1(\alpha_1, Y) \rightarrow \dots \rightarrow D_R^{(i)} = A_i(\alpha_i, D_S^{(i)}) \rightarrow \dots \rightarrow D_R^{(N_Q)} = \\ = A_{N_Q}(\alpha_{N_Q}, D_S^{(N_Q)}) = \hat{s}, \end{aligned} \quad (4.2)$$

где $A_i(\cdot)$ — алгоритм преобразования исходных данных $D_S^{(i)}$ в результат $D_R^{(i)}$ обработки на i -й стадии; α_i — параметры настройки алгоритма, причем под исходными данными i -й стадии в общем случае понимается не только результат $D_R^{(i-1)}$, но и любые предшествующие ему результаты.

Если структура (4.2) определена и конструктор ИТ имеет полную информацию о каждой i -й стадии обработки $D_R^{(i)} = A_i(D_S^{(i)})$, то синтез ИТ сводится к традиционным этапам жизненного цикла разработки программного обеспечения. Однако опыт показывает, что при синтезе новых ИТ обработки сигналов, порождаемых сложными объектами, основное время затрачивается не столько на программную реализацию вычислительных алгоритмов, сколько на экспериментальные исследования, направленные на поиск подходящих алгоритмов и их оптимальную настройку.

Рассмотрим на примере решения задачи предварительной обработки сигналов возможные подходы к построению архитектуры инструментальной системы, позволяющей ускорить этот процесс.

В соответствии с методологией объектно-ориентированного подхода (ООП) данные и методы их обработки следует инкапсулировать в единый объект, описывающий некоторую сущность предметного мира. Разумеется, прикладная система, являясь

сущностью предметного мира, может сама рассматриваться как объект. Однако такой объект будет слишком велик и требует декомпозиции, т. е. разложения на объекты более низкого уровня.

В соответствии с (4.1) для подавления внешних искажений необходимо реализовать цепочку элементарных стадий обработки, направленных на подавление аддитивных помех определенного типа. Поэтому представляется естественным проводить декомпозицию системы с точки зрения этих стадий, представив подсистему предварительной обработки в виде последовательности взаимодействующих объектов элементарной обработки (ЭО).

В данном случае предполагается, что каждый объект ЭО:

- реализует отдельный алгоритм фильтрации;
- при необходимости отображает результат в графическом виде на экране монитора;
- исполняет роль приемника и/или источника информации при реализации каскадного алгоритма фильтрации.

Для осуществления такого подхода в рамках ООП необходимо разработать абстрактный класс, реализующий в самых общих чертах обобщенную модель элементарной обработки (ЭО). В этом классе входные и выходные данные объединяются в единый блок данных, а методы обработки и визуализации — в единый блок методов. Далее следует построить иерархию наследников класса, в которых последовательно конкретизируются представления о данных, методах их обработки и визуализации.

Таким образом, на базе производных классов могут быть разработаны конкретные экземпляры объектов ЭО, которые реализуют отдельные стадии фильтрации и визуализацию полученных результатов.

Отсюда следует, что абстрактный класс ЭО должен содержать как минимум два поля (свойства) — «Исходные данные (ИД)» и «Результат обработки (РО)», а также виртуальные методы работы с этими полями — «Обработать ИД», «Отобразить РО» и «Передать РО».

Понятно, что для организации взаимодействия различных объектов ЭО необходим внешний координатор, который также желательно реализовать в виде отдельного объекта — менеджера системы (МС). Такой менеджер должен обеспечить:

- порождение требуемого экземпляра ЭО и регистрацию его в своем поле;
- передачу ЭО исходных данных;
- вызов метода ЭО «Выполнить обработку»;

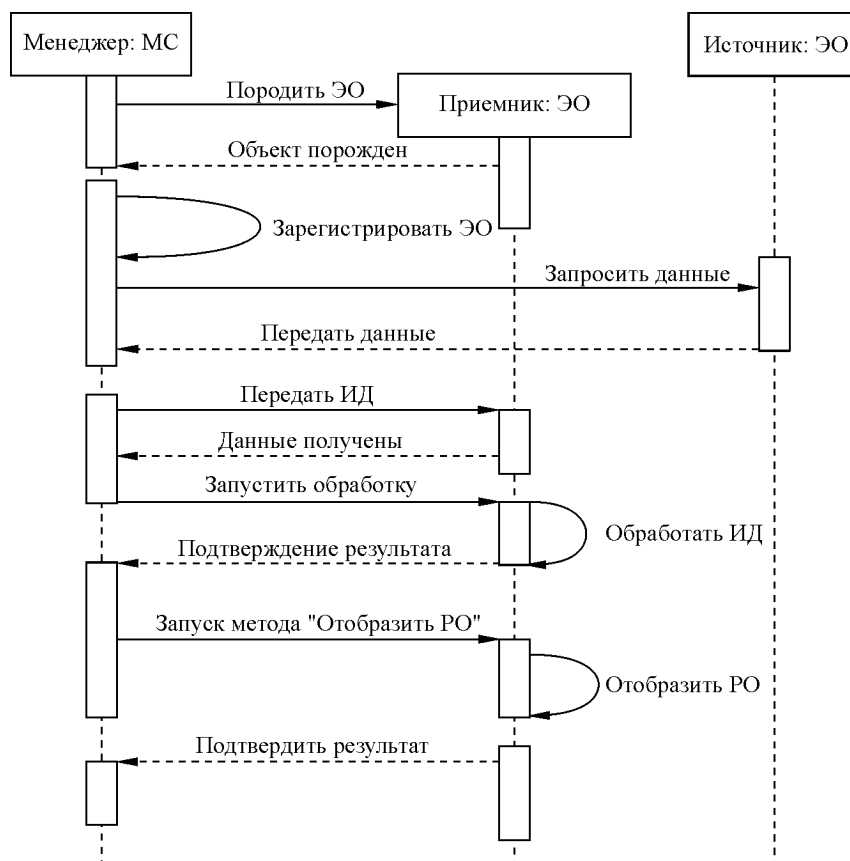


Рис. 4.3. Диаграмма взаимодействия менеджера системы с объектами ЭО (в нотациях универсального языка моделирования UML)

- передачу ЭО команды на отображение результата (при необходимости).

На рис. 4.3 показан фрагмент укрупненной схемы взаимодействия (диаграммы последовательности в терминах унифицированного языка моделирования UML) объекта МС с двумя объектами ЭО, один из которых служит источником информации, а другой — приемником.

Рассмотренная модель построения инструментальной системы представляется достаточно простой и вполне логичной с точки зрения объектно-ориентированного подхода: достаточно

ТАБЛИЦА 4.1

Структура производных классов ЭО

Поле, метод	Производные классы			
	класс ЭО 1	класс ЭО 2	класс ЭО 3	класс ЭО 4
Поле ИД	—	Исходная ЭКГ	Сглаженная ЭКГ	Сглаженная ЭКГ
Метод обработки	Чтение файла данных	Сглаживание	Сегментация	Измерение признака
Поле РО	Исходная ЭКГ	Сглаженная ЭКГ	Границы фрагментов	Значение признака
Метод визуализации	Отображение исходной ЭКГ	Отображение сглаженной ЭКГ	Отображение границ фрагментов	Отображение значения признака

разработать нужные классы ЭО, а взаимодействие объектов этих классов в системе реализовать с помощью объекта МС.

Однако при более глубоком анализе этой модели оказывается, что она не обладает требуемой гибкостью. Для того чтобы это продемонстрировать, рассмотрим конкретный пример.

Пусть требуется обработать сигнал с локально сосредоточенными признаками (например, цикл ЭКГ), представленный последовательностью дискретных значений $y[1], \dots, y[K]$, которые хранятся в файле данных на диске. Предположим для простоты, что задача обработки состоит в сглаживании сигнала, разбиении его на отдельные фрагменты и оценке значения признака, сосредоточенного на одном из информативных фрагментов. Результаты обработки должны быть отображены на экране монитора.

Для построения такой системы на основе рассмотренного подхода необходимо разработать четыре наследника класса ЭО:

- ЭО 1: «Чтение и отображение исходного сигнала»;
- ЭО 2: «Сглаживание и отображение сглаженного сигнала»;
- ЭО 3: «Выделение фрагментов и отображение их границ»;
- ЭО 4: «Оценка и отображение на экране значения признака».

Поскольку эти классы являются наследниками одного абстрактного класса ЭО, их реализация требует переопределения методов абстрактного класса согласно табл. 4.1.

Однако в этом случае каждый из производных объектов, помимо реализации методов обработки, выполняет также функцию контейнера исходных данных и результата обработки, что, как

видно из табл. 4.1, приводит к ненужному дублированию информации.

Рассмотренный пример представляет собой достаточно простую схему последовательной обработки. Поэтому в данном случае ненужное дублирование информации в полях наследников ЭО не так уж страшно: с помощью менеджера можно уничтожить объект-источник после того как необходимая информация будет передана объекту-приемнику.

Однако в общем случае для выполнения i -й стадии обработки может потребоваться не только результат $i - 1$ -й стадии, но и какой-либо другой из предшествующих стадий. В таких ситуациях мы вынуждены оставлять активным объект ЭО только лишь с целью сохранения результата обработки. Вполне понятно, что такую организацию системы нельзя признать удовлетворительной.

Кроме того, разнообразие алгоритмов, которые могут потребоваться для реализации отдельных стадий ИТ, существенно больше, чем разнообразие методов визуализации результатов обработки.

Принимая во внимание эти факты, предлагается несколько иначе взглянуть на декомпозицию системы: вычленив данные и методы их визуализации из класса ЭО и рассмотреть их как независимый класс — модель носителя данных (МНД), но оставить за ЭО возможность ссылаться на нужный экземпляр объекта МНД (рис. 4.4).

При объединении в одном объекте данных и методов их визуализации открываются возможности:

- получать доступ к данным посредством экранных графических образов, отображающих эти данные;
- обеспечить необходимую привязку графического образа, визуализирующего результат обработки, к графическому образу исходных данных.

Такие возможности весьма полезны, а порой и просто необходимы при синтезе прикладных ИТ обработки сигналов.

Для решения этой задачи в рамках ООП предлагается разработать абстрактный класс МНД, реализующий в самых общих чертах обобщенную модель носителя данных, а далее построить иерархию наследников этого класса, в которых последовательно конкретизировать представления о данных и методах их визуализации.

При таком подходе каждый отдельный экземпляр наследника МНД может исполнять роль либо источника, либо приемника

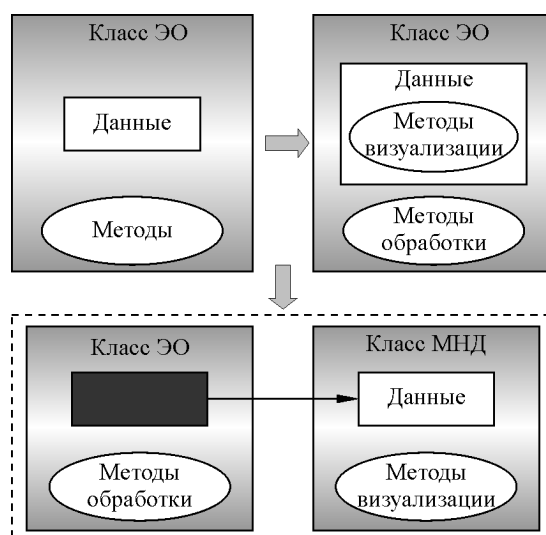


Рис. 4.4. Декомпозиция класса элементарной обработки данных (ЭО)

данных, а совокупность порожденных объектов МНД образует виртуальную базу данных, управление которой разумно поручить менеджеру системы (МС).

Поскольку экземпляры объектов МНД должны тесно взаимодействовать с объектами ЭО, целесообразно сам метод порождения объектов МНД «делегировать» от менеджера непосредственно к объектам ЭО, сохранив за менеджером лишь функции регистрации. Такое распределение координационных функций может быть реализовано передачей соответствующих ссылок (указателей) на нужные типы объектов.

В дополнение к этому предлагается также сосредоточить процедуры, реализующие вычислительные алгоритмы обработки данных на разных стадиях, в библиотеке вычислительных процедур (БВП), которая может разрабатываться независимо от структуры проектируемой системы, оставив в объекте ЭО лишь возможность ссылаться на конкретную вычислительную процедуру.

В результате мы приходим к трехуровневой схеме организации управления объектами системы (рис. 4.5). Менеджер системы (МС) порождает объекты ЭО и взаимодействует с ними. Объекты ЭО в свою очередь порождают объекты МНД. При этом первые вызывают требуемую вычислительную процедуру из библиотеки БВП, передают результат ее выполнения в объекты МНД,

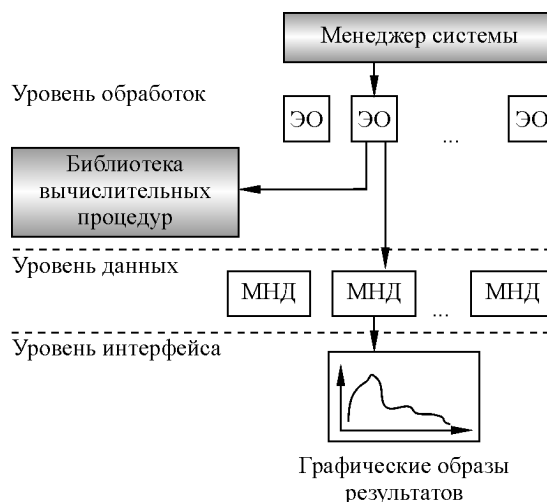


Рис. 4.5. Трехуровневая схема построения инструментальной системы

которые обеспечивают хранение и, при необходимости, отображение результата обработки на экране монитора.

Укрупненная модель взаимодействия менеджера системы (МС) с объектами элементарной обработки (ЭО) и носителя данных (НД) показана на рис. 4.6.

В отличие от ранее рассмотренной схемы (см. рис. 4.3), теперь уже линия жизни объекта ЭО всегда завершается после выполнения очередной стадии обработки. Активным же остается объект НД-приемник и, возможно, объект НД-источник, если содержащиеся в нем данные потребуются на последующих стадиях обработки.

Для практической реализации такой схемы должны быть детализированы поля и методы базовых программных классов — обобщенной модели обработки и обобщенной модели носителя данных, составляющих основу ядра инструментальной системы. Поэтому рассмотрим эти вопросы более подробно.

4.2.3. Объектно-ориентированные модели ядра инструментальной системы

В основу построения обобщенной модели носителя данных (МНД) положена идея разработки абстрактного класса, на базе которого создаются объекты инструментальной системы предва-

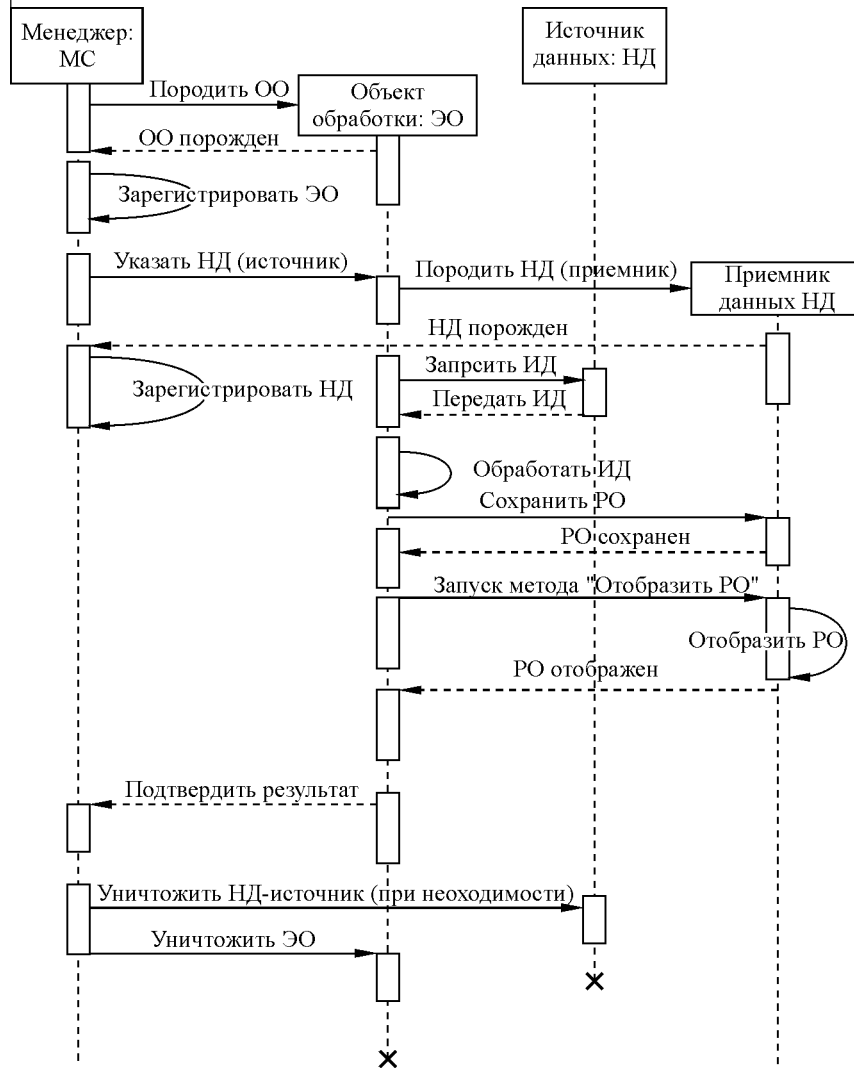


Рис. 4.6. Диаграмма взаимодействия менеджера системы (МС) с объектами элементарной обработки (ЭО) и носителя данных (НД)

рительной обработки сигналов. Эти объекты должны обеспечивать (рис. 4.7):

- хранение содержательных данных, необходимых для организации элементарных вычислительных процедур на различных

стадиях предварительной обработки сигнала;

- отображение результатов обработки отдельных стадий в виде экранных графических образов и текстовой информации;
- доступ к данным посредством их графических образов;
- модификацию графических образов (перемещение, масштабирование, изменение типа и т. п.).

С этой целью предлагается [213] рассматривать МНД в виде тройки:

$$\text{МНД} = \langle D_C, D_G, G_M \rangle, \quad (4.3)$$

где D_C — поля содержательных данных; D_G — поля графических данных; G_M — методы отображения и модификации графических образов.

Упрощенная структура абстрактного класса, реализующего эту модель, показана на рис. 4.8, а. Идентификационный номер ID позволяет регистрировать порождаемые объекты МНД. Содержательные данные D_C включают в себя указатели для хранения массивов заданного размера K , причем для хранения одномерного массива ($Dim = 1$) используется указатель MY , двумерного ($Dim = 2$) — указатели MX и MY , а трехмерного ($Dim = 3$) — указатели MXX , MYY и MZZ . В последнем случае поля MX и MY заполняются данными, необходимыми для построения проекции трехмерного графического образа, соответствующего данным полей MXX , MYY и MZZ .

Графические данные D_G составляют поля (свойства) класса МНД, которые определяют цвет порождаемого графического образа, признак его видимости на экране монитора, координаты привязки к экранным координатам, коэффициенты масштабирования, элементы матрицы проектирования трехмерного графического образа и т. п.

Широкий набор методов отображения и модификации графических образов (перемещения, масштабирования, проектиро-

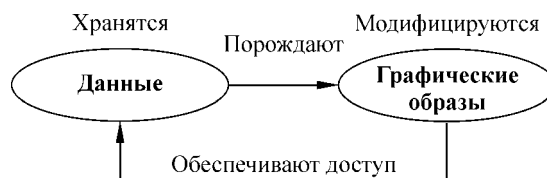


Рис. 4.7. Концептуальная идея построения МНД

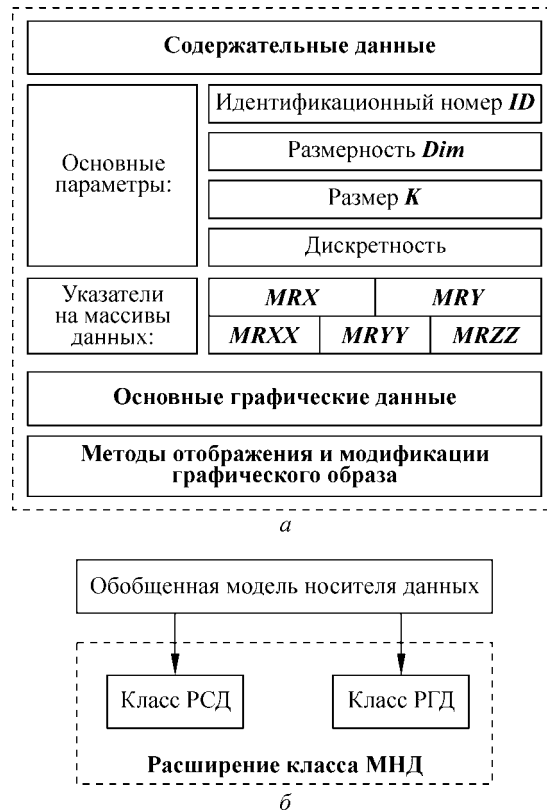


Рис. 4.8. Объектно-ориентированная модель носителя данных:

а — упрощенная структура абстрактного класса МНД; *б* — механизм динамического расширения класса

вания и др.), инкапсулированных в модель, дает возможность пользователю инструментальной системы обращаться к полям графических данных через элементы управления интерфейса инструментальной системы. Например, перемещать и масштабировать нужным образом графические образы на экране монитора и т. п. В этом случае инструментальная система исполняет роль «интеллектуального» осциллографа, а конструктор ИТ, получая требуемое отображение результатов обработки, может вместе с экспертом предметной области оценивать эффективность той или иной вычислительной процедуры не только численно, но и визуально.

Следует отметить, что в реальной системе абстрактная модель носителя данных (МНД) используется лишь в форме производных классов, полученных от нее путем наследования. Однако, несмотря на разнообразие возможных наследников, роль обобщенной модели в этой иерархии более чем значительна: именно в ней определена главная часть содержательных и графических данных, а также связанная с ними функциональность объектов. И лишь относительно небольшой объем дополнительных данных и необходимых методов доопределяется в наследниках класса МНД.

Для того чтобы ограничить иерархию наследников класса МНД и в то же время придать этому классу требуемую гибкость и возможность дальнейшего развития, в МНД предложено ввести универсальный механизм расширения (рис. 4.8, б). Этот механизм реализуется соответствующими ссылками на два дополнительных класса, обеспечивающих расширение содержательных (РСД) и графических данных (РГД).

Поле данных класса РСД представляет собой динамический вариантный массив, который поддерживается современными языками программирования, в частности Delphi. В этом случае объекты РСД обеспечивают хранение различных атрибутов очередного этапа обработки, таких как наименование (дескриптор) процедуры обработки, ее тип, параметры настройки и т. п. Для удобства доступа к атрибутам класс РСД содержит минимальный набор необходимых методов, в том числе «Добавить атрибут», «Извлечь атрибут», «Удалить атрибут» и т. п.

Объекты РГД обеспечивают хранение дополнительных графических данных в виде различных знаков (меток), которые привязываются к изображению порождаемого графического образа и тем самым повышают его наглядность и информативность.

Каждая графическая метка характеризуется типом и свойствами, определяющими ее вид на экране. Для управления механизмом отображения меток в базовой модели РГД предусмотрены методы, позволяющие добавить и удалить метку заданного типа. При необходимости варианты меток могут быть расширены в производных классах базовой модели РГД.

Поскольку трехуровневая архитектура построения инструментальной системы (рис. 4.5) предполагает, что данные $D_S^{(i)}$ и $D_R^{(i)}$ хранятся в объектах МНД, а процедура, реализующая элементарную стадию обработки $A_i(\cdot)$, сосредоточена в библиотеке

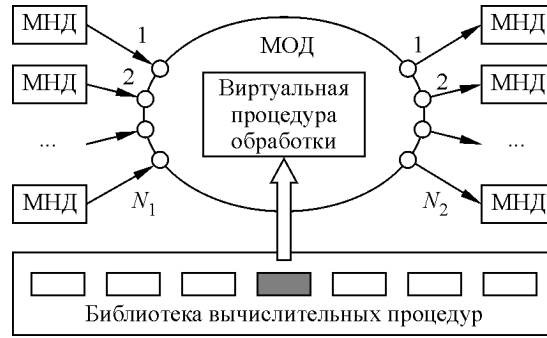


Рис. 4.9. Упрощенная схема компоновки элементарной стадии обработки: МОД — модель обработки данных; МНД — модель носителя данных

вычислительных процедур, необходим объект, обеспечивающий взаимодействие этих сущностей.

Идея построения обобщенной модели обработки данных (МОД) как раз и состоит в разработке абстрактного класса с универсальными полями, свойствами и методами, инвариантными относительно конкретной процедуры обработки. При этом МОД содержит виртуальный метод «пустой» обработки, который доопределяется в наследниках этого класса (рис. 4.9).

Абстрактный класс МОД позволяет унифицировать понятие «элементарная обработка данных», упорядочить процесс создания программных компонент, реализующих отдельные стадии обработки, и тем самым организовать удобный интерфейс взаимодействия между активными экземплярами (объектами) класса МНД и конкретной процедурой обработки.

Класс МОД поддерживает унифицированный интерфейс вызова любой вычислительной процедуры:

$$\langle \text{Имя} \rangle (I_1, \dots, I_{N_1}; L_1^{(I)}, \dots, L_{N_1}^{(I)}; O_1, \dots, O_{N_1}; \\ L_1^{(O)}, \dots, L_{N_2}^{(O)}; \alpha_1, \dots, \alpha_{N_{1\alpha}}; \text{Mess}), \quad (4.4)$$

где $I_1, \dots, I_{N_1}$; $O_1, \dots, O_{N_2}$ — указатели на массивы входных и выходных данных; $L_1^{(I)}, \dots, L_{N_1}^{(I)}$; $L_1^{(O)}, \dots, L_{N_2}^{(O)}$ — число элементов в соответствующих массивах; $\alpha_1, \dots, \alpha_{N_{1\alpha}}$ — настроечные параметры процедуры обработки; Mess — строка сообщения об исключительной (аварийной) ситуации.

Такой интерфейс позволяет распараллелить работу программистов, участвующих в создании инструментальной системы. Прикладные программисты могут сосредоточиться на их главной задаче — реализации эффективных вычислительных процедур обработки данных, не заботясь об организации диалога с пользователем и методах графического представления результатов, а системные программисты — рассматривать процедуру обработки как некий «черный ящик».

4.2.4. Базовые компоненты инструментальной системы

Предложенная обобщенная модель обработки данных позволила нам на базе производных классов МОД реализовать широкий набор вычислительных компонент, необходимых при построении прикладных ИТ обработки сигналов с локально сосредоточенными признаками. При этом оценка эффективности разработанных базовых компонент проводилась на модельных и реальных данных согласно обобщенной схеме экспериментов, показанной на рис. 4.10.

Для организации таких экспериментов в состав инструментальной системы включены компоненты, обеспечивающие решения типовых задач, с которыми чаще всего сталкивается конструктор прикладных ИТ, в том числе:

- ввода реальных сигналов;
- генерации тестовых сигналов;
- формирования смесей и цепочек фрагментов сигнала для имитации типовых искажений;
- восстановления формы полезного сигнала по искаженной реализации;
- выделения фрагментов сигналов (во временной области и в фазовом пространстве);

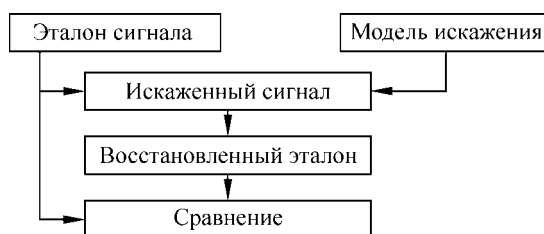


Рис. 4.10. Общая схема организации экспериментов

- измерения значений диагностических признаков, сосредоточенных на фрагментах.

Компоненты *ввода* реальных сигналов обеспечивают 16-канальный ввод однополярных аналоговых сигналов трех стандартных уровней: $0...2,5V$, $0...5,0V$ и $0...10,0V$, либо 8-канальный ввод биполярных сигналов уровней $-2,5V...+2,5V$, $-5,0V...+5,0V$, $-10,0V...+10,0V$.

Преобразование аналоговых сигналов в цифровую форму осуществляется с помощью быстродействующего 14-разрядного АЦП (время преобразования 2 мкс). В тех случаях, когда требуется вводить сигналы нестандартных уровней, в системе дополнительно используются нормирующие преобразователи.

Предусмотрены также средства ввода предварительно оцифрованных сигналов через стандартные интерфейсы COM и USB, а также импорта дискретных значений сигнала из файлов различных форматов.

Компоненты *генерации* тестовых сигналов обеспечивают формирование временных последовательностей, соответствующих различным моделям детерминированных и вероятностных процессов.

В системе реализованы генераторы типовых одиночных сигналов и их последовательностей, в том числе генератор полигармонического сигнала, последовательности независимых одинаково распределенных величин с различными законами распределения (равномерный, нормальный, логнормальный, Коши, Парето, Максвелла, Релея) и др.

Реализованы также генераторы специальных сигналов, в частности, имитатор циклических последовательностей, описанный в пункте 2.3.3.

Базовые вычислительные компоненты системы (более 100 программных модулей) включают в себя широкий набор стандартных методов цифровой обработки сигналов: нормализации, сплайн-интерполяции, аппроксимации, дифференцирования на основе регуляризирующих процедур и др.

Естественно, что в число базовых включены и вычислительные компоненты, реализующие оригинальные процедуры обработки сигналов в фазовом пространстве (см. гл. 3).

В системе обеспечена возможность «наложения» гармонических, случайных и импульсных помех на реальные и тестовые сигналы, а также имитации различных других искажений, с которыми приходится иметь дело при решении практических задач

обработки сигналов. Использование этих средств позволяет проводить эксперименты согласно схеме, показанной на рис. 4.10.

При разработке инструментальной системы естественным требованием было построение дружелюбного пользовательского интерфейса. Обеспечение этого требования и породило название системы «СИДИГРАФ» (Сигнал—Диалог—График), в которой режим диалога с пользователем реализует принцип «Укажи и выбери». Для обеспечения доступа к данным достаточно указать на соответствующий график в рабочем окне, а для вызова требуемой вычислительной процедуры — выбрать соответствующий пункт меню. Тем самым конструктору прикладной системы предоставлена возможность реализации цепочки вычислительных процедур из готовых компонент самой инструментальной системы (рис. 4.11, а).

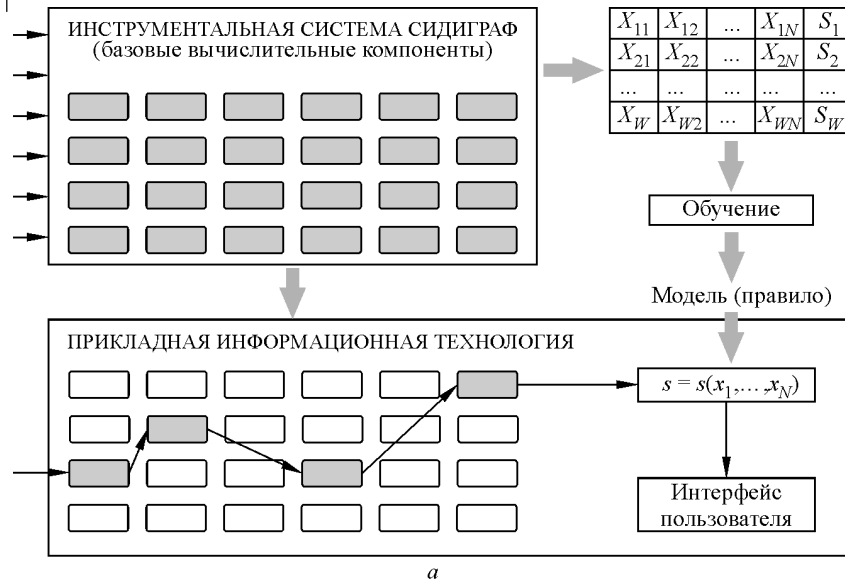
Полезной находкой системы, на наш взгляд, является подход, обеспечивающий интерактивную настройку параметров вычислительных процедур с использованием идей когнитивной компьютерной графики [89, 90]. В тех случаях, когда отсутствуют формальные методы определения оптимальных значений настроечных параметров, предложенный подход позволяет быстро определить эти параметры экспериментальным путем.

На рис 4.11, б показано рабочее окно системы при настройке вычислительного модуля сглаживания зашумленного сигнала.

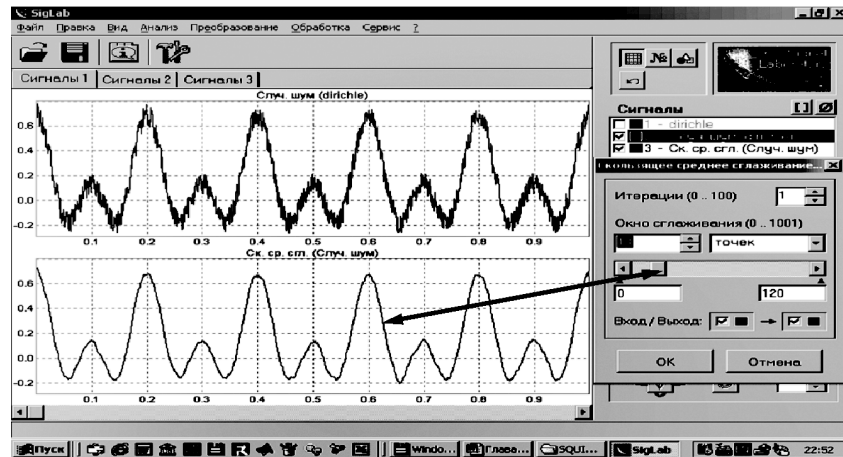
Благодаря «*виртуальной*» связи органа управления и графического образа, отображающего результат обработки (нижний график), пользователь может визуально определить «оптимальное» значение параметра настройки вычислительной процедуры (в данном случае — алгоритма сглаживания). При нажатии клавиши ОК найденное таким образом значение параметра настройки сохраняется в поле соответствующего объекта носителя данных. Аналогичным способом осуществляется выбор и настройка других вычислительных процедур, реализующих отдельные стадии ИТ обработки сигналов с локально сосредоточенными признаками.

При разработке базовых вычислительных компонент удалось внести ряд полезных усовершенствований в традиционные алгоритмы, в частности процедуры фильтрации частотных и случайных помех.

Поскольку при обработке сигналов с локально сосредоточенными признаками требуется не только эффективно подавить



a



б

Рис. 4.11. Иллюстрация метода интерактивного синтеза ИТ:

а — организация цепочки вычислительных процедур; б — настройка параметров вычислительных компоненты

мешающую помеху, но и в максимально возможной степени сохранить форму информативных фрагментов полезного сигнала, рассмотрим эти усовершенствования более подробно.

4.3. Эффективная процедура фильтрации сосредоточенных гармонических помех

4.3.1. Общие замечания, касающиеся частотно-избирательной фильтрации

Предположим, что аддитивная помеха имеет частотную природу и на интервале наблюдения может быть представлена в виде суммы стационарных гармонических колебаний:

$$h[k] = \sum_{g=1}^{M_G} a_g \sin(2\pi f_g k + \varphi_g), \quad (4.5)$$

где a_g , f_g , φ_g — соответственно частота, амплитуда и начальная фаза g -й гармоники.

В общем случае для подавления таких помех используют частотно-избирательные фильтры, которые пропускают сигналы в определенной полосе частот и задерживают остальные частоты. Принято различать [9]:

- фильтры нижних частот (ФНЧ), пропускающие частоты, меньшие некоторой частоты среза f_0 ;
- фильтры верхних частот (ФВЧ), пропускающие частоты, большие некоторой частоты среза f_0 ;
- полосовые фильтры (ПФ), пропускающие частоты в некотором интервале $[f_1, f_2]$;
- режекторные (заграждающие) фильтры (РФ), пропускающие все частоты, кроме тех, которые лежат в некотором интервале $[f_1, f_2]$.

Для подавления помехи (4.5) достаточно воспользоваться каскадом режекторных фильтров, настроенных на заграждение g -й гармоники.

При обработке сигнала в реальном масштабе времени алгоритм фильтрации может быть физически реализуем, если для получения фильтрованного значения $y_{\text{вых}}[k]$ в k -й момент времени используется информация, полученная в моменты времени $k, k-1, k-2$ и т. д. В общем случае дискретный фильтр суммирует (с весовыми коэффициентами) некоторое количество входных значений сигнала и некоторое количество предыдущих фильтрованных значений:

$$y_{\text{вых}}[k] = b_0 y_{\text{вх}}[k] + b_1 y_{\text{вх}}[k-1] + \dots + b_m y_{\text{вх}}[k-m] - a_1 y_{\text{вых}}[k-1] - a_2 y_{\text{вых}}[k-2] - \dots - a_n y_{\text{вых}}[k-n]. \quad (4.6)$$

В частном случае, когда $a_1 = a_2 = \dots = a_n = 0$, т. е. алгоритм не использует предыдущие фильтрованные значения, уравнение фильтра сводится к дискретной свертке:

$$y_{\text{вых}}[k] = \sum_{i=0}^m b_i y_{\text{вх}}[k-i]. \quad (4.7)$$

Фильтр, реализующий алгоритм (4.7), называют нерекурсивным фильтром m -го порядка, который, в отличие от рекурсивного фильтра (4.6), всегда является устойчивым [174, с. 203].

Задача проектирования (синтеза) дискретного фильтра сводится к выбору порядка и определению значений параметров a_i, b_i . В тех случаях, когда фазовые характеристики фильтра не важны, такая задача сводится к аппроксимации желаемой амплитудно-частотной характеристики, доставляющей минимум p -нормы отклонения, в частности среднеквадратической ошибке ($p = 2$) либо максимальному отклонению по модулю ($p \rightarrow \infty$).

Поскольку в общем случае такая задача не имеет аналитического решения [174, с. 319], на практике используют различные численные алгоритмы, в частности, метод, основанный на решении системы уравнений Юла—Уокера, метод экспоненциального оценивания Прони и ряд других субоптимальных методов синтеза фильтра.

Следует, однако, заметить, что даже при высоком порядке фильтра невозможно получить «идеальную» амплитудно-частотную характеристику (АЧХ) прямоугольной формы (рис. 4.12, а), обеспечивающую полное подавление частот в полосе заграждения без искажения сигнала в полосе пропускания. Помимо основных параметров (частотных границ полос пропускания f_p и задержания f_s), на практике приходится проектировать заградительный фильтр с учетом дополнительных требований — допустимой неравномерности АЧХ в полосе пропускания R_p и минимально необходимого затухания в полосе заграждения R_s (рис. 4.12, б).

Особые сложности синтеза режекторных фильтров возникают в тех случаях, когда необходимо обеспечить заграждение узкополосной помехи, сосредоточенной в полосе «полезных» час-

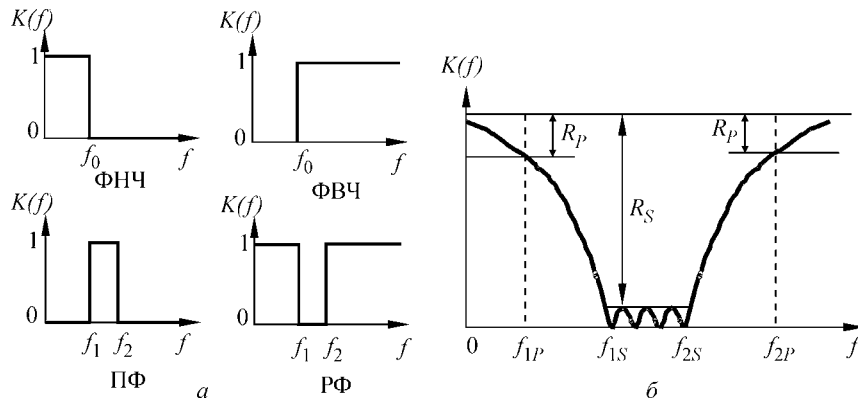


Рис. 4.12. Амплитудно-частотные характеристики фильтров:
 а — «идеальных» фильтров; б — реального заградительного фильтра

тот. При уменьшении разности $\Delta_f = f_{2S} - f_{1S}$ неизбежно увеличиваются параметры R_p, R_s . Это приводит к недопустимым искажениям значений диагностических признаков, сосредоточенных на локальных фрагментах сигнала.

Рассмотрим иной возможный путь подавления сосредоточенных гармонических помех, который в научной литературе называют методом фильтрации с изменением области преобразования сигналов [329]. В отличие от рассмотренных выше методов последовательной обработки сигнала во временной области, такие фильтры проводят блочную обработку всего массива данных $y[1], \dots, y[K]$ в частотной области.

Алгоритм блочной фильтрации может быть основан на применении процедур прямого FFT и обратного $IFFT$ дискретного преобразования Фурье. Известно [87], что процедуры FFT и $IFFT$ взаимно обратимы (с точностью до малых погрешностей округления), т. е. для любого вектора $Y = (y[1], \dots, y[K])$ справедливо равенство

$$IFFT(FFT(Y)) = Y. \quad (4.8)$$

Из (4.8), казалось бы, следует, что для заграждения стационарной гармонической помехи (4.5) достаточно выполнить последовательность трех операций (рис. 4.13):

- построить спектр наблюдаемого сигнала на основе FFT ;

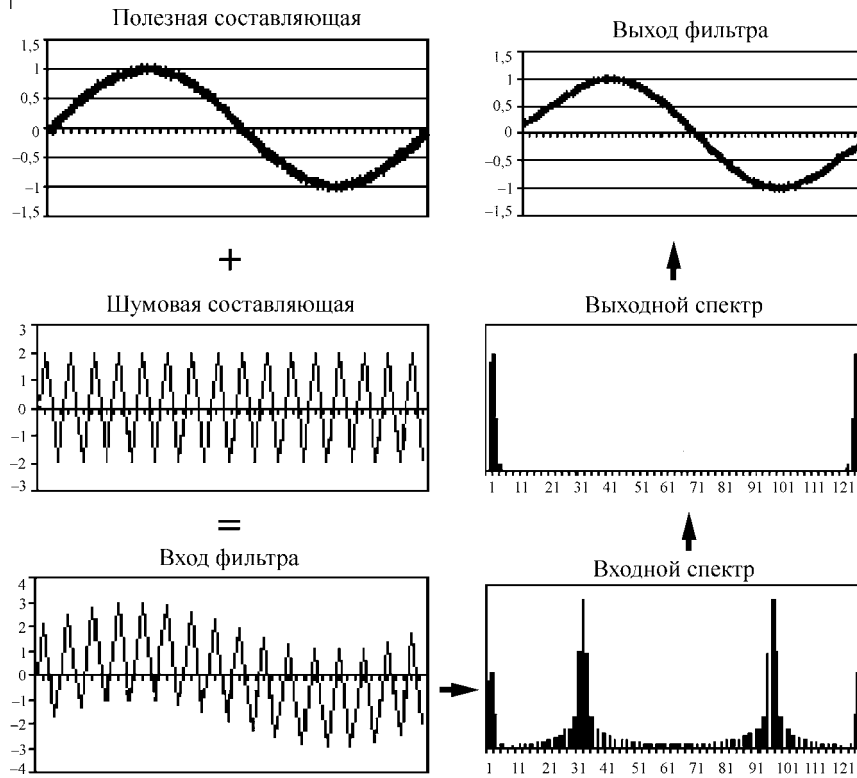


Рис. 4.13. Идея заградительного фильтра, построенного на основе ДПФ

- подавить составляющие спектра, соответствующие частотам помехи $h(\cdot)$;
- провести обратное преобразование на основе $IFFT$ для восстановления полезного сигнала во временной области.

Однако оказывается, что такой прием не всегда эффективен. Дело в том, что спектр дискретного гармонического сигнала

$$h_g[k] = a_g \sin(2\pi f_g k + \varphi_g),$$

построенный на основе прямого ДПФ, «адекватен» истинному спектру помехи $h_g(\cdot)$ только в том случае, когда ее частота f_g совпадает с одной из частот разложения Фурье.

Если же это условие не выполняется, то спектр гармонической помехи будет «распределяться» на целый ряд смежных час-

тот. Такой эффект в научной литературе принято называть растеканием спектра (spectrum leakage) дискретного гармонического сигнала [174, с. 270].

Для иллюстрации этого эффекта рассмотрим спектры двух дискретных гармонических сигналов

$$h_1[k] = \sin(2\pi f_1 \Delta k) \text{ и } h_2[k] = \sin(2\pi f_2 \Delta k), \quad k = 1, \dots, K, \quad (4.9)$$

где $f_1 = 16,5$ Гц и $f_2 = 16,5165432768543$ Гц, а шаг квантования по времени $\Delta = 0,001$ секунд (частота дискретизации $f_d = 1/\Delta = 1000$ Гц).

На рис. 4.14 показаны фрагменты спектров этих сигналов, построенные на основе ДПФ по $K = 30\,000$ точкам. Как видно, только в первом случае, когда частота сигнала совпала с одной из частот разложения Фурье, построенный спектр адекватен реальной ситуации (рис. 4.14, а). Заметим, что частоты f_1 и f_2 сигналов практически совпадают.

Отсюда следует, что если в общем случае для фильтрации гармонической помехи подавлять все гармоники, на которые распределена нагрузка помехи, то при близких частотах помехи и сигнала неизбежно придется затронуть гармоники, несущие информацию о полезном сигнале, и в результате после обратного преобразования Фурье полезный сигнал будет искажен.

Понятно, что для построения эффективной процедуры подавления сосредоточенных гармонических помех (4.5) требуется каким-то образом устранить или хотя бы снизить эффект расте-

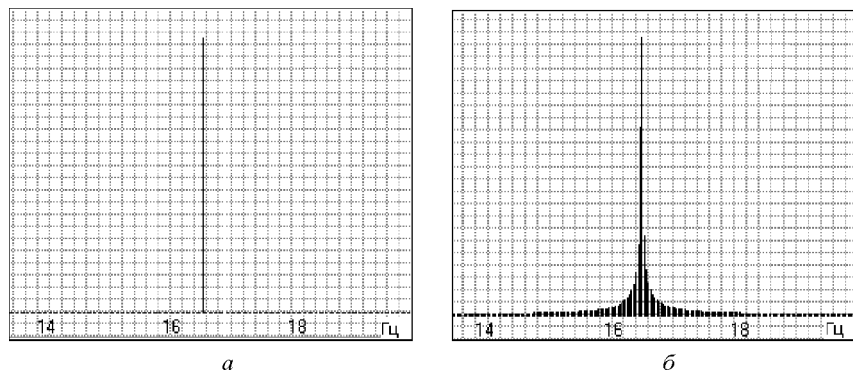


Рис. 4.14. Спектры дискретных гармонических сигналов (4.9):

$a - f_1$ — кратная частота; $b - f_2$ — некратная частота

кания спектра дискретного гармонического сигнала. Поскольку этот эффект вытекает из самой природы ДПФ [174, с. 270], остановимся на нем более подробно.

Рассмотрим дискретный гармонический сигнал

$$h[k] = a \cos(\omega k \Delta + \varphi), \quad k = 0, \dots, K-1, \quad (4.10)$$

где a — амплитуда; Δ — шаг квантования по времени; ω — круговая частота.

Если конечная последовательность (4.10) содержит *целое* число периодов анализируемого сигнала, т. е. если отношение $K\omega\Delta/2\pi$ является *целым* числом, то периодически продолженный сигнал представляет собой гармоническое колебание (без скачков), а подстановка (4.10) в формулу ДПФ (1.20) показывает, что вычисленное ДПФ содержит лишь два спектральных отсчета, отличных от нуля:

$$Y_n = \begin{cases} \frac{aK}{2} e^{i\varphi} & \text{при } n = \frac{\omega\Delta}{2\pi} K, \\ \frac{aK}{2} e^{-i\varphi} & \text{при } n = (1 - \frac{\omega\Delta}{2\pi}) K, \\ 0 & \text{— в остальных случаях.} \end{cases} \quad (4.11)$$

Таким образом, аналогично спектру непрерывного гармонического сигнала, ДПФ отличается от нуля всего для двух значений n . Заметим, что на рис. 4.14, *а* показана лишь одна из этих двух составляющих спектра (при положительном значении частоты).

Однако если отношение $K\omega\Delta/2\pi$ не является целым числом, то периодически продолженная последовательность (4.10) уже не соответствует набору дискретных значений непрерывной косинусоиды. Поэтому, в полном соответствии со свойствами ДПФ, спектр оказывается значительно более богатым: в нем появляются дополнительные составляющие (рис. 4.14, *б*).

Известным приемом, обеспечивающим уменьшение эффекта растекания спектра, является умножение сигнала на весовую функцию $W(k)$ [174, с. 273], которая имеет максимум в центральной точке сигнала ($k = K/2$) и плавно спадает к краям ($k = 0$ и $k = K-1$). В этом случае формула ДПФ принимает вид

$$Y_n(W) = \sum_{k=0}^{K-1} W[k]y[k]e^{-j\frac{2\pi nk}{K}}.$$

Считается, что при удачном выборе $W(k)$ можно несколько ослабить эффекты, связанные с возникновением скачков сигнала при периодическом повторении анализируемой конечной последовательности, и таким образом уменьшить эффект растекания спектра. Однако неизбежной платой за использование весовой функции $W(k)$ является искажение спектра полезного сигнала, что недопустимо при обработке сигналов с локально сосредоточенными признаками.

В связи с изложенным возникает естественное желание попытаться найти способ борьбы с самим явлением, породившим растекание спектра, а не с его следствием. Иными словами, поставим задачу: модифицировать массив обрабатываемых данных, чтобы, не используя весовую функцию $W(k)$, получить адекватный спектр гармонического колебания на основе традиционного алгоритма ДПФ.

4.3.2. Предлагаемый алгоритм

Для решения поставленной задачи предлагается дополнить алгоритм фильтрации сигналов на основе прямого и обратного ДПФ специальной процедурой, обеспечивающей автоматический подбор оптимального числа $K_{\text{opt}} \leq K$ элементов массива, при обработке которых растекание спектра сосредоточенной гармонической помехи будет устранено или, по крайней мере, уменьшено [211].

Как уже говорилось, ДПФ предполагает, что последовательность отсчетов анализируемого сигнала является периодически продолженной вперед и назад по времени. Однако при неизвестной частоте ω помехи невозможно аналитически определить значение K_{opt} , удовлетворяющее этому требованию. Более того, для любого значения частоты ω всегда можно указать такой шаг квантования по времени Δ , при котором периодическое продолжение *конечной* последовательности дискретных значений гармонического сигнала $h[k] = a \sin(\omega \Delta k + \varphi)$, $k = 0, \dots, K-1$, не будет соответствовать дискретным значениям непрерывной синусоиды $h(t) = a \sin(\omega t + \varphi)$.

Поэтому для определения K_{opt} предлагается использовать поисковую процедуру, идея построения которой состоит в следующем.

Поскольку мы предполагаем, что помеха сосредоточена на некоторой фиксированной частоте, но в результате растекания спектра ее мощность распределена на ряд смежных частот, то при уменьшении эффекта растекания неизбежно происходит увеличение амплитуды соответствующей спектральной компоненты ДПФ.

Отсюда следует, что если построить семейство спектральных плотностей при постепенно уменьшающейся длине обрабатываемого массива, то среди них найдется такая спектральная плотность, при которой гармоника, соответствующая частоте помехи, будет иметь наибольшее значение. Найденная таким образом спектральная плотность и определит оптимальное значение K_{opt} .

Для ускорения поиска K_{opt} предлагается ограничить число итераций, задав границы $f_{\min}, f_{\max}$ интервала частот (Гц), на котором предположительно сосредоточена гармоническая помеха, и допустимое уменьшение числа точек δ_K в обрабатываемом массиве.

В результате мы приходим к следующему алгоритму.

Шаг 1. Последовательно сокращаем длину массива, уменьшая число K_j обрабатываемых точек от начального значения K до $K - \delta_K$.

Шаг 2. Для каждого $K_j \in [K - \delta_K, K]$ вычисляем спектральные компоненты $C_n(K_j)$, $n = -K_j/2, \dots, 0, \dots, K_j/2$ (значения модулей ДПФ);

Шаг 3. Определяем оптимальное значение K_{opt} по условию

$$K_{\text{opt}} = \arg \max_{K_j \in [K - \delta_K, K]} \left\{ \frac{\max_{n \in \Omega_h} C_n(k)}{\sum_{n \in \Omega_h} C_n(k)} \right\},$$

где Ω_h — множество номеров спектральных компонент с частотами, принадлежащими интервалу $[f_{\min}, f_{\max}]$.

Шаг 4. В качестве оценки частоты $\hat{f}_g$ помехи принимаем значение максимальной спектральной компоненты на интервале $[f_{\min}, f_{\max}]$, вычисленной при K_{opt} .

Рис. 4.15. Структура узкополосного заградительного фильтра

Процедуру, реализующую описанный выше алгоритм, назовем процедурой повышения контраста спектра помехи. На ее основе реализован эффективный узкополосный заградительный фильтр (рис. 4.15), который включен в состав базовых инструментальных средств системы СИДИГРАФ.



4.3.3. Результаты модельных экспериментов

Узкополосный заградительный фильтр построен на наследнике класса МОД. При реализации фильтра использована процедура Матео Фридо и Стевена Джонсона [325] быстрого вычисления преобразования Фурье.

Проверка эффективности компонент частотной фильтрации проводилась экспериментально на модельных и реальных данных согласно общей схеме организации экспериментов (см. рис. 4.10).

Приведем некоторые из результатов экспериментов.

Эксперимент 1. «Заграждение низких частот». Модели эталона $y_0[k]$ и помехи $h[k]$ представляли собой гармонические колебания с частотами $f_0 = 16,0123456789$ и $f_h = 16,5123456789$ Гц. Сигналы генерировались в $K = 30\,000$ точках при частоте дискретизации $f_D = 1000$ Гц. Отметим, что различие частот сигнала и помехи не превышает 0,1 % частоты Найквиста.

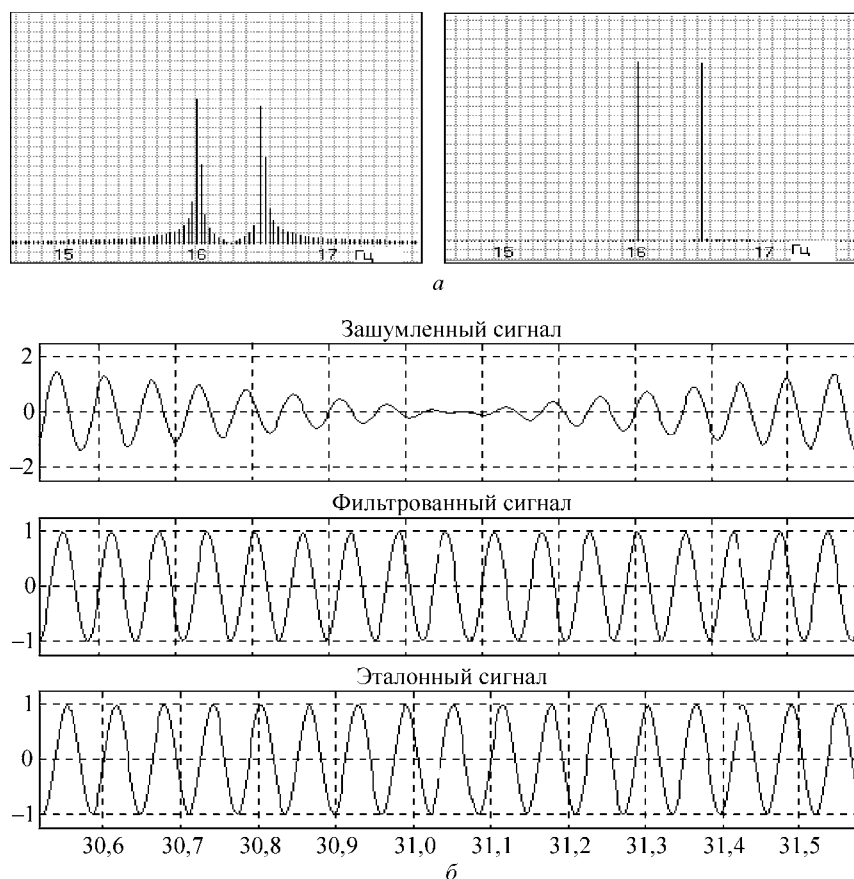
На рис. 4.16, а показан фрагмент спектра суммы указанных гармоник до (слева) и после (справа) применения поисковой процедуры, обеспечивающей уменьшение эффекта растекания спектра. Результат фильтрации представлен на рис. 4.16, б. Для сравнения на этом же рисунке представлен фрагмент эталонного сигнала.

Сравнение фильтрованного и эталонного сигнала показало, что на всей области определения различие их значений не превышало 1 % диапазона изменения эталона.

Эксперимент 2. «Заграждение высоких частот». В условиях данного эксперимента частота полезного сигнала составляла $f_0 = 250,123456789$, а помехи — $f_h = 251,123456789$ Гц. Сумма сигналов генерировалась в $K = 30\,000$ точках при частоте дискретизации $F_D = 1000$ Гц.

При фильтрации высокочастотной помехи получен практически такой же результат (рис. 4.16, в), как и в эксперименте 1.

Эксперимент 3. «Фильтрация сигналов стандартной формы». В качестве эталонов генерировались сигналы стандартной формы в 30 000 точках при частоте дискретизации $F_D = 1000$ Гц. На сигнал



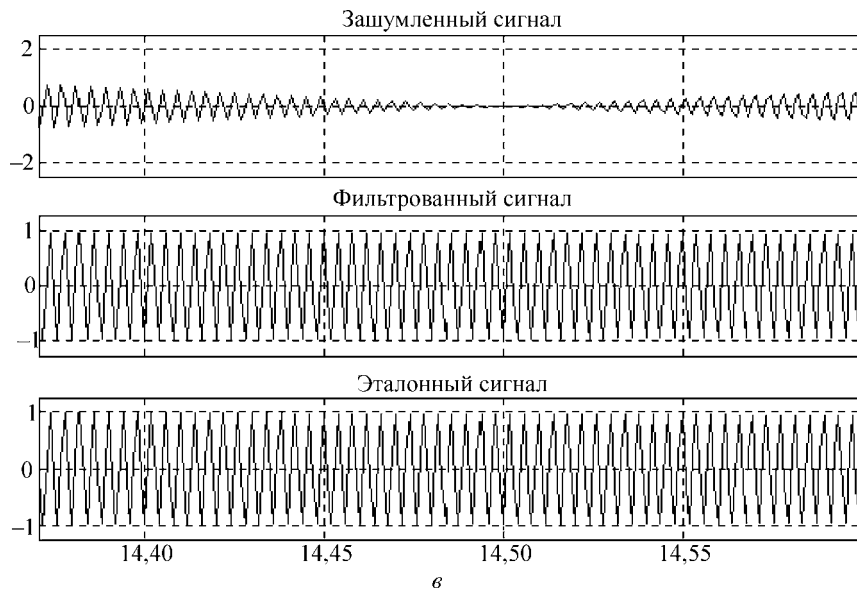


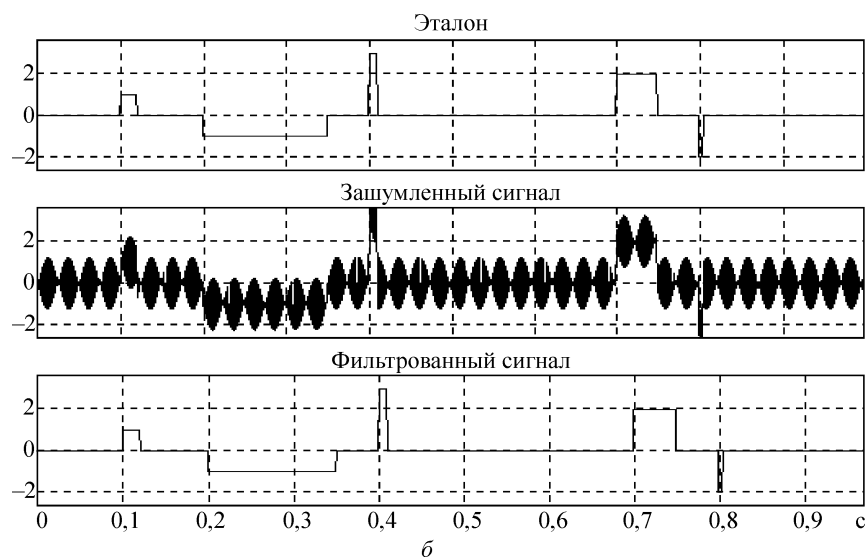
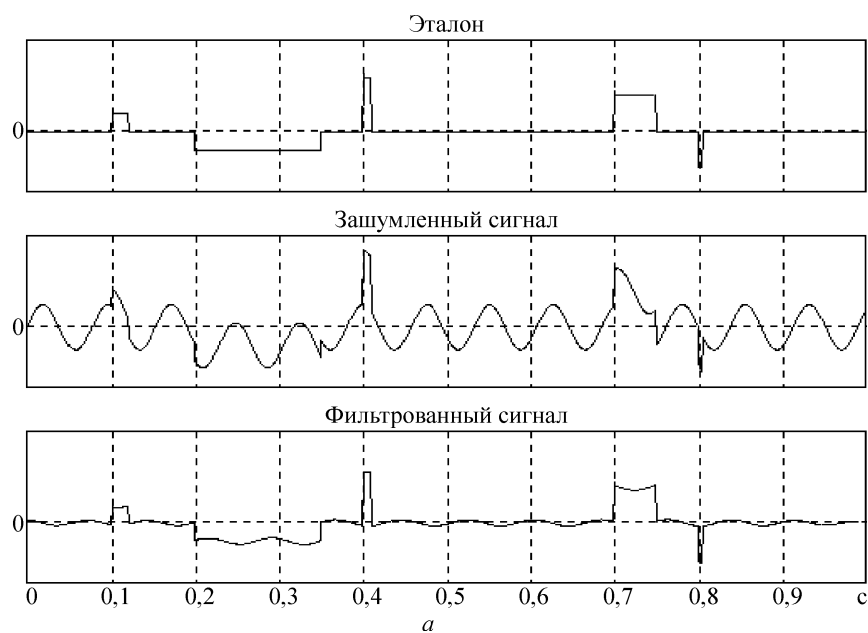
Рис. 4.16. Результат фильтрации частотных помех:

a — спектры сигнала до (слева) и после (справа) использования предложенной процедуры; *б* — фильтрация низкочастотной помехи (эксперимент 1); *в* — фильтрация высокочастотной помехи (эксперимент 2)

накладывалась 50-процентная гармоническая помеха в области низких и высоких частот, причем частота помехи выбиралась не кратной частотам гармоник разложения ДПФ.

При фильтрации меандра, искаженного гармонической помехой с частотой 13,137 Гц (рис. 4.17, *a*), среднеквадратичная ошибка (по отношению к эталону) уменьшилась в 7,67 раза, а максимальное расхождение — в 6,79 раза. На всей области определения расхождение фильтрованного сигнала по отношению к эталону не превышало 7,36 %.

Еще более впечатляющими были результаты фильтрации высокочастотных помех. В результате фильтрации помехи с частотой 479,22 Гц (рис. 4.17, *б*), приближающейся к частоте Найквиста, среднеквадратичная ошибка уменьшилась в 172 раза, а максимальное расхождение — в 122 раза. На всей области определения расхождение фильтрованного сигнала по отношению к эталону не превышало 1,03 %.



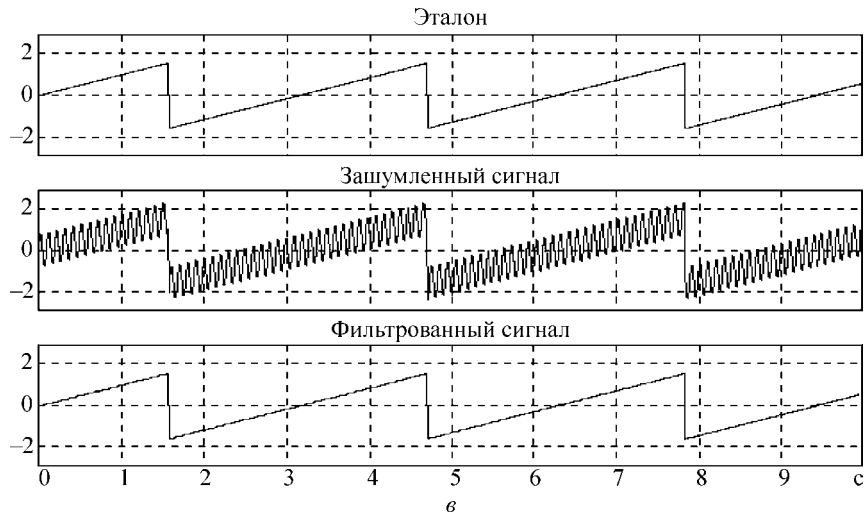


Рис. 4.17. Результаты фильтрации сигналов стандартной формы

Аналогичные результаты получены при фильтрации стандартных сигналов иной формы, в частности пилообразного сигнала, зашумленного помехой 479,22 Гц (рис. 4.17, в).

Эксперимент 4. «Выделение полигармонического сигнала на фоне случайной помехи». В данном эксперименте исследовалась эффективность предложенной процедуры для построения порогового фильтра, обеспечивающего выделение гармонических компонент на фоне случайного шума.

Полезный сигнал представлял собой сумму трех гармонических колебаний

$$y_0[k] = 12 \sin[2\pi f_{01}k] + 7 \sin[2\pi f_{02}k + 0,3] + 9 \sin(2\pi f_{03}k + 0,8)$$

с частотами $f_{01} = 15$ Гц, $f_{02} = 50$ Гц, $f_{03} = 75$ Гц, а аддитивная помеха — последовательность независимых величин, равномерно распределенных на интервале, равном 10 % диапазона изменения $y_0[k]$.

Последовательности $y_0[k]$ и $h[k]$ генерировались в 30 000 точках при частоте дискретизации $F_D = 1000$ Гц.

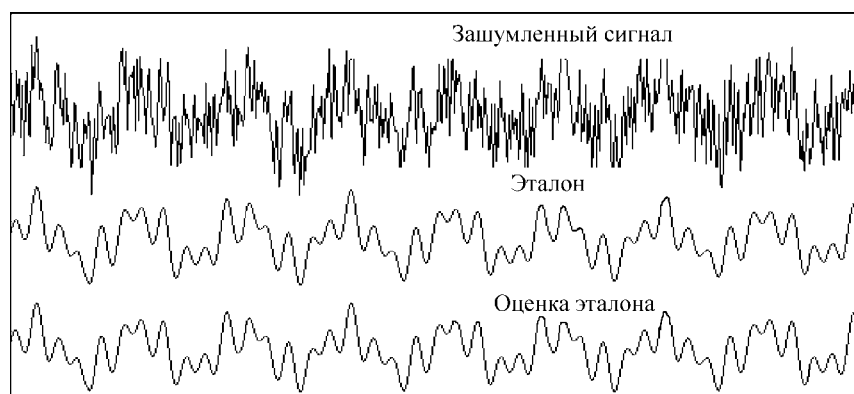


Рис. 4.18. Результат выделения полигармонического сигнала на фоне случайных искажений

Результат обработки (рис. 4.18) еще раз подтвердил эффективность предложенной процедуры.

Эксперимент 5. «Фильтрация электрокардиограммы». В качестве эталона была использована реальная ЭКГ, зашумленная 75 гармоническими помехами с частотами 16,68 Гц и 50,12 Гц. Отметим, что указанные частоты лежат в области информативных частот ЭКГ (рис. 4.19, а), что вызывает известные проблемы при их подавлении традиционным способом.

Применение разработанного узкополосного режекторного фильтра на основе прямого и обратного ДПФ позволило практически без искажений восстановить эталонный сигнал (рис. 4.19, б).

Эксперимент 6. «Фильтрация магнитокардиограммы». В данном эксперименте проводилась фильтрация реальной МКГ здорового пациента, которая в течение 30 секунд регистрировалась с частотой дискретизации $F_D = 1000$ Гц.

Поскольку измерения проводились в реальных клинических условиях с использованием сверхчувствительного измерителя, наблюдаемые сигналы существенно исказились (рис. 4.19, в).

Для фильтрации использовалась последовательность фильтров, построенных на основе предложенной процедуры, в том числе:

- фильтра верхних частот с частотой среза 0,6 Гц;
- узкополосных режекторных фильтров на 16,7 Гц и 50 Гц;
- фильтра нижних частот с частотой среза 90 Гц.

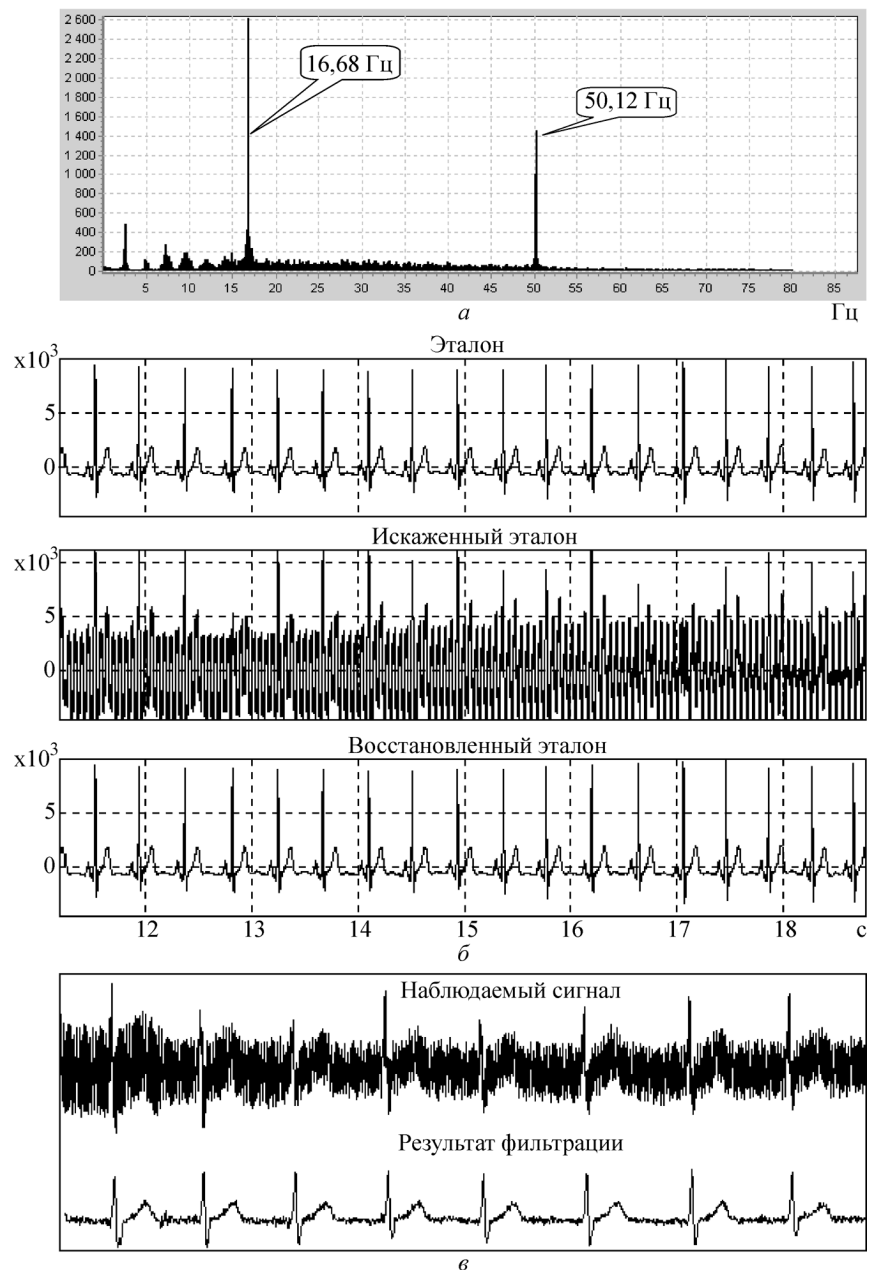


Рис. 4.19. Результат фильтрации физиологических сигналов:
a — спектры наблюдаемых сигналов; *б* — фильтрация ЭКГ; *в* — фильтрация МКГ

Ввиду отсутствия эталона результат фильтрации оценивался визуально по внешнему виду фильтрованной МКГ. Критерием оценки служило наличие на сигнале характерных фрагментов, отражающих работу предсердий и желудочков сердца в виде зубца P , комплекса QRS и сегмента $S-T$.

Таким образом, предложенная процедура обеспечивает эффективное подавление гармонических помех, в том числе расположенных в области информативных частот полезного сигнала.

4.4. Эффективная процедура подавления ограниченных по уровню случайных помех

4.4.1. Анализ традиционных алгоритмов сглаживания

Рассмотрим теперь не менее важный для практики случай, когда аддитивная помеха имеет случайную природу.

В этом случае модель помехи представим последовательностью $h[k]$, $k = 1, 2, \dots, K$, независимых случайных величин, которые неизвестным законом распределены на фиксированном интервале $[-h_0, h_0]$, причем $M\{h\} = 0$, $M\{h[i]h[j]\} = 0$, $i \neq j$, где $M\{\cdot\}$ — знак математического ожидания, а $h_0 > 0$ — константа, ограничивающая уровень помехи.

Понятно, что для подавления случайного шума неэффективно применять частотно-избирательные фильтры, поскольку в общем случае спектр $h[k]$ распределен во всей полосе частот от нуля до частоты Найквиста.

Один из известных подходов к получению оценки полезного сигнала $\tilde{y}_0[k]$ по наблюдению $y[k] = y_0[k] + h[k]$ состоит в использовании алгоритмов сглаживания данных, среди которых наибольшую популярность получили метод экспоненциального сглаживания и метод скользящего среднего.

Простейший метод экспоненциального сглаживания (*Simple Exponential Smoothing*), предложенный Р. Брауном [264], дает оценку $\tilde{y}_0[k]$ в виде

$$\tilde{y}_0[k] = \alpha y[k] + (1 - \alpha)\tilde{y}_0[k - 1], \quad (4.12)$$

где α — параметр сглаживания, $\alpha = \text{const}$, $0 < \alpha < 1$.

Рекуррентное соотношение (4.12) можно представить в эквивалентной форме

$$\tilde{y}_0[k] = \tilde{y}_0[k-1] + \alpha(y[k] - \tilde{y}_0[k-1]),$$

откуда следует, что оценка сглаженного значения в k -й момент определяется оценкой, полученной в $k-1$ -й момент времени, плюс доля α разницы между текущим наблюдением и оценкой в предшествующий момент времени.

Если последовательно использовать рекуррентное соотношение (4.12), то оценку $\tilde{y}_0[k]$ можно выразить посредством значений временного ряда [126]:

$$\begin{aligned} \tilde{y}_0[k] &= \alpha y[k] + \alpha(1-\alpha)y[k-1] + \alpha(1-\alpha)^2 y[k-2] + \dots + \\ &+ (1-\alpha)^i y[k-i] + \dots + (1-\alpha)^{W_0} \tilde{y}_{00} = \\ &= \alpha \sum_{i=0}^{W_0-1} (1-\alpha)^i y[k-i] + (1-\alpha)^{W_0} \tilde{y}_{00}, \end{aligned}$$

где W_0 — количество членов ряда; $\tilde{y}_{00}$ — величина, характеризующая начальные условия для первого применения формулы (4.12) при $k=1$.

Поскольку $0 < \alpha < 1$, то $(1-\alpha)^{W_0} \rightarrow 0$ при $W_0 \rightarrow \infty$, а значит,

$$\tilde{y}_0[k] = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i y[k-i]. \quad (4.13)$$

Таким образом, оценка $\tilde{y}_0[k]$ представляет собой взвешенную сумму всех предыдущих членов временного ряда с экспоненциально уменьшающимися весами.

Пусть $y[k] = y_0[k] + h[k]$, причем $y_0[k] \equiv y_0 = \text{const}$, а $h[k]$ — последовательность независимых случайных величин с нулевым средним и дисперсией σ^2 . В этом случае (4.13) можно записать в виде

$$\tilde{y}_0[k] = y_0 + \alpha \sum_{i=0}^{\infty} (1-\alpha)^i h[k-i]. \quad (4.14)$$

Из (4.14) следует [126, с. 8], что математическое ожидание оценки $\tilde{y}_0$ равно y_0 , а дисперсия определяется соотношением

$$\begin{aligned} D\{\tilde{y}_0\} &= M\{(\tilde{y}_0 - y_0)^2\} = M\{(\alpha \sum_{i=0}^{\infty} (1-\alpha)^i h[k-i])^2\} = \\ &= \alpha^2 \sum_{i=0}^{\infty} (1-\alpha)^{2i} \sigma^2 = \frac{\alpha}{2-\alpha} \sigma^2. \end{aligned} \quad (4.15)$$

Поскольку $0 < \alpha < 1$, то на основании (4.15) можно заключить, что дисперсия оценки $\tilde{y}_0$ меньше дисперсии шума, т. е. $D\{\tilde{y}_0\} < \sigma^2$, причем чем меньше α , тем в большей степени сокращается дисперсия $\tilde{y}_0$.

Известно несколько разновидностей метода экспоненциального сглаживания, в частности двухпараметрический метод Холта (*Holt's method*) [303]:

$$\begin{aligned}\tilde{y}_0[k] &= \alpha y[k] + (1 - \alpha)(\tilde{y}_0[k - 1] + t[k - 1]), \\ t[k] &= \beta(y[k] - y[k - 1]) + (1 - \beta)t[k - 1],\end{aligned}$$

где β — параметр сглаживания тренда, а также трехпараметрический метод Уинтерса (*Winter's method*) [382]:

$$\begin{aligned}\tilde{y}_0[k] &= \alpha(y[k] / s[k - p]) + (1 - \alpha)(\tilde{y}_0[k - 1] + t[k - 1]), \\ t[k] &= \beta(\tilde{y}_0[k] - \tilde{y}_0[k - 1]) + (1 - \beta)t[k - 1], \\ s[k] &= \gamma(y[k] / \tilde{y}_0[k]) + (1 - \gamma)s[k - 1],\end{aligned}$$

в котором фигурирует дополнительная константа γ сглаживания циклов с периодом p .

Традиционный алгоритм скользящего среднего (*Moving Average*) обеспечивает сглаживание зашумленных данных в соответствии с формулой

$$\tilde{y}_0[k] = \frac{1}{W_0} \sum_{j=0}^{W_0-1} y[k - j], \quad (4.13)$$

где W_0 — ширина окна сглаживания. Видно, что степень сглаживания увеличивается с увеличением W_0 , причем $D\{\tilde{y}\} = \sigma^2 / W_0$.

Поскольку

$$\sum_{j=0}^{W_0-1} y[k - j] = \sum_{j=0}^{W_0-1} x[k - 1 - j] + y[k] - y[k - W_0],$$

то формула (4.13) легко может быть преобразована в рекуррентную форму

$$\tilde{y}_0[k] = \tilde{y}_0[k - 1] + \lambda(y[k] - y[k - W_0]), \quad (4.14)$$

где $\lambda = 1 / W_0$.

Преимущество формулы (4.14) по сравнению с (4.13) очевидно — скорость работы алгоритма не зависит от ширины окна сглаживания W_0 .

Эксперименты показали, что, несмотря на различие формул (4.12) и (4.14), при соответствующей настройке параметров эти процедуры обеспечивают практически одинаковую степень сглаживания зашумленных данных.

Отметим, что процедуры (4.12) и (4.14) используют лишь предшествующие значения зашумленного сигнала. Поэтому они могут быть применены в системах реального времени, что является их несомненным преимуществом.

В тех же случаях, когда допустимо вести обработку сразу всего массива $y[1], \dots, y[K]$, для оценки сглаженного значения $\tilde{y}_0[k]$ целесообразно использовать не только предшествующие, но и последующие значения зашумленного сигнала. В этом случае процедуру скользящего среднего можно записать в виде

$$\tilde{y}_0[k] = \frac{1}{2W_0 + 1} \sum_{j=-W_0}^{W_0} y[k - j]. \quad (4.15)$$

После очевидных преобразований формулу (4.15) можно представить в рекуррентной форме

$$\tilde{y}_0[k] = \tilde{y}_0[k - 1] + \mu(y[k + W_0] - y[k - 1 - W_0]), \quad (4.16)$$

где $\mu = 1/(1 + 2W_0)$. Понятно, что процедура (4.16) позволяет оценивать сглаженные значения сигнала при $W_0 \leq k \leq K - W_0$.

Отметим, что при сглаживании, согласно (4.15), все точки в окне $1 + 2W_0$ усредняются с равным весом. Такая процедура может рассматриваться как дискретная свертка обрабатываемого сигнала с ядром прямоугольной формы. Однако можно воспользоваться и другими симметричными ядрами, в частности известными из [336, 337] окнами Бартлетта (*Bartlett window*), Гаусса (*Gaussian window*), Хемминга (*Hamming window*), Блэкмана (*Blackman window*), Ханна (*Hann window*) и др. В этом случае вес более далеких от центра окна точек постепенно уменьшается, приближаясь к нулю на границах окна.

Покажем, что аналогичный эффект может быть достигнут многократным применением процедуры (4.15). Для простоты предположим, что окно сглаживания $1 + 2W_0$ равно трем точкам. Тогда, согласно (4.15), на первом шаге итерации сглаженные

значения сигнала в моменты времени $k-1$, k и $k+1$ определяются выражениями

$$\tilde{y}_0^{(1)}[k-1] = \frac{1}{3}y[k-2] + \frac{1}{3}y[k-1] + \frac{1}{3}y[k],$$

$$\tilde{y}_0^{(1)}[k] = \frac{1}{3}y[k-1] + \frac{1}{3}y[k] + \frac{1}{3}y[k+1],$$

$$\tilde{y}_0^{(1)}[k+1] = \frac{1}{3}y[k] + \frac{1}{3}y[k+1] + \frac{1}{3}y[k+2].$$

Применяя повторно процедуру (4.15) к полученным величинам $\tilde{y}_0^{(1)}[k-1]$, $\tilde{y}_0^{(1)}[k]$, $\tilde{y}_0^{(1)}[k+1]$, оценим сглаженное значение $\tilde{y}_0^{(2)}[k]$ на втором шаге итерации:

$$\begin{aligned}\tilde{y}_0^{(2)}[k] &= \frac{1}{3}\tilde{y}_0^{(1)}[k-1] + \frac{1}{3}\tilde{y}_0^{(1)}[k] + \frac{1}{3}\tilde{y}_0^{(1)}[k+1] = \\ &= \frac{1}{9}y[k-2] + \frac{2}{9}y[k-1] + \frac{1}{3}y[k] + \frac{2}{9}y[k+1] + \frac{1}{9}y[k+2].\end{aligned}$$

Легко заметить, что теперь уже весовая функция становится треугольной.

Следует, однако, обратить внимание на один общий недостаток рассмотренных выше традиционных процедур сглаживания, который продемонстрируем модельным примером.

Пусть наблюдаемый сигнал (ЭКГ) зашумлен 20-процентной случайной помехой с нулевым математическим ожиданием (рис. 4.20, а). При сглаживании этого сигнала процедурой (4.15) с окном из 7 точек не удастся получить требуемое сокращение шумовой компоненты на информативном фрагменте $S-T$ (рис. 4.20, б). Только при ширине окна из 30 точек сглаживание приводит к желательному сокращению шумовой компоненты на фрагменте $S-T$, однако в этом случае происходит недопустимое снижение амплитуды другого информативного фрагмента — комплекса QRS (рис. 4.20, в), измерение которого также важно для интерпретации ЭКГ.

Таким образом, при использовании традиционного алгоритма скользящего среднего мы сталкиваемся с *проблемой*, обусловленной противоречием выбора оптимального значения параметра сглаживания W_0 : при уменьшении W_0 не обеспечивается желательный эффект подавления шумовой компоненты, а при уве-

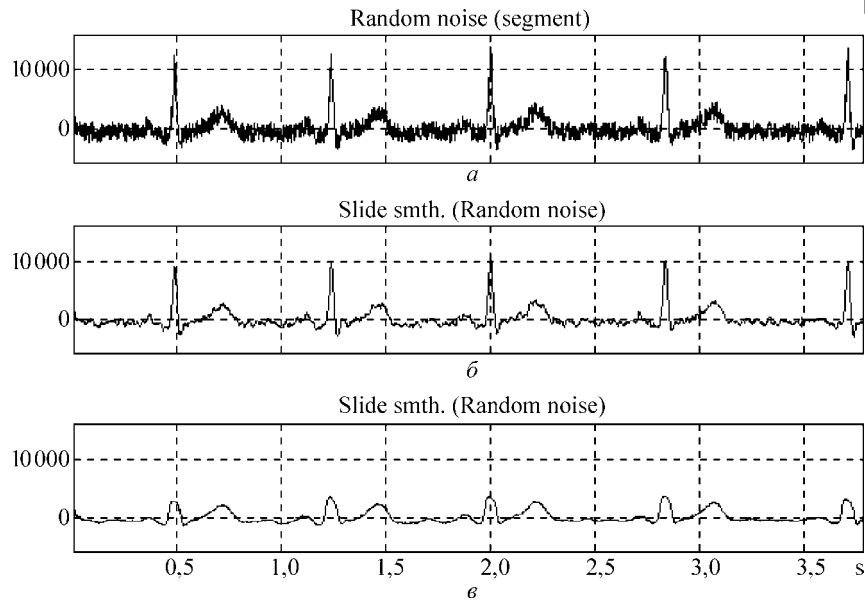


Рис. 4.20. Результаты применения процедуры (4.15) для сглаживания ЭКГ: исходный сигнал (а); результат сглаживания при $W_0 = 7$ (б) и $W_0 = 30$ (в)

личении W_0 происходят недопустимые искажения полезного сигнала.

Аналогичное противоречие возникает и при выборе оптимального значения параметра α алгоритма экспоненциального сглаживания (4.12).

Поэтому в задачах обработки сигналов с локально сосредоточенными признаками традиционные алгоритмы сглаживания можно использовать лишь при весьма малых уровнях шумовой компоненты, что, естественно, ограничивает область их практического применения.

4.4.2. Алгоритм адаптивного сглаживания

Для устранения отмеченного недостатка в работе [210] предложен усовершенствованный алгоритм скользящего среднего. Основная идея модификации состоит в следующем. Поскольку предполагается, что уровень случайной помехи $h(k)$ по модулю ограничен величиной h_0 , всякое отклонение сглаженного значе-

ния $\tilde{y}_0[k]$ от наблюдаемого $y[k]$, превышающее h_0 , можно объяснить лишь тем, что сигнал чрезмерно сглажен. Поэтому предлагается приспособить (адаптировать) параметр W к обрабатываемому сигналу таким образом, чтобы, с одной стороны, обеспечить максимальную степень сглаживания, а с другой — не допустить искажение сигнала, превышающее заданный порог h_0 .

Формально предлагаемый алгоритм реализуется формулой

$$\tilde{y}_0[k] = \frac{1}{2W_k + 1} \sum_{j=-W_k}^{W_k} y[k - j], \quad (4.17)$$

которая отличается от (4.15) тем, что число усредняемых точек в окне зависит от номера k обрабатываемой точки.

Значение параметра W_k , фигурирующего в (4.17), определяется из условий

$$W_k \leq W_0, \quad (4.18)$$

$$|\tilde{y}[k] - y[k]| \leq h_0, \quad (4.19)$$

$$|W_k - W_{k-1}| \leq 1,$$

где W_0 — ограничение на параметр сглаживания, задаваемый пользователем.

В отличие от традиционного алгоритма скользящего среднего (4.15), предлагаемый алгоритм предусматривает два этапа обработки массива зашумленных данных. На первом этапе определяются оптимальные значения $W_1, \dots, W_K$ для всех обрабатываемых точек сигнала $y[k]$, $k = 1, \dots, K$, а на втором — сглаживание входного массива с использованием оптимальных параметров.

Процедура определения W_k реализована следующим образом. При заданном значении W_0 выбирается первый блок $1 + 2W_0$ обрабатываемых точек с порядковыми номерами $j = 1, 2, \dots, 1 + 2W_0$ и оценивается сглаженное значение в точке с порядковым номером $i = W_0$ по формуле

$$\tilde{y}[i] = \frac{1}{2W_j + 1} \sum_{j=-W_j}^{W_j} y_{i+j}, \quad (4.20)$$

при начальном значении параметра сглаживания $W_i = W_0$.

Если выполняется условие

$$|y[i] - \tilde{y}[i]| > h_0, \quad (4.21)$$

то значение W_i уменьшается на единицу и формула (4.20) применяется к массиву точек с порядковыми номерами $j = 1, 2, \dots, 2W_0$, т. е. первый блок уменьшается на одну точку.

Описанный итерационный процесс сокращения длины обрабатываемого блока продолжается до тех пор, пока первый раз не нарушится условие (4.21). Как только это произойдет, выбирается следующий блок $1 + 2W_0$ точек с порядковыми номерами $j = 2, \dots, 2 + 2W_0$, которые обрабатываются аналогичным образом, и т. д. до конца массива.

В результате для каждой k -й точки выбирается значение $W_k^{(1)}$, удовлетворяющее условиям (4.18) и (4.19). Окончательным шагом адаптации параметров сглаживания является модификация найденных значений $W_k^{(1)} \rightarrow W_k^{(2)} \rightarrow W_k^{(3)}$, такая, чтобы обеспечить выполнение «плавности» изменения окон по ходу обработки сигнала.

Предположим, например, что $W_0 = 30$ и на первом этапе получена последовательность параметров

$$W_k^{(1)} = \{25, 28, 30, 15, 18, 23, 27, 26, 30, 29, \\ 18, 24, 30, \dots, 24, 28, 26\},$$

удовлетворяющих условиям (4.18), (4.19). Тогда переход от $W_k^{(1)}$ к $W_k^{(2)}$, обеспечивающий «правостороннюю плавность» значений, осуществляется следующим образом:

$$W_{i+1}^{(2)} = \begin{cases} W_i^{(2)} + 1, & \text{если } W_{i+1}^{(1)} > W_i^{(1)}, \\ W_{i+1}^{(1)} & \text{в противном случае.} \end{cases}$$

В результате массив параметров приобретает вид

$$W_k^{(2)} = \{25, 26, 27, 15, 16, 17, 18, 19, 20, 21, \\ 18, 19, 20, \dots, 24, 25, 26\}.$$

Далее осуществляется модификация $W_k^{(2)}$, обеспечивающая «левостороннюю плавность»:



Рис. 4.21. Интерфейс компоненты адаптивного сглаживания

$$W_{i-1}^{(3)} = \begin{cases} W_i^{(3)} - 1, & \text{если } W_{i-1}^{(2)} < W_i^{(2)}, \\ W_{i-1}^{(2)} & \text{в противном случае.} \end{cases}$$

В результате формируется конечный массив $W_k^{(3)}$ оптимальных значений параметров сглаживания с «двухсторонней» плавностью:

$$W_k^{(3)} = \{18, 17, 16, 15, 16, 17, 18, 19, 20, 19, \\ 18, 19, 20, \dots, 24, 25, 26\},$$

который и используется для сглаживания сигнала.

Для удобства настройки алгоритма в инструментальной системе предусмотрен орган плавного управления порогом h_0 при одновременном просмотре графических образов исходного и сглаженного сигналов (рис. 4.21).

4.4.3. Результаты экспериментальных исследований

Оценки эффективности предложенного алгоритма проводилось на модельных и реальных данных согласно общей схеме организации экспериментов (см. рис. 4.10).

В качестве эталонов использовались типовые сигналы стандартной формы в виде одиночных импульсов и их последова-

тельностью с изменяющимися во времени параметрами. Помеха генерировалась в виде последовательности независимых случайных величин, равномерно распределенных на интервале $[-h_0, h_0]$, величина которого задавалась в процентах от диапазона Δ_0 эталона y_0 .

Результаты сглаживания традиционным и адаптивным алгоритмами оценивались максимальными $\Delta_{\max 1}, \Delta_{\max 2}$ и среднеквадратическими σ_1, σ_2 отклонениями сглаженных сигналов относительно эталона (табл. 4.2).

Приведем более подробно результаты сглаживания (рис. 4.21 и 4.22). В верхней части рисунков показан эталонный сигнал (слева) и сигнал, зашумленный помехой (справа). В нижней части рисунков представлены результаты сглаживания традиционным алгоритмом (слева) и адаптивным (справа).

Рис. 4.22, а иллюстрирует результат сглаживания импульса единичной амплитуды, ширина которого составляла 1/3 области определения сигнала. Сигнал был сгенерирован в 1000 точках и за-

ТАБЛИЦА 4.2

Результаты тестирования алгоритма адаптивного сглаживания

Эталонный сигнал y_0			Помеха h		Результаты сглаживания			
Модель (форма)	Δ_0	σ_0	%	σ_h	Традиционный		Адаптивный	
					$\Delta_{\max 1}$	σ_1	$\Delta_{\max 2}$	σ_2
Прямоугольный импульс (широкий)	1	0,47	20	0,058	0,49	0,08	0,07	0,01
Прямоугольный импульс (узкий)	1	0,08	20	0,056	0,84	0,08	0,09	0,01
Сложный меандр	5	0,69	5	0,073	0,91	0,13	0,06	0,02
			20	0,285	1,01	0,26	0,52	0,06
			50	0,720	2,25	0,34	0,76	0,09
Пилообразное колебание	3,14	0,89	20	0,182	1,52	0,09	0,26	0,03
Треугольное колебание	3,14	0,86	20	0,181	0,15	0,03	0,11	0,03
Затухающая синусоида	1,57	0,22	5	0,023	0,033	0,008	0,021	0,007
Последовательность гауссовых импульсов	1,27	0,18	25	0,121	0,22	0,06	0,11	0,03
Последовательность затухающих гауссовых импульсов	1,7	0,13	10	0,049	0,38	0,08	0,23	0,03

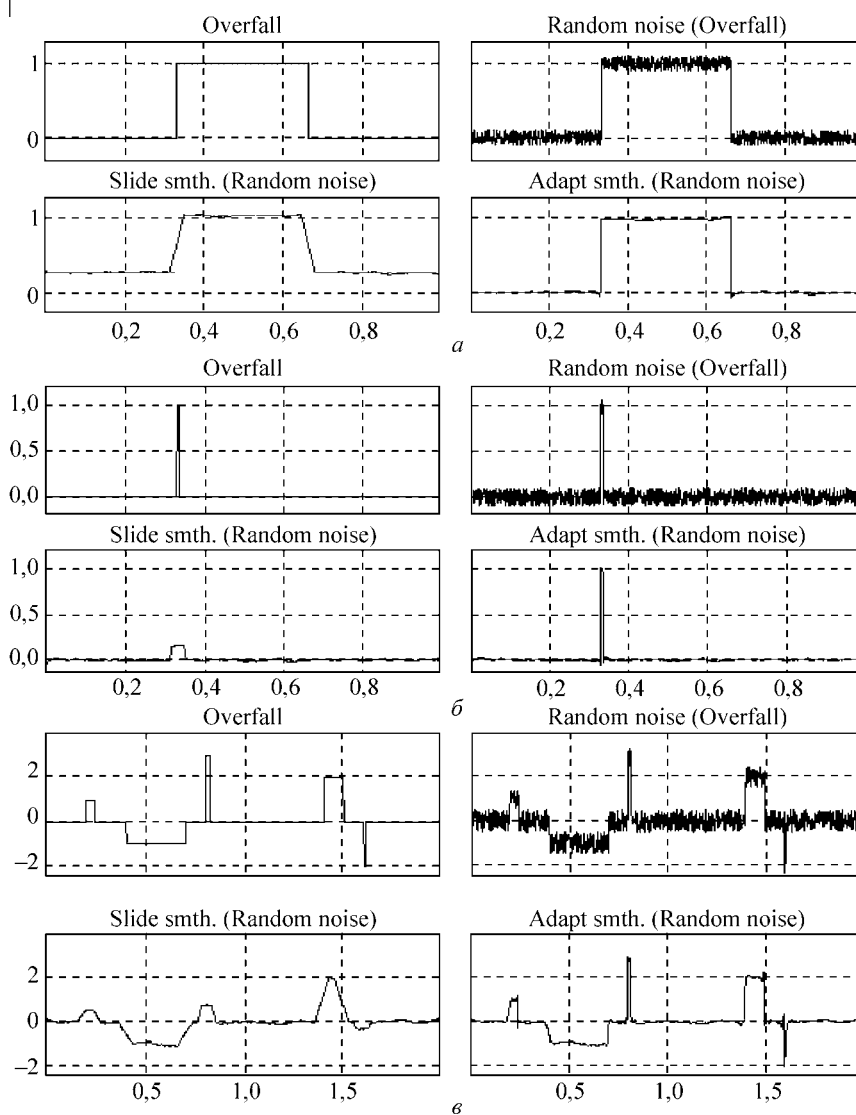


Рис. 4.22. Результаты сглаживания сигналов прямоугольной формы

шумлен 20-процентной помехой. Параметры настройки составляли $W_0 = 18$ для обоих алгоритмов и $h_0 = 10\%$ для адаптивного алгоритма.

Эксперименты показали, что среднеквадратическая ошибка традиционного алгоритма составила 0,08 ед. (8 % вариации эталонного сигнала), тогда как ошибка адаптивного алгоритма была равна 0,01 ед. (1 % вариации эталонного сигнала).

На рис. 4.22, б показаны результаты, полученные при сглаживании узкого прямоугольного импульса, ширина которого составляла всего 6 точек (0,6 % области определения сигнала). Уровень помехи по-прежнему равен 20 %.

В данном случае преимущества предлагаемого алгоритма оказались еще более убедительными: после сглаживания традиционным алгоритмом амплитуда импульса уменьшилась более чем в пять раз, в то время как при сглаживании адаптивным алгоритмом амплитуда короткого импульса практически не изменилась.

Для исследования эффективности алгоритмов при обработке последовательностей прямоугольных импульсов различной продолжительности и амплитуды генерировался эталонный сигнал в виде сложного меандра. Сигнал генерировался в 1000 точках и содержал 5 прямоугольных импульсов, ширина которых составляла 5, 10, 20 и 50 точек, а амплитуды изменялись от -2 ед. до $+3$ ед. Эксперимент проводился с 5, 2 и 50-процентной помехой (см. табл. 4.2).

Результат обработки эталонного меандра (рис. 4.22, в) при уровне шума 20 % показал, что среднеквадратическая ошибка традиционного алгоритма составила 0,26 ед. (5,2 % вариации эталонного сигнала), тогда как ошибка адаптивного алгоритма была равна 0,06 ед. (1,2 % вариации эталонного сигнала).

При сглаживании сигналов другой формы (рис. 4.23) были получены практически те же результаты. Хотя при сглаживании пилообразного колебания, зашумленного 20-процентной помехой (рис. 4.23, а), оба алгоритма продемонстрировали внешне достаточно близкие результаты, однако численный анализ показал, что на задних фронтах импульсов отклонение ошибки традиционного алгоритма было в 5,8 раза выше, чем адаптивного. При этом в первом случае среднеквадратическая ошибка составляла 0,09 ед. (2,9 % от диапазона эталонного сигнала), а во втором — всего 0,03 ед. (0,96 % от диапазона эталонного сигнала).

Для исследования эффективности алгоритмов при обработке гармонических колебаний использовались различные эталонные сигналы, в том числе затухающая синусоида и последовательности гауссовых гармонических колебаний.

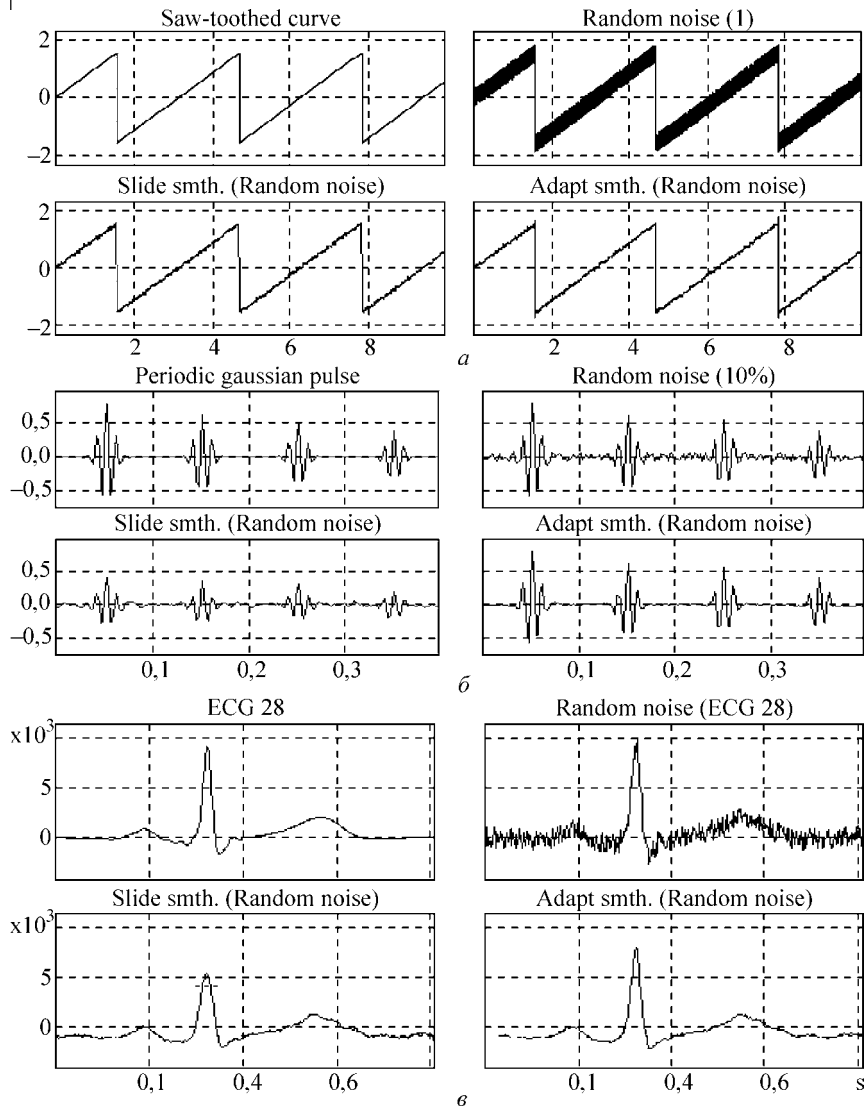


Рис. 4.23. Результаты сглаживания пилообразного колебания (*a*) затухающих гармонических гауссовых импульсов (*б*) и цикла ЭКГ (*в*)

Результат одного из таких экспериментов — сглаживание последовательности гармонических гауссовых импульсов с центральной частотой 100 Гц и шириной полосы 50 %, которые по-

вторялись с периодичностью 10 Гц. При этом амплитуда каждого последующего импульса снижалась на 20 % по отношению к амплитуде предыдущего.

Эталон, искаженный 10-процентной помехой (рис. 4.23, б), сглаживался традиционным и адаптивным алгоритмами. Легко заметить, что и в этом случае адаптивный алгоритм показал лучшие результаты, что подтверждается данными, приведенными в табл. 4.2.

При сглаживании эталонного цикла ЭКГ, зашумленного 20-процентной помехой (рис. 4.23, в), адаптивный алгоритм по сравнению с традиционным позволил более чем в 3 раза уменьшить искажение амплитуды комплекса *QRS* при достаточно хорошем сглаживании помехи, искажающей фрагмент *S—T*, причем остаточная среднеквадратическая ошибка при сглаживании традиционным алгоритмом составила 4,11 %, а адаптивным — 2,2 % диапазона изменения ЭКГ.

Алгоритм адаптивного сглаживания оказался полезным при регуляризации процедур численного дифференцирования. Поскольку процедуры оценки производных дискретно заданного сигнала использованы при обработке сигнала в фазовом пространстве, рассмотрим их более подробно.

Известно, что в общем случае дифференцирование зашумленных сигналов относится к числу некорректно поставленных задач [191]. Это легко продемонстрировать на следующем характерном примере.

Предположим, что наблюдаемый сигнал

$$y(t) = y_0(t) + h(t) \quad (4.22)$$

представляет собой сумму полезного сигнала $y(t)$ и помехи $h(t)$, причем

$$h(t) = h_0 \sin(\omega t), \quad (4.23)$$

где $h_0 = \text{const}$, $\omega = \text{const}$.

Из (4.22) с учетом (4.23) следует, что

$$|y(t) - y_0(t)| \leq h_0. \quad (4.24)$$

В соответствии с (4.22) имеем

$$\dot{y}(t) = \dot{y}_0(t) + \omega h_0 \cos(\omega t), \quad (4.25)$$

и, следовательно,

$$|\dot{y}(t) - \dot{y}_0(t)| \leq \omega h_0. \quad (4.26)$$

Из сравнения (4.24) и (4.26) видно, что при сколь угодно малом отклонении Δ_y функций $y(t)$ и $y_0(t)$, вызванном аддитивной помехой $|h(t)| \leq h_0$, возможно сколь угодно большое отклонение $\Delta_{\dot{y}}$ производных $\dot{y}(t)$ и $\dot{y}_0(t)$, которое зависит от частоты ω .

Отметим также, что дифференцирование оцифрованного сигнала является некорректной процедурой даже при отсутствии внешней шумовой компоненты $h[k]$, поскольку из-за ошибок квантования по уровню дискретные значения оцифрованного сигнала $y[k]$, $k = 1, \dots, K$, отличаются от истинных значений функции $y(t)$ в моменты времени $t = t_k$.

В то же время для дискретно заданной функции в предположении о существовании ее производных можно вычислить приближенные значения производных на основе конечно-разностных аппроксимаций (приближений).

Один из известных методов численной оценки значения производной $\dot{y}[k]$ в k -й точке дискретного сигнала $y[k]$ основан на дифференцировании интерполяционного полинома Лагранжа, построенного по n узлам в окрестности k -й точки. При этом ошибка аппроксимации становится наименьшей, если k -я точка является центральной среди узлов интерполяции [197, с. 83].

Используя общую формулу, приведенную в [197], легко показать, что при числе узлов $n = 7$ и единичном шаге квантования оценка производной в k -й точке может быть получена по формуле

$$\begin{aligned} \dot{y}[k] = 1/60 (y[k+3] - 9y[k+2] + 45y[k+1] - \\ - 45y[k-1] + 9y[k-2] - y[k-3]), \end{aligned} \quad (4.27)$$

причем точность такой оценки характеризует остаточный член

$$R = \frac{1}{140} y_v^{(7)},$$

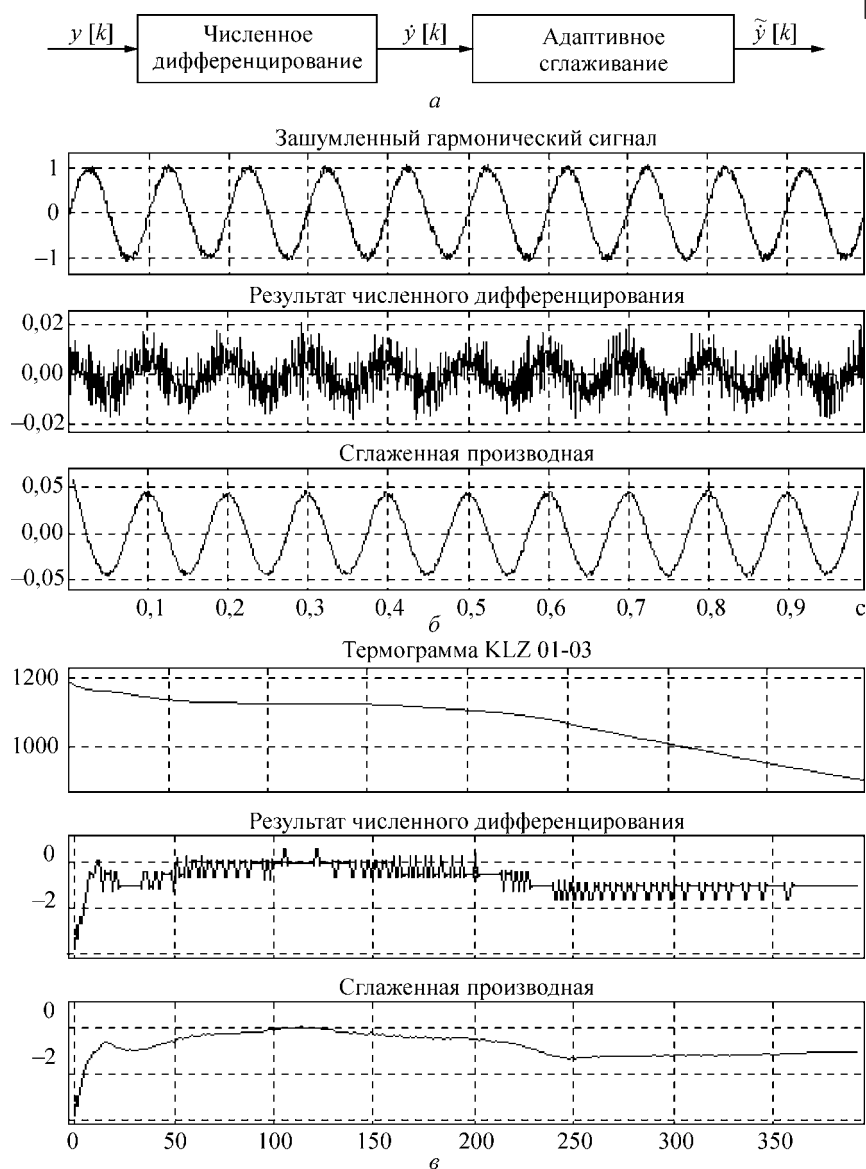


Рис. 4.24. Компонента численного дифференцирования сигналов:

a — структура компоненты; *б* — результат численного дифференцирования гармонического сигнала, зашумленного 10-процентной случайной помехой; *в* — результат численного дифференцирования реальной термограммы кристаллизации пробы чугуна

где $y_v^{(7)}$ — значение производной седьмого порядка функции $y(t)$ в некоторой внутренней точке v текущего интервала интерполяции.

Понятно, что оценить точность вычисления производной по приведенной выше формуле не представляется возможным, поскольку нам не известны истинные значения производных сигнала, заданного массивом дискретных значений. Однако для решения задачи построения фазовых траекторий сигналов важна не столько точность оценки производной $y(t)$, сколько сохранение «формы» функции, аппроксимирующей производную так, чтобы особые точки исходного сигнала (экстремумы, изломы и т. п.) адекватно отображались на этой функции.

Таким образом, в инструментальную систему включена вычислительная компонента, предусматривающая обработку сигнала в два этапа: оценку производной предварительно фильтрованного сигнала по формуле (4.27) и адаптивное сглаживание полученных данных на основе алгоритма (4.17), который обеспечивает регуляризацию процедуры численной оценки производной (рис. 4.24, *а*). При этом настройка параметров алгоритма адаптивного сглаживания определяется экспериментально для конкретного типа обрабатываемых сигналов.

Эксперименты подтвердили, что при ограниченном уровне шума, который в экспериментах достигал 10 % диапазона изменения сигнала, форма функции, аппроксимирующей производную, адекватно отражала особые точки, отделяющие один фрагмент сигнала от другого (рис. 4.24, *б*, *в*).

4.5. Каскадная процедура предварительной обработки сигналов

Как уже отмечалось, помимо гармонических (узкополосных) и случайных (широкополосных) помех, на практике приходится подавлять и помехи иного рода. Поэтому в состав инструментальной системы включена вычислительная компонента, которая на основе производного класса МОД реализует каскадную процедуру предварительной обработки сигналов.

При этом пользователь имеет возможность включать или отключать любой из этапов обработки (рис. 4.25), а также разрешать или запрещать визуализацию результата отдельной стадии.

Рис. 4.25. Каскадная обработка сигнала



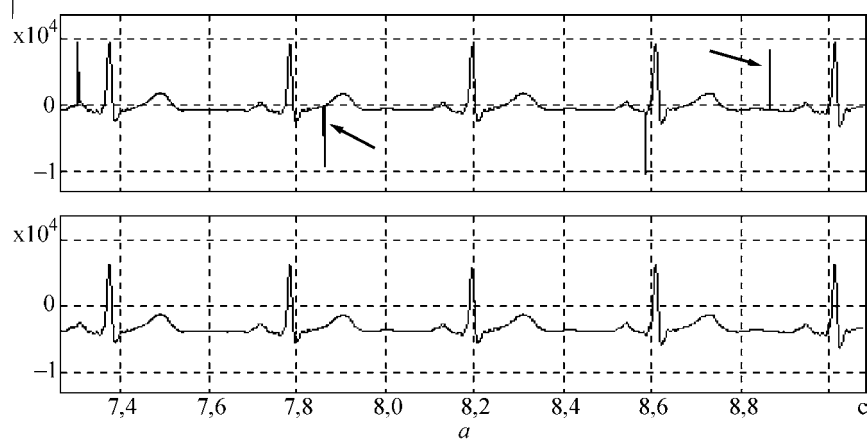
Поскольку процедуры частотной фильтрации и адаптивного сглаживания, реализованные в каскадном фильтре, уже были рассмотрены выше, остановимся кратко на вычислительных процедурах, реализующих первые три стадии комплексной обработки, и приведем результаты их тестирования при обработке реальных сигналов (рис. 4.26).

Для подавления импульсной помехи (рис. 4.26, а) применен известный прием ранговой фильтрации сигнала, реализованный такой процедурой: фильтрованное значение сигнала $\hat{y}[k]$ в k -й момент времени определяется как центральное значение (медиана) упорядоченного массива $y[k-m], \dots, y[k], \dots, y[k+m]$ в скользящем окне с нечетным числом точек. Для подавления одиночных выбросов достаточно последовательно просматривать по три точки сигнала, положив $m=1$. Тогда, после упорядочивания массива $y[k-1], y[k], y[k+1]$, выброс «переместится» в первую или третью точку.

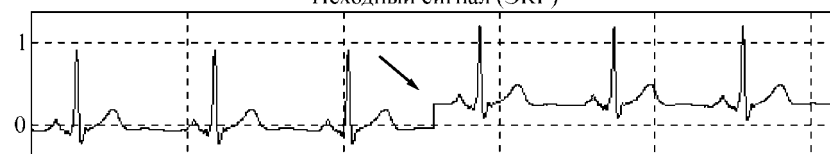
В разработанной вычислительной компоненте предусмотрена настройка параметра m . В этом случае обеспечивается возможность устранения m -точечных выбросов. Однако численные эксперименты показали, что при $m > 5$ применение процедуры становится неэффективным из-за искажений фронтов полезного сигнала.

Удаление резких перепадов (скачков) базового уровня сигнала реализовано следующим образом. Строится профиль изменения среднего значения сигнала, вычисленного «левосторонним» и «правосторонним» окнами:

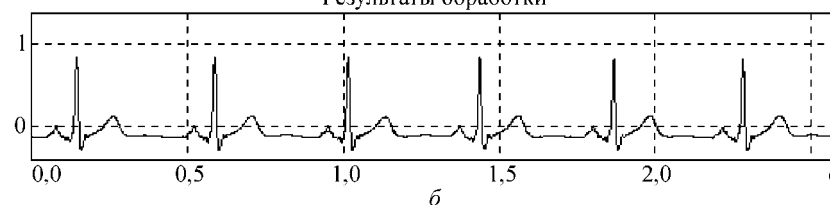
$$\chi[k] = 1/W_0 \left(\sum_{j=1}^{W_0} y[k-j] - \sum_{j=1}^{W_0} y[k+j] \right), \quad (4.28)$$



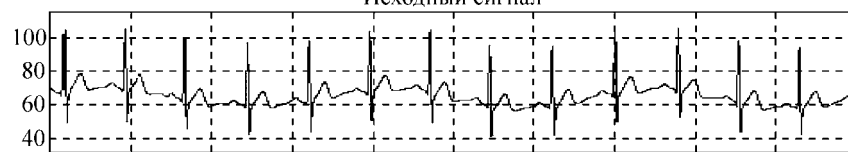
Исходный сигнал (ЭКГ)



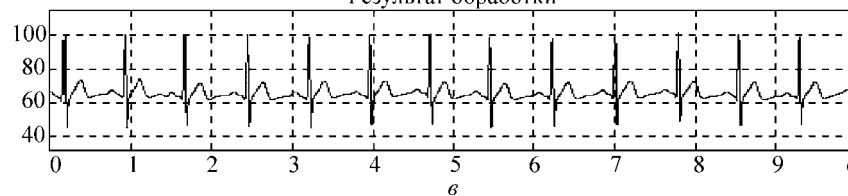
Результаты обработки



Исходный сигнал



Результат обработки



где $W_0 = \text{const}$ — ширина окна. Затем определяется подмножество точек: $\Omega_\chi = \{k \in [1, K] : \chi[k] > \chi_0\}$, где $\chi_0 = \text{const}$ — заданный порог. Если оказывается, что $\Omega_\chi \neq \emptyset$, то вычисляется значение точки $k^* = \arg \max_{k \in \Omega_\chi} \chi[k]$, в которой наблюдается резкий «скачок» сигнала, и вводится поправка:

$$\tilde{y}[k] = \begin{cases} y[k] & \text{при } k < k^*, \\ y[k] + \chi[k^*] & \text{при } k \geq k^*. \end{cases} \quad (4.29)$$

Понятно, что параметры W_0, χ_0 должны подбираться опытным путем для конкретных сигналов. Эксперименты показали, что при правильном выборе этих параметров описанная процедура эффективна для решения прикладных задач (рис. 4.26, б).

Для устранения дрейфа изолинии реальных ЭКГ обычно применяют фильтр верхних частот с частотой среза $f_0 = 0,05$ — $0,5$ Гц. Такой фильтр легко может быть реализован на основе прямого и обратного ДПФ. Однако эксперименты показали, что для решения этой задачи более эффективно использовать иной подход.

Вначале оценим дрейф изолинии, обработав исходный массив $y[1], \dots, y[K]$ следующим образом:

$$y_B[k] = \frac{1}{2W_B + 1} \sum_{j=-W_B}^{W_B} y[k-j], \quad k \in [W_B, K - W_B],$$

где $W_B = \text{const}$ — настраиваемый параметр.

Затем введем соответствующую коррекцию в значения массива, что можно сделать, например, так:

$$\tilde{y}[k] = y[k] - y_B[k] + \Delta_y, \quad (4.30)$$

где

$$\Delta_y = \frac{\max_{W_B \leq k \leq K - W_B} y[k] - \min_{W_B \leq k \leq K - W_B} y[k]}{2}. \quad (4.31)$$

◀ **Рис. 4.26.** Результаты экспериментов:

a — удаление одиночных импульсов; *б* — резкого скачка; *в* — дрейфа изолинии

Очевидно, что и в данном случае эффективность процедуры зависит от правильности выбора параметра W_B , который должен обеспечить усреднение точек в окне, ширина которого заведомо превышает продолжительность отдельных циклов обрабатываемого сигнала. При решении практических задач настройка параметра W_B проводится экспериментально с использованием графических средств инструментальной системы СИДИГРАФ.

Один из результатов тестирования процедуры (4.30) при обработке реальной ЭКГ показан на рис. 4.26, в.

5.1. Критерии эффективности диагностических тестов

5.1.1. Постановка задачи

Как уже отмечалось во второй главе, важным этапом синтеза ИТ обработки сигналов является выбор совокупности полезных признаков, используемых для построения диагностических правил (тестов). При этом часто приходится исследовать новые свойства сигналов, которые предположительно несут дополнительную диагностическую информацию о состоянии объекта.

Например, наши многолетние наблюдения за фазовыми траекториями электрокардиограмм различных пациентов породили гипотезу о возможности расширения традиционного пространства диагностических признаков ЭКГ [180]. Было обнаружено, что после физических нагрузок у нетренированных людей увеличивается разброс фазовых траекторий относительно усредненной траектории, а также наблюдается характерное смещение вправо фрагмента фазовой траектории, соответствующего периоду реполяризации, в то время как у спортсменов такие отклонения отсутствуют.

Естественно, появилось желание разработать компьютерные алгоритмы для автоматического анализа таких признаков и использовать эти измерения в новых ИТ обработки ЭКГ. Однако в соответствии с современными представлениями о доказательной медицине [52] введение новых диагностичес-

ких тестов требует обоснований на клиническом материале с использованием формальных критериев.

Поэтому в настоящей главе рассмотрим результаты наших исследований [206, 208], направленных на усовершенствование методов формальной оценки эффективности новых диагностических тестов. Рассмотрим следующую постановку этой задачи.

Пусть объект Z находится в одном из двух возможных состояний V_1, V_2 , например «болен» или «здоров». Есть алгоритм (диагностический тест), который на основании измерения совокупности признаков $x = (x_1, \dots, x_N)$, $N \geq 1$, принимает решения в виде индикаторной переменной

$$\delta(x) = \begin{cases} 1, & \text{если принято решение в пользу } V_1; \\ 2, & \text{если принято решение в пользу } V_2. \end{cases} \quad (5.1)$$

Предполагается, что тест может ошибаться, т. е. допускает ложноотрицательные и ложноположительные ошибки. Такие ошибки, которые в теории распознавания образов принято называть ошибками пропуска цели и ложной тревоги, будем считать случайными событиями

$$\begin{aligned} E_1 &= (Z \in V_1) \wedge (\delta = 2), \\ E_2 &= (Z \in V_2) \wedge (\delta = 1). \end{aligned}$$

Предположим, что по некоторой представительной выборке наблюдений, для которых известны истинные состояния Z , получены оценки вероятностей $P(E_1)$ и $P(E_2)$. Будем, как это принято в теории статистических решений, характеризовать возможные потери платежной матрицей вида

$$L = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}, \quad (5.2)$$

в которой L_{11}, L_{22} — потери, связанные с правильными решениями, а L_{12}, L_{21} — потери от ошибок пропуска цели и ложной тревоги соответственно.

Тогда математическое ожидание потерь (средний риск) определяет взвешенная сумма указанных потерь с учетом вероятностей их появления:

$$R = \sum_{k=1}^2 \sum_{j=1}^2 L_{kj} P(V_k, \delta = j), \quad (5.3)$$

где величина $P(V_k, \delta = j)$ обозначает вероятность совместного выполнения двух случайных событий: Z находится в состоянии V_k , $k = 1, 2$, а алгоритм принял решение $\delta = j$, отнеся его к классу V_j , $j = 1, 2$.

Введем следующее определение.

Определение 5.1. Диагностический тест полезен, если средний риск R принимаемых на его основе решений меньше риска R_0 априорных решений, принимаемых по вероятностям $P(V_1)$ и $P(V_2)$, т. е. выполняется строгое неравенство

$$R < R_0. \quad (5.4)$$

Ставится задача — получить условия, гарантирующие выполнение неравенства (5.4).

5.1.2. Достаточные условия

полезности диагностического теста

Введем обозначение

$$\omega = \frac{L_{12} - L_{11}}{L_{21} - L_{22}}, \quad (5.5)$$

полагая, что $L_{12} > L_{11}$ и $L_{21} > L_{22}$, а значит, $\omega > 0$. Такое предположение вполне логично, поскольку потери от ошибочных результатов всегда превышают потери от правильных решений. Отметим, что в частном случае, когда $L_{11} = L_{22} = 0$, величина ω просто определяет соотношение потерь от ошибок пропуска цели и ложной тревоги. Однако в дальнейших исследованиях такое дополнительное предположение не обязательно.

Докажем справедливость такого утверждения.

Утверждение 5.1. Диагностический тест полезен в смысле строгого неравенства (5.4) в том и только в том случае, когда

$$[1 - P(V_1)]P(E_2) < \omega P(V_1)[1 - P(E_1)] \quad (5.6)$$

при $P(V_1)(1 + \omega) < 1$ либо

$$[1 - P(V_1)][1 - P(E_2)] > \omega P(V_1)P(E_1) \quad (5.7)$$

при $P(V_1)(1 + \omega) > 1$.

Доказательство. Запишем совместные вероятности $P(V_k, \delta = j)$, фигурирующие в (5.3), в эквивалентной форме

$$P(V_k, \delta = j) = P(V_k)P(\delta = j / V_k), \quad \delta, j, k \in \{1, 2\}.$$

Поскольку по определению

$$P(\delta = 2 / V_1) = P(E_1) \quad \text{и} \quad P(\delta = 1 / V_2) = P(E_2),$$

причем

$$P(\delta = 1 / V_1) = 1 - P(E_1) \quad \text{и} \quad P(\delta = 2 / V_2) = 1 - P(E_2),$$

то, согласно (5.3), средний риск решений (5.1) можно представить в виде

$$\begin{aligned} R = & L_{11}P(V_1)[1 - P(E_1)] + L_{22}[1 - P(V_1)][1 - P(E_2)] + \\ & + L_{12}P(V_1)P(E_1) + L_{21}[1 - P(V_1)]P(E_2). \end{aligned} \quad (5.8)$$

Если же не использовать диагностический тест, то априорная «стратегия» решений о неизвестных состояниях Z сводится к выбору одного из двух вариантов. Либо всегда считать, что $Z \in V_1$ и тогда риск будет равен $L_{11}P(V_1) + L_{21}[1 - P(V_1)]$, либо всегда считать, что $Z \in V_2$ и тогда риск будет равен $L_{22}[1 - P(V_1)] + L_{12}P(V_1)$.

Принимая во внимание обозначение (5.5), заключаем, что если выполняется условие

$$P(V_1)(1 + \omega) < 1, \quad (5.9)$$

то для обеспечения минимума априорного риска всегда следует принимать решение $Z \in V_2$, а если выполняется условие

$$P(V_1)(1 + \omega) > 1, \quad (5.10)$$

то лучшим оказывается решение $Z \in V_1$.

Следовательно, априорный риск R_0 , фигурирующий в правой части (5.4), можно представить в виде

$$R_0 = \begin{cases} L_{22}[1 - P(V_1)] + L_{12}P(V_1), & \text{если } P(V_1)(1 + \omega) < 1 \\ L_{11}P(V_1) + L_{21}[1 - P(V_1)], & \text{если } P(V_1)(1 + \omega) > 1 \end{cases}. \quad (5.11)$$

Рис. 5.1. Области полезности теста:

a — при $P(V_1)(1 + \omega) > 1$; b — при $P(V_1)(1 + \omega) < 1$

Подстановка (5.8) и (5.11) в (5.4) после простых преобразований приводит к условиям (5.6), (5.7). Утверждение доказано.

На рис. 5.1 показаны границы областей полезного теста, определяемых соотношениями (5.6) и (5.7) соответственно.

Из утверждения 5.1 следует, что если не выполняется условие (5.6) или (5.7), то с точки зрения уменьшения априорного риска диагностический тест абсолютно бесполезен. Этот факт позволяет сформулировать еще два не менее важных утверждения.

Утверждение 5.2. Пусть выполняется условие $P(V_1)(1 + \omega) < 1$. Тогда при сколь угодно малой вероятности ошибки пропуска цели диагностический тест заведомо бесполезен, если вероятность ошибки ложной тревоги удовлетворяет условию

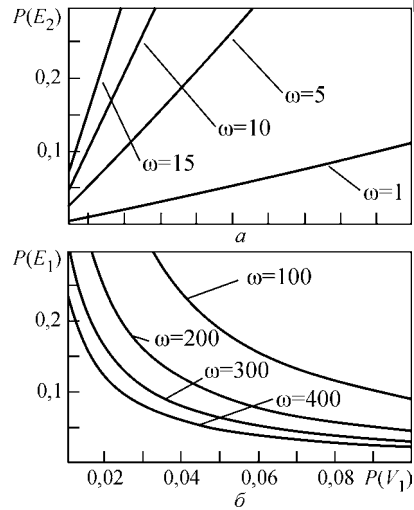
$$P(E_2) \geq \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (5.12)$$

Доказательство. Принимая во внимание, что $P(V_1) \neq 1$, запишем условие (5.6), гарантирующее полезность теста в области $P(V_1)(1 + \omega) < 1$, следующим образом:

$$P(E_2) < \frac{\omega P(V_1)[1 - P(E_1)]}{1 - P(V_1)}. \quad (5.13)$$

Поскольку всегда $P(E_1) \geq 0$, то усиление (5.13) путем подстановки $P(E_1) = 0$ приводит к оценке сверху допустимого значения вероятности ошибки ложной тревоги полезного теста:

$$\sup_{0 \leq P(E_1) \leq 1} P(E_2) = \frac{\omega P(V_1)}{1 - P(V_1)} \quad (5.14)$$



для области $P(V_1)(1 + \omega) < 1$. Отсюда непосредственно следует справедливость утверждения.

Утверждение 5.3. Пусть выполняется условие $P(V_1)(1 + \omega) > 1$. Тогда при любой сколь угодно малой вероятности ошибки ложной тревоги диагностический тест заведомо бесполезен с точки зрения уменьшения априорного риска, если вероятность ошибки пропуска цели удовлетворяет условию

$$P(E_1) \geq \frac{1 - P(V_1)}{\omega P(V_1)}. \quad (5.15)$$

Доказательство. Поскольку $P(V_1) \neq 0$, то условие (5.7), гарантирующее полезность теста для области $P(V_1)(1 + \omega) > 1$, можно представить в эквивалентной форме:

$$P(E_1) < \frac{[1 - P(V_1)][1 - P(E_2)]}{\omega P(V_1)}. \quad (5.16)$$

Если учесть, что всегда $P(E_2) \geq 0$, то усиление (5.16) путем подстановки $P(E_2) = 0$ приводит к оценке сверху вероятности ошибки пропуска цели полезного теста

$$\sup_{0 \leq P(E_2) \leq 1} P(E_1) = \frac{1 - P(V_1)}{\omega P(V_1)} \quad (5.17)$$

для области $P(V_1)(1 + \omega) > 1$. Утверждение доказано.

5.1.3. Условия полезности совокупности диагностических признаков

Любой алгоритм диагностики основан на измерении некоторой совокупности $N \geq 1$ признаков $x_1, \dots, x_N$, образующих N -мерное описание $x^{(N)} = (x_1, \dots, x_N)$ распознаваемых классов V_1 и V_2 . В простейшем случае для диагностики используется только один признак ($N = 1$). Тогда, если известны оценки вероятностей ошибок пропуска цели и ложной тревоги, допускаемых этим признаком, и оценка априорной вероятности $P(V_1)$, то при заданном значении величины ω полезность отдельного признака можно оценить на основе полученных выше условий.

Для оценки полезности признаков в совокупности $N > 1$ признаков введем следующее определение.

Определение 5.2. Диагностический признак $x_n (1 \leq n \leq N)$ полезен в совокупности с остальными $N - 1$ признаками, если его исключение из описания $x^{(N)} = (x_1, \dots, x_N)$ приводит к увеличению среднего риска, т. е.

$$R_{N-1} > R_N, \quad (5.18)$$

где R_N и R_{N-1} — средний риск диагностики по описанию $x^{(N)}$ и сокращенному описанию $x^{(N-1)}$, которое не содержит x_n .

Рассмотрим условия, гарантирующие выполнение строгого неравенства (5.18), для бинарных признаков (симптомов), характерных для задач медицинской диагностики.

Пусть для диагностики некоторого заболевания используется симптом x_1 , который может отсутствовать (случайное событие x_1^-) или присутствовать (случайное событие x_1^+) у пациента Z . Предположим, что симптом x_1 характерен для всех без исключения больных (класс V_1) и только для части здоровых (класс V_2). Формально это означает, что условные вероятности $P(x_1^+ / V_1)$, $P(x_1^+ / V_2)$ удовлетворяют соотношениям

$$P(x_1^+ / V_1) = 1, \quad 0 < P(x_1^+ / V_2) < 1. \quad (5.19)$$

Из (5.19) следует, что $P(x_1^- / V_1) = 0$. Значит, при отсутствии симптома ($x_1 = x_1^-$) однозначно нужно принимать решение в пользу V_2 . Ясно, что в этом случае вероятность ошибочного отнесения больного к группе здоровых (вероятность ошибки «пропуска цели») равна нулю, т. е. $P(E_1) = 0$.

Предположим теперь, что вероятность $P(x_1^+ / V_2)$ удовлетворяет условию

$$P(x_1^+ / V_2) \geq \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (5.20)$$

Заметим, что для фиксированных значений $P(V_1)$ и ω , при которых выполняется условие $P(V_1)(1 + \omega) < 1$, неравенство (5.20) не противоречит второму из ограничений (5.19).

Если принимать решение в пользу класса V_1 , когда $x_1 = x_1^+$, т. е. считать пациента больным при наличии симптома, то величина $P(x_1^+ / V_2)$ определит вероятность ошибки ложной тревоги, т. е. $P(E_2) = P(x_1^+ / V_2)$. Но тогда при выполнении условия (5.20) в соответствии с утверждением 5.2 признак x_1 заведомо бесполезен с точки зрения уменьшения априорного риска.

Для того чтобы еще раз убедиться в этом, сравним условные риски возможных решений, которые принимаются при $x_1 = x_1^+$.

Риск решения $Z \in V_1$ равен

$$R_1 = L_{11}P(V_1)P(x_1^+ / V_1) + L_{21}[1 - P(V_1)]P(x_1^+ / V_2),$$

а риск решения $Z \in V_2$ —

$$R_2 = L_{22}[1 - P(V_1)]P(x_1^+ / V_2) + L_{12}P(V_1)P(x_1^+ / V_1).$$

Если учесть, что $P(x_1^+ / V_1) = 1$, то с учетом обозначения (5.5) разность $\Delta R = R_2 - R_1$ определяется соотношением

$$\Delta R = (L_{21} - L_{22})[\omega P(V_1) - [1 - P(V_1)]P(x_1^+ / V_2)], \quad (5.21)$$

из которого в силу (5.20) и того, что $L_{21} > L_{22}$, следует $R_2 \leq R_1$.

Отсюда можно сделать вывод, что и при наличии симптома, т. е. при $x_1 = x_1^+$, необходимо принимать решение $Z \in V_2$.

Таким образом, при любом возможном значении признака $x_1 = x_1^-$ и $x_1 = x_1^+$ принимается одно и то же решение в пользу класса V_2 , которое для фиксированных значений $P(V_1)$ и ω в области $P(V_1)(1 + \omega) < 1$ совпадает с априорным решением. Естественно, что такой признак-симптом сам по себе абсолютно бесполезен.

Предположим теперь, что в нашем распоряжении есть два таких «бесполезных» в отдельности признака. Возникает естественный вопрос: в каких случаях их совокупность может оказаться полезной с точки зрения определения 5.3.

В работе [277] нами было показано, что совокупность неинформативных в отдельности признаков может быть полезна, если они статистически связаны в классах.

Покажем теперь, что совокупность бесполезных в отдельности признаков-симптомов может быть полезна и в том случае, когда такая связь отсутствует. Этот факт в точных формулировках определяет следующее утверждение.

Утверждение 5.4. Пусть x_1, x_2 — два признака, каждый из которых принимает два возможных значения x_1^-, x_1^+ и x_2^-, x_2^+ , причем

$$P(x_i^+ / V_1) = 1, \quad i = 1, 2. \quad (5.22)$$

Пусть, кроме того, признаки статистически независимы в обоих классах, т. е.

$$P(x_1, x_2 / V_k) \equiv P(x_1 / V_k)P(x_2 / V_k), \quad k = 1, 2. \quad (5.23)$$

Тогда для любых $P(V_1)$ и ω , удовлетворяющих условию $P(V_1)(1 + \omega) < 1$, существуют такие значения вероятностей $P(x_i^+ / V_2)$, $i = 1, 2$, при которых каждый из этих признаков бесполезен сам по себе, но их совокупность полезна.

Доказательство. Предположим, что

$$P(x_i^+ / V_2) \geq \rho, \quad i = 1, 2, \quad (5.24)$$

где

$$\rho = \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (5.25)$$

Как было показано ранее, в этом случае рассматриваемые признаки сами по себе бесполезны. Рассмотрим условные риски возможных решений при комбинации значений признаков $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$.

Принимая во внимание (5.23), заключаем, что при такой комбинации значений симптомов риск решения $Z \in V_1$ определяется выражением

$$\begin{aligned} R_1(x_1^+, x_2^+) &= L_{11}P(V_1)P(x_1^+ / V_1)P(x_2^+ / V_1) + \\ &+ L_{21}[1 - P(V_1)]P(x_1^+ / V_2)P(x_2^+ / V_2), \end{aligned} \quad (5.26)$$

а риск решения $Z \in V_2$ —

$$R_2(x_1^+, x_2^+) = L_{22}[1 - P(V_1)]P(x_1^+ / V_2)P(x_2^+ / V_2) +$$

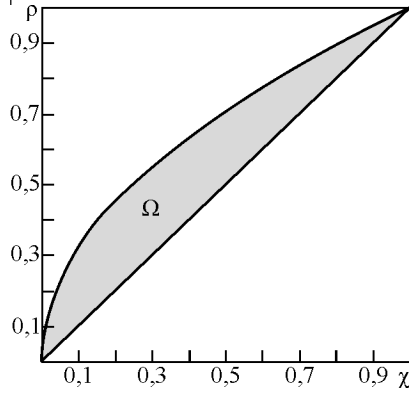


Рис. 5.2. Область полезной совокупности признаков

$$+ L_{12}P(V_1)P(x_1^+ / V_1)P(x_2^+ / V_1). \quad (5.27)$$

При выполнении условия $P(V_1)(1 + \omega) < 1$ для минимизации априорного риска всегда следует принимать решение $Z \in V_2$. Потребуем, чтобы при комбинации значений признаков $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$ оптимальным было противоположное решение, т. е. $Z \in V_1$. Иными словами, потребуем, чтобы выполнялось строгое неравенство

$$R_1(x_1^+, x_2^+) < R_2(x_1^+, x_2^+).$$

На основании (5.26), (5.27) с учетом обозначения (5.25) последнее неравенство можно представить в виде

$$P(x_1^+ / V_2)P(x_2^+ / V_2) < \rho. \quad (5.28)$$

Однако в соответствии с (5.24)

$$P(x_1^+ / V_2)P(x_2^+ / V_2) \geq \rho^2. \quad (5.29)$$

Поскольку при выполнении условий $P(V_1) \neq 0$ и $P(V_1)(1 + \omega) < 1$ величина ρ лежит в пределах $0 < \rho < 1$, то всегда найдутся такие значения вероятностей $P(x_1^+ / V_2)$ и $P(x_2^+ / V_2)$, удовлетворяющие (5.24), при которых неравенства (5.28), (5.29) выполняются совместно. Утверждение доказано.

На рис. 5.2 в координатах ρ и $\chi = P(x_1^+ / V_2)P(x_2^+ / V_2)$ показана область Ω значений, удовлетворяющих (5.28), (5.29).

5.1.4. Модельный пример

Пусть $P(V_1) = 0,04$, $L_{11} = L_{22} = 0$, $L_{12} = 4$, $L_{21} = 1$. Следовательно, $\omega = 4$, а $\rho \approx 0,167$. Предположим, что есть два бинарных признака $x_1 = x_1^+, x_1^-$ и $x_2 = x_2^+, x_2^-$, которые статистически незави-

ТАБЛИЦА 5.1

**Условные распределения признаков
в первом классе**

x_1	x_2		$P(x_1/V_1)$
	$x_2 = x_2^-$	$x_2 = x_2^+$	
$x_1 = x_1^-$	0	0	0
$x_1 = x_1^+$	0	1	1
$P(x_2/V_1)$	0	1	—

ТАБЛИЦА 5.2

**Условные распределения признаков
во втором классе**

x_1	x_2		$P(x_1/V_2)$
	$x_2 = x_2^-$	$x_2 = x_2^+$	
$x_1 = x_1^-$	0,49	0,21	0,7
$x_1 = x_1^+$	0,21	0,09	0,3
$P(x_2/V_2)$	0,7	0,3	—

симы в обоих классах. Условные распределения признаков представлены в табл. 5.1 и 5.2.

Видно, что для каждого из признаков выполняется условие (5.24), а значит, эти признаки сами по себе бесполезны. В этом легко убедиться, если сравнить условные риски возможных решений при различных значениях этих признаков.

Например, принятие решения $Z \in V_1$ по признаку x_1 приводит к риску $R_1(x_1^+) = 0,288$ при наличии симптома ($x_1 = x_1^+$) и риску $R_1(x_1^-) = 0,672$ при отсутствии симптома ($x_1 = x_1^-$), в то время как решение $Z \in V_2$ приводит к риску $R_2(x_1^+) = 0,16$ при $x_1 = x_1^+$ и риску $R_2(x_1^-) = 0$ при $x_1 = x_1^-$.

Следовательно, и при наличии, и при отсутствии симптома x_1 с точки зрения минимума среднего риска должны приниматься решения в пользу класса V_2 , которые при заданных значениях $P(V_1) = 0,04$ и $\omega = 4$ совпадают с априорными решениями. Аналогичная ситуация и для признака x_2 .

Однако из условия данного примера следует, что $\chi = P(x_1^+/V_2)P(x_2^+/V_2) = 0,09$. Поскольку $\rho \approx 0,167$, то выполняется условие $\chi < \rho \leq \sqrt{\chi}$. А это, как следует из приведенного выше доказательства, свидетельствует о том, что совокупность признаков полезна.

В самом деле, при обнаружении обоих симптомов, когда $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$, решение $Z \in V_1$ приводит к риску $R_1(x_1^+, x_2^+) = 0,0864$, тогда как решение $Z \in V_2$ приводит к большему

риску $R_2(x_1^+, x_2^+) = 0,16$. Следовательно, при такой комбинации признаков оптимальное решение не совпадает с априорным, что и обеспечивает уменьшение среднего риска.

5.1.5. Частный случай

Пусть $L_{11} = L_{22} = 0$, а $L_{12} = L_{21} = 1$. Известно, что в этом случае средний риск (5.3) равен средней вероятности ошибки, а оптимальные решения, обеспечивающие минимум этой вероятности, должны приниматься по правилу максимума апостериорных вероятностей.

Если в дополнение к условиям утверждения 5.4 допустить, что вероятности наличия каждого из симптомов для объектов второго класса одинаковы, т. е.

$$P(x_1^+ / V_2) = P(x_2^+ / V_2) = P_2^+,$$

то, как следует из доказательства, при выполнении условия

$$\frac{P(V_1)}{1 - P(V_1)} \leq P_2^+ < \sqrt{\frac{P(V_1)}{1 - P(V_1)}} \quad (5.30)$$

такие признаки будут бесполезны сами по себе, но полезны в совокупности.

На рис. 5.3, а в координатах $P_2^+ - P(V_1)$ показана область Ω значений, удовлетворяющих условию (5.30). Например, если $P(V_1) = 0,1$ и симптом $x = x_1$ или симптом $x = x_2$ присутствует для всех объектов первого класса ($P_1^+ = 1$) и для 20 % объектов второго класса ($P_2^+ = 0,2$), то точка с такими координатами попадает в область Ω .

Результаты проверки по формуле Байеса убеждают в том, что $P(V_1 / x^+) \approx 0,357$, в то время как $P(V_2 / x^+) \approx 0,642$. Вместе с тем $P(V_1 / x^-) = 0$, а $P(V_2 / x^-) = 1$, т. е.

$$P(V_2 / x) > P(V_1 / x) \quad \forall x = x^+, x^-.$$

Следовательно, при любом значении отдельно взятого симптома оптимальным решением по правилу максимума апостериорных вероятностей $P(V_1 / x)$, $P(V_2 / x)$ оказывается решение

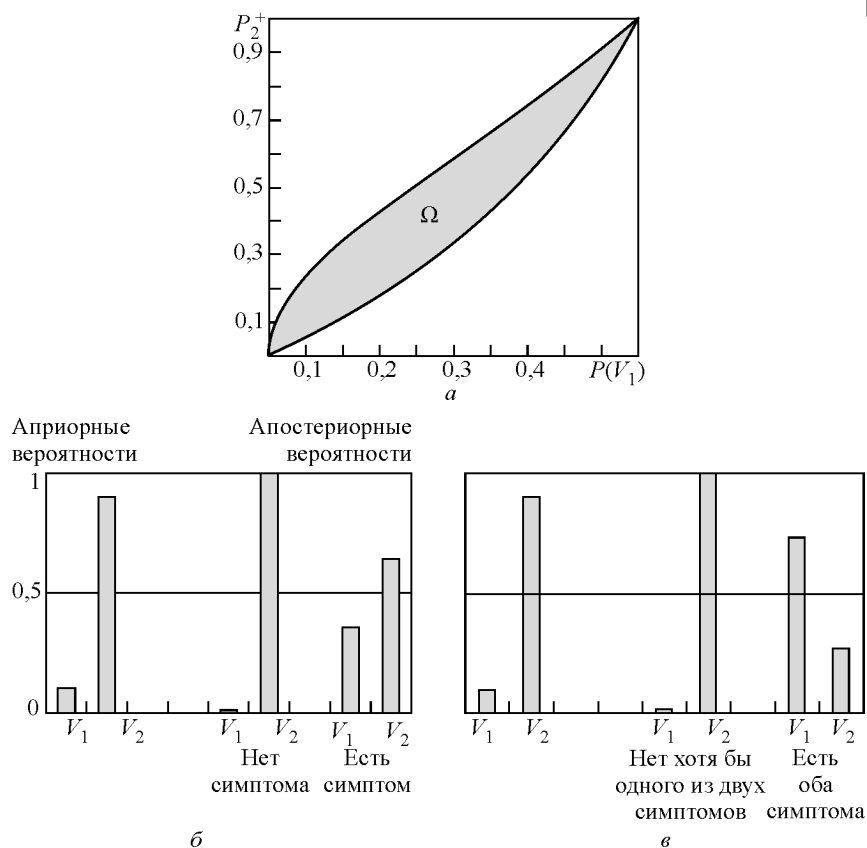


Рис. 5.3. Полезная совокупность «бесполезных» симптомов:

область Ω (а); апостериорные вероятности классов при анализе отдельных признаков (б) и их совокупности (в)

$Z \in V_2$, совпадающее с решением принимаемых по максимуму априорных вероятностей $P(V_1)$ и $P(V_2) = 1 - P(V_1)$ распознаваемых классов (рис. 5.3, б). Значит, с точки зрения средней вероятности ошибки такой признак-симптом абсолютно бесполезен.

В то же время, если два подобных симптома статистически независимы в классах, то их совместное использование полезно. При обнаружении обоих симптомов оптимальным оказывается решение в пользу класса V_1 , поскольку $P(V_1 / x_1^+, x_2^+) \approx 0,74$, а $P(V_2 / x_1^+, x_2^+) \approx 0,26$ (рис. 5.3, в).

5.2. Оценка полезности диагностических тестов в задачах скрининга заболеваний

5.2.1. Общая схема

Переосмыслим теперь полученные результаты применительно к задаче оценки эффективности диагностических тестов, ориентированных на задачи скрининга — массовое профилактическое обследование населения с целью выявления того или иного заболевания.

Рассмотрим задачу выявления больных людей в группе, где есть и больные (класс V_1), и здоровые (класс V_2). При этом, естественно, подразумевается, что V_2 включает в себя не абсолютно здоровых людей, а лиц, у которых нет рассматриваемого заболевания.

Пусть для выявления больных предполагается использовать некоторый диагностический тест, в соответствии с которым для конкретного пациента Z из обследуемой группы принимаются одно из двух решений: положительное решение $\delta = 1$ («Болен») или отрицательное решение $\delta = 2$ («Здоров»).

В соответствии с определением 5.1 будем оценивать полезность диагностического теста по схеме, показанной на рис. 5.4.

Предположим, что проведен экзамен диагностического теста по некоторой репрезентативной группе пациентов с заранее известными диагнозами. Результаты такого экзамена можно представить в виде 2×2 матрицы (табл. 5.3), в которой приняты следующие обозначения:

- TP (*True Positive*) — число правильных диагнозов «Болен»;
- TN (*True Negative*) — число правильных диагнозов «Здоров»;
- FP (*False Positive*) — число здоровых, ошибочно признанных больными;

ТАБЛИЦА 5.3

Результаты тестирования репрезентативной группы

Истинный диагноз	Результат тестирования	
	Решение «Болен» ($\delta = 1$)	Решение «Здоров» ($\delta = 2$)
Класс V_1 — «Болен»	TP	FN
Класс V_2 — «Здоров»	FP	TN

• FN (*False Negative*) — число больных, ошибочно признанных здоровыми.

По данным табл. 5.3 легко определить принятые в медицинской диагностике операционные характеристики теста — чувствительность и специфичность [53].

Чувствительность (sensitivity) S_E определяется долей (процентом) предварительно верифицированных больных, которые тест признал больными:

$$S_E = \frac{TP}{TP + FN}, \quad (5.31)$$

а *специфичность* (specificity) S_P определяется долей (процентом) здоровых, которые тест отнес к группе здоровых:

$$S_P = \frac{TN}{TN + FP}. \quad (5.32)$$

Очевидно, что если выборка репрезентативна, то S_E является оценкой величины $1 - P(E_1)$, а S_P — оценкой величины $1 - P(E_2)$, где $P(E_1)$, $P(E_2)$ — вероятности ошибок пропуска цели и ложной тревоги.

Из утверждения 5.1 следует, что для оценки полезности теста по описанной выше схеме достаточно проверить выполнение одного из условий

$$S_E > \theta(1 - S_P), \quad \text{если } \theta \geq 1; \quad (5.33)$$

$$S_E > 1 - \theta + \theta(1 - S_P), \quad \text{если } \theta < 1, \quad (5.34)$$

где параметр θ определяется по формуле

$$\theta = \frac{1 - P}{\omega P},$$

ω — соотношение потерь (5.5), а $P \neq 0$ — частота (преваленс) заболевания.

Поскольку численные примеры часто бывают более убедительны, чем формальные рассуждения, рассмотрим модельный пример.

Пусть планируется провести скрининг заболевания, встречаемость (преваленс) которого составляет 2 %, т. е. $P = 0,02$. Для диагностики предлагается использовать тест, имеющий чувстви-

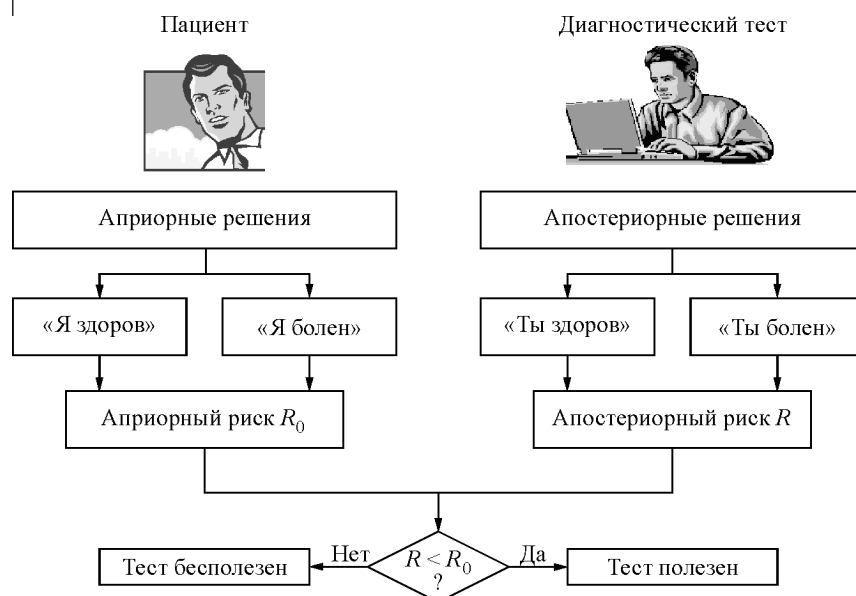


Рис. 5.4. Общая схема оценки полезности диагностического теста



Рис. 5.5. Ожидаемые результаты скрининга

тельность $S_E = 0,9$ и специфичность $S_p = 0,9$. Отметим, что такие операционные характеристики обычно считаются достаточно высокими для медицинских диагностических тестов.

Предположим, что потери L_{11}, L_{22} от правильных решений одинаковы и равны нулю, а потери от ложноотрицательных и ложноположительных результатов составляют соответственно $L_{12} = 5$ и $L_{21} = 1$, т. е. $\omega = 5$.

Поскольку при таких исходных данных $\theta > 1$, то проверке подлежит условие полезности (12). Легко убедиться в том, что это условие не выполняется, а значит, тест абсолютно бесполезен для скрининга.

На рис. 5.5 представлены ожидаемые результаты тестирования группы в 10 000 человек.

Легко заметить, что априорные потери, связанные с тем, что 200 больных будут не выявлены, составляет $5 \cdot 200 = 1000$, т. е. $R_0 = 0,1$. В то же время суммарные потери, связанные с ложноположительными результатами диагностики для 980 здоровых пациентов и ложноотрицательными результатами для 20 человек, составят $1 \cdot 980 + 5 \cdot 20 = 1080$, т. е. $R = 0,108$.

Следовательно, несмотря на высокие показатели чувствительности и специфичности, скрининг на основе такого теста только увеличит риск.

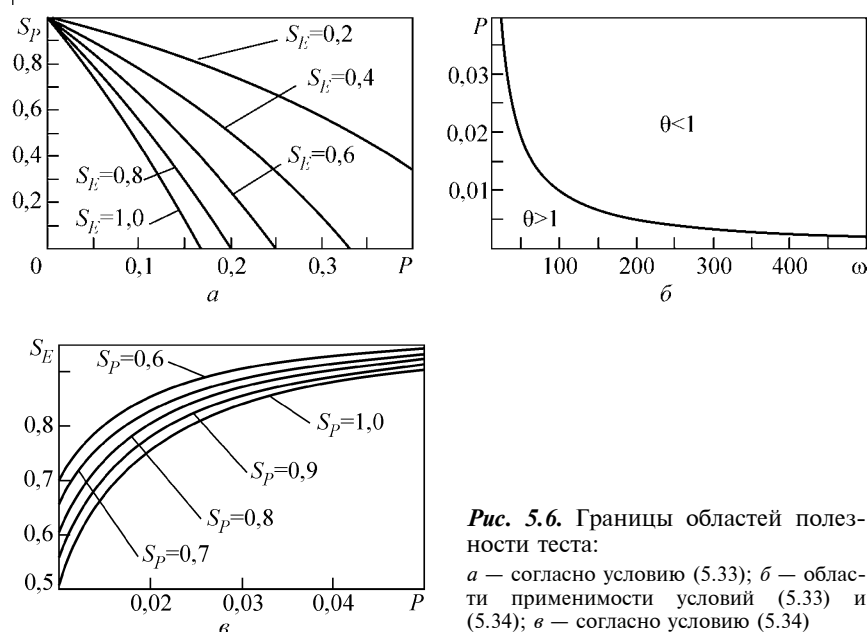
Если изменить условия примера, положив $L_{12} = L_{21} = 1$, то в результате применения теста увеличение риска оказывается еще более впечатляющим. В этом случае ожидаемый средний риск диагностики составит $R = 0,1$, в то время как априорный риск — $R_0 = 0,02$.

Важная особенность скрининга и профилактических обследований — это низкие значения преваленса больных в обследуемой группе. В частности, согласно [53], преваленс сахарного диабета составляет 2 %, а рака молочной железы в возрасте более 50 лет — 1 % [118].

Если к тому же допустить, что соотношение потерь ω от ложноотрицательных и ложноположительных ошибок не очень велико, то выполняется условие $\theta \geq 1$. В этих предположениях достаточно проверить лишь условие (5.33), чтобы убедиться в полезности теста.

На рис. 5.6, а показаны границы областей полезности теста, построенные, согласно (5.33), при $\omega = 5$ для преваленса $0 < P \leq 0,4$.

Однако, согласно [53, с. 104], при диагностике особо опасных заболеваний соотношение потерь от ложноотрицательных и



ложноположительных результатов иногда достигает $\omega \approx 400$. При таких больших значениях ω условие $\theta \geq 1$ может не выполняться даже при низких значениях преваленса (рис. 5.6, б).

На рис. 5.6, в показаны области полезности теста, построенные в соответствии с условием (5.34), для случая, когда преваленс имеет низкие значения, но $\theta > 1$, поскольку $\omega = 200$.

5.2.2. Допустимые границы специфичности и чувствительности полезного теста

В общем случае между специфичностью S_p и чувствительностью S_E диагностического теста не существует однозначной связи. Известно лишь то, что при изменении порога скрининга повышается чувствительность и снижается специфичность или наоборот.

Такая ситуация характерна при использовании диагностических тестов в виде пороговых правил:

$$\delta = \begin{cases} 1, & \text{если } x > x_0, \\ 2, & \text{если } x < x_0, \end{cases}$$

где x — измеряемый физиологический параметр; x_0 — пороговое значение, которое принято называть уровнем скрининга.

В выборе оптимального значения порога x_0 заключается основная проблема настройки диагностических тестов, используемых при массовых обследованиях населения для выявления того или иного заболевания [53].

Покажем, что полученные результаты позволяют обоснованно подходить к разрешению этой важной проблемы.

Из условия утверждения 5.2 прямо следует, что для значений P и ω , удовлетворяющих условию $P(1 + \omega) < 1$, или, что то же самое, $\theta > 1$, диагностический тест абсолютно бесполезен, если его специфичность удовлетворяет условию

$$S_p \leq S_p^0, \quad (5.35)$$

где

$$S_p^0 = \frac{1 - P(1 + \omega)}{1 - P}. \quad (5.36)$$

Как видно по графикам (рис. 5.7, а), при малых значениях prevalence ($P < 0,02$) специфичность полезного теста должна быть выше 90 % уже при $\omega \leq 5$, а в условиях равенства потерь от ложноположительных и ложноотрицательных результатов — выше 98 %.

Поскольку для реальных тестов $S_E < 1$, причем по мере снижения чувствительности растет число ложноотрицательных результатов, то для обеспечения полезности теста величина специфичности должна быть выше приведенных граничных величин, что еще более сужает область полезности теста.

Заметим, что согласно (5.33), для низкого prevalence даже при $\omega = 1$ снижение специфичности теста оказывается намного более критичным, чем уменьшение чувствительности.

Например, при $P = 0,05$ и $\omega = 1$ допустимая величина специфичности S_p изменяется всего от 95,7 % до 97,9 % при существенном снижении чувствительности S_E от 80 % до 40 % соответственно (рис. 5.7, б).

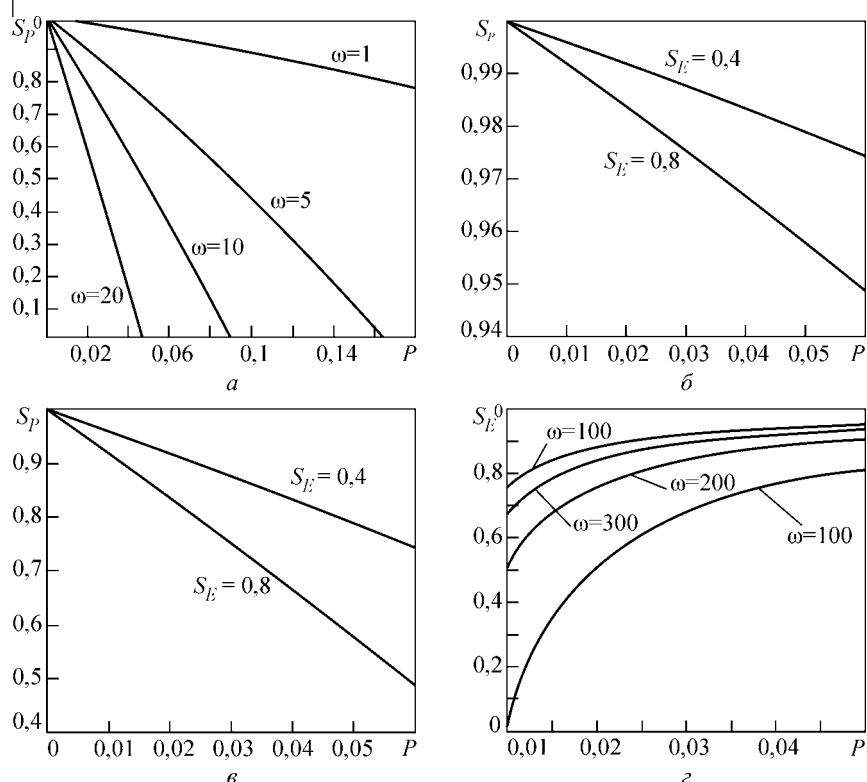


Рис. 5.7. Области полезности теста:

зависимость допустимой специфичности S_p^0 от преваленса P (а); фрагменты областей полезности при $\omega = 1$ (б) и $\omega = 10$ (в); зависимость допустимой чувствительности S_E^0 от преваленса P (г)

Если же $\omega = 10$, то при аналогичном изменении S_E и таком же преваленсе допустимая величина специфичности S_p изменяется уже от 56,5 % до 78,5 % (рис. 5.7, в).

Из условия утверждения 5.3 следует, что для значений P и ω , удовлетворяющих условию $P(1 + \omega) > 1$, или, что то же самое, $\theta < 1$ диагностический тест абсолютно бесполезен, если его чувствительность удовлетворяет условию

$$S_T \leq S_E^0, \quad (5.37)$$

где

$$S_E^0 = 1 - \frac{1 - P}{\omega P}. \quad (5.38)$$

На рис. 5.7, *г* представлены графики, отражающие зависимость допустимой чувствительности S_E^0 от преваленса P при различных значениях ω .

Примечательно, что, в отличие от граничного значения специфичности S_p^0 (см. рис. 5.7, *а*), допустимая граница чувствительности уменьшается с уменьшением P и ω .

В данном случае (при $\theta < 1$) снижение специфичности теста оказывается менее критичным, чем уменьшение чувствительности. Например, при $P = 0,02$ допустимая величина S_E может изменяться всего на 10 % при снижении чувствительности на 40 %.

5.2.3. Взаимосвязь полученных результатов с традиционным ROC-анализом

Для оценки эффективности диагностических тестов в медицинской практике широко используется известный метод, основанный на анализе так называемых характеристических кривых ROC (*Receiver Operating Characteristic curve*) [294, 326, 328, 339]. При этом считается, что традиционный ROC-анализ в определенной мере позволяет разрешить проблему компромисса между чувствительностью S_E и специфичностью S_p диагностического теста.

Эмпирическая ROC-кривая отображает зависимость чувствительности S_E от величины $1 - S_p$. Для пороговых решающих правил такая зависимость строится по дискретным значениям пар S_E и S_p , полученным на экзаменационной выборке при различных значениях порога x_0 (рис. 5.8, *а*).

Если условные распределения признаков в классах различны, то

$$S_E > 1 - S_p \quad (5.39)$$

и точки расположены выше диагонали. Поэтому считается, что

площадь U_{ROC} под ROC -кривой характеризует диагностическую ценность признака.

Величина U_{ROC} представляет собой среднюю чувствительность теста при возможных значениях специфичности и наоборот, причем $U_{ROC} = 1$ для «идеального» теста и $U_{ROC} = 0,5$ для бесполезного теста, операционные характеристики которого не удовлетворяют условию (5.39).

Однако на основании проведенных выше исследований можно заключить, что в общем случае выполнение строгого неравенства (5.39) есть необходимым, но не достаточным условием полезности диагностического теста с точки зрения снижения априорного риска. Поэтому рассмотрим возможность **усиления** традиционного ROC -анализа за счет выделения в ROC -пространстве области, гарантирующей полезность теста.

В соответствии с условием (5.33) для случая $P(1 + \omega) < 1$ такую область ограничивает прямая OA , которая проходит от точки с координатами $(0, 0)$ в точку с координатами $(0, 1/\theta)$ (рис. 5.8, б).

Легко видеть, что если prevalence исследуемого заболевания равен 15 %, а $\omega = 1$, то тест бесполезен, поскольку практически вся ROC -кривая расположена ниже ограничивающей прямой. Однако если допустить, что стоимость ложноотрицательных результатов в четыре раза выше стоимости ложноположительных (при эквивалентности потерь от правильных результатов), т. е. положить $\omega = 4$, то ограничивающая прямая OB уже пересекает ROC -кривую.

Следовательно, диагностический тест с чувствительностью $S_E = 62,5\%$ и специфичностью $S_P = 80\%$ может быть рекомендован для практического применения, поскольку в этом случае соответствующая точка C принадлежит допустимой части ROC -кривой.

Аналогично на основании условия (5.34) можно ограничить ROC -кривую и для случая $P(1 + \omega) > 1$, но тогда ограничивающий отрезок прямой проходит от точки с координатами $(1, 1)$ к точке с координатами $(0, 1 - \theta)$ и, следовательно, ограничивает ROC -кривую уже с другой стороны (рис. 5.8, в).

Усиленный ROC -анализ позволяет упростить процедуру оценки эффективности диагностического теста, придав ей наглядную интерпретацию в виде когнитивного графического образа.

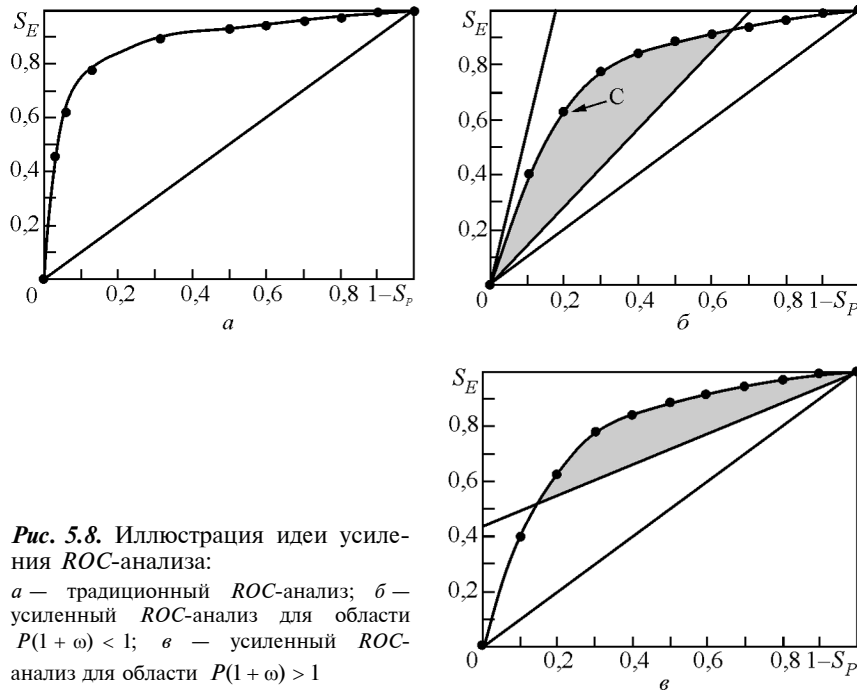


Рис. 5.8. Иллюстрация идеи усиления ROC-анализа:

a — традиционный ROC-анализ; *б* — усиленный ROC-анализ для области $P(1 + \omega) < 1$; *в* — усиленный ROC-анализ для области $P(1 + \omega) > 1$

Отметим также, что условия (5.33) и (5.34) позволяют решить и обратную задачу: определить допустимый интервал соотношения потерь ω , при котором тест с известными операционными характеристиками — чувствительностью S_E и специфичностью S_P — можно считать полезным для скрининга заболевания с известным преваленсом P .

Этот интервал определяется формулой

$$\frac{1 - P}{P} \frac{1 - S_P}{S_E} \leq \omega \leq \frac{1 - P}{P} \frac{S_P}{1 - S_E}. \quad (5.40)$$

Например, для заболевания с преваленсом $P = 15\%$ при чувствительности $S_E = 62,5\%$ и специфичности $S_P = 80\%$ теста его можно рекомендовать для практического применения, если допустимо соотношение потерь от ложноотрицательных и ложноположительных ошибок в пределах $2,66 \leq \omega \leq 17,7$ в предположении об эквивалентности стоимости правильных диагнозов.

5.3. Модель формирования коллективного решения в условиях противоречий

5.3.1. Постановка задачи

Рассмотрим теперь задачу построения коллективного решающего правила, интегрирующего частные решения группы независимых диагностических тестов.

Пусть некоторый объект находится в одном из M возможных состояний (классов) $V_1, \dots, V_M$ с известными априорными вероятностями $P(V_1), \dots, P(V_M)$, $\sum_{j=1}^M P(V_j) = 1$. Ясно, что если не

располагать какой-либо дополнительной информацией и всегда относить состояние объекта к классу, имеющему наибольшую априорную вероятность, то величина

$$P_0 = 1 - \max \{P(V_1), \dots, P(V_M)\} \quad (5.41)$$

определяет априорную вероятность ошибочной классификации.

Предположим, что есть N алгоритмов (диагностических тестов) $A_1, \dots, A_N$, которые на основании дополнительной информации независимо один от другого принимают решения δ_i о состоянии объекта в виде индикаторных переменных

$$\delta_i = k, \text{ если } A_i \text{ решает в пользу } V_k, i = 1, \dots, N, k = 1, \dots, M. \quad (5.42)$$

Будем характеризовать «квалификацию» i -го теста вероятностью $P(A_i)$ ошибочной классификации, которая считается известной для всех N тестов на основании предыдущего опыта. При этом естественно допустить, что вероятности $P(A_i)$ удовлетворяют условиям

$$P(A_i) < P_0 \quad \forall i = 1, \dots, N. \quad (5.43)$$

Ставится задача построения коллективного решающего правила $D = k$, если коллектив решает в пользу V_k , $k = 1, \dots, M$, которое минимизирует среднюю вероятность ошибки на множестве возможных комбинаций частных решений (5.42).

5.3.2. Байесова схема разрешения противоречий

Рассмотрим вначале простейший случай, когда число тестов $N = 2$ и число классов $M = 2$. В этом случае возможны четыре комбинации частных решений экспертов:

$$S_{11}: \delta_1 = 1, \delta_2 = 1;$$

$$S_{12}: \delta_1 = 1, \delta_2 = 2;$$

$$S_{21}: \delta_1 = 2, \delta_2 = 1;$$

$$S_{22}: \delta_1 = 2, \delta_2 = 2.$$

Как видно, в ситуациях S_{12} и S_{21} решения экспертов противоречивы. Возникает естественный вопрос: какое решение следует принимать, чтобы минимизировать вероятность ошибочной классификации?

На первый взгляд может показаться, что в условиях противоречий следует выбрать то решение, которое принял тест, допускающий меньшие частные вероятности ошибки.

Покажем, что в общем случае такой подход неправилен [209].

С этой целью рассмотрим условные (апостериорные) вероятности $P(V_1/S_{12})$ и $P(V_2/S_{12})$ классов в ситуации S_{12} . При этом для минимизации средней вероятности ошибочной классификации будем принимать окончательное решение в пользу класса V_1 , если

$$P(V_1/S_{12}) > P(V_2/S_{12}), \quad (5.44)$$

и решение в пользу V_2 в противном случае.

По формуле Байеса имеем

$$P(V_1 / S_{12}) = \frac{P(V_1)P(S_{12} / V_1)}{P(S_{12})},$$

$$P(V_2 / S_{12}) = \frac{P(V_2)P(S_{12} / V_2)}{P(S_{12})}.$$

Очевидно, что неравенство (5.44) справедливо в том и только в том случае, когда

$$P(V_1)P(S_{12}/V_1) > P(V_2)P(S_{12}/V_2). \quad (5.45)$$

По определению условная вероятность $P(S_{12}/V_1)$ — не что иное, как вероятность того, что в ситуации, когда имеет место класс V_1 , тест A_2 принял правильное решение, а тест A_1 ошибся. Поскольку предполагалось, что решения тестов независимы, то по формуле произведения вероятностей

$$P(S_{12}/V_1) = [1 - P(A_1)]P(A_2). \quad (5.46)$$

Аналогичным образом

$$P(S_{12}/V_2) = P(A_1)[1 - P(A_2)]. \quad (5.47)$$

Неравенство (5.45) с учетом (5.46), (5.47), можно представить в виде

$$P(V_1)[1 - P(A_1)] P(A_2) > P(V_2) P(A_1) [1 - P(A_2)]. \quad (5.48)$$

Из (5.48) вытекает, что в ситуации S_{12} , когда решения тестов противоречивы, объект следует относить к классу V_1 в том и только том случае, когда

$$P(A_2) > \frac{\lambda_0 P(A_1)}{1 - P(A_1)(1 - \lambda_0)}, \quad (5.49)$$

где $\lambda_0 = P(V_2)/(P(V_1) + P(V_2))$ — отношение априорных вероятностей классов.

Если же выполняется соотношение

$$P(A_2) < \frac{\lambda_0 P(A_1)}{1 - P(A_1)(1 - \lambda_0)}, \quad (5.50)$$

то в ситуации S_{12} объект следует относить к классу V_2 .

Аналогичным образом легко показать, что в ситуации S_{21} решение в пользу класса V_1 следует принимать в том и только в том случае, когда

$$P(A_1) > \frac{\lambda_0 P(A_2)}{1 - P(A_2)(1 - \lambda_0)}, \quad (5.51)$$

а решение в пользу класса V_2 , когда

$$P(A_1) < \frac{\lambda_0 P(A_2)}{1 - P(A_2)(1 - \lambda_0)}. \quad (5.52)$$

Отметим, из (5.49)—(5.52) непосредственно следует, что только при равновероятных классах, когда $\lambda_0 = 1$, окончательное решение совпадает с решением того из тестов, который имеет меньшую вероятность ошибки. В остальных же случаях, когда $\lambda_0 \neq 1$, т. е. $P(V_1) \neq P(V_2)$, окончательное решение определяется не только «квалификацией» тестов, но и соотношением априорных вероятностей классов. При этом окончательное решение может не совпадать с решением более «квалифицированного» теста.

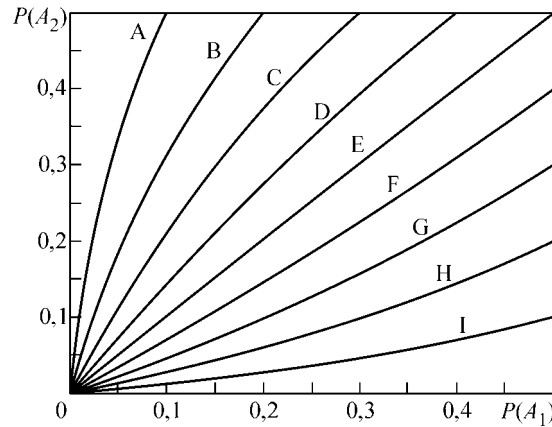


Рис. 5.9. Границы областей коллективных решений в ситуации S_{12} :
 $\lambda_0 = 9$ (A); $\lambda_0 = 4$ (B); $\lambda_0 = 2,33$ (C); $\lambda_0 = 1,5$ (D); $\lambda_0 = 1$ (E); $\lambda_0 = 0,67$ (F); $\lambda_0 = 0,43$ (G);
 $\lambda_0 = 0,25$ (H); $\lambda_0 = 0,11$ (I).

На рис. 5.9 показаны границы областей решений, построенные для ситуации S_{12} согласно условиям (5.49), (5.50) при различных значениях λ_0 . Область решения в пользу класса V_1 расположена выше соответствующей границы, а решений в пользу класса V_2 — ниже соответствующей границы.

Рассмотрим модельный пример. Пусть $P(V_1) = 0,3$, $P(V_2) = 0,7$, т. е. $\lambda_0 = 2,33$. Пусть на основании предыдущего опыта известно, что первый тест ошибается в 5 % случаев ($P(A_1) = 0,05$), а второй — в 8 % случаев ($P(A_2) = 0,08$).

Предположим, что тест A_1 отнес объект к классу V_1 , а тест A_2 — к классу V_2 , т. е. наблюдаем ситуацию S_{12} противоречивых решений. Заметим, что первый тест более «квалифицированный», поскольку $P(A_1) < P(A_2)$.

Как видно из рисунка 5.9, точка с координатами $P(A_1) = 0,05$ и $P(A_2) = 0,08$ расположена ниже границы, соответствующей $\lambda_0 = 2,33$. Следовательно, объект должен быть отнесен к классу V_2 , хотя более квалифицированный тест A_1 принял противоположное решение.

Для проверки обоснованности решения в пользу V_2 определим по формуле Байеса апостериорные вероятности классов в рассматриваемой ситуации S_{12} :

$$P(V_1 / S_{12}) = \frac{P(V_1)P(S_{12} / V_1)}{P(S_{12})} =$$

$$= \frac{0,3 \cdot (1 - 0,05) \cdot 0,08}{0,3 \cdot (1 - 0,05) \cdot 0,08 + 0,7 \cdot 0,05 \cdot (1 - 0,08)} = 0,415 \quad (5.53)$$

и

$$P(V_2 / S_{12}) = \frac{P(V_2)P(S_{12} / V_2)}{P(S_{12})} =$$

$$= \frac{0,7 \cdot 0,05 \cdot (1 - 0,08)}{0,3 \cdot (1 - 0,05) \cdot 0,08 + 0,7 \cdot 0,05 \cdot (1 - 0,08)} = 0,585. \quad (5.54)$$

Как видим, $P(V_1 / S_{12}) < P(V_2 / S_{12})$, значит, объект действительно следует отнести к классу V_2 .

Изменим в условиях примера соотношения априорных вероятностей классов, положив $P(V_1) = 0,4$, $P(V_2) = 0,6$. В этом случае $\lambda = 0,67$ и, как видно из рис. 5.9, точка с координатами $P(A_1) = 0,05$ и $P(A_2) = 0,08$ попадает уже в область решений в пользу класса V_1 . В самом деле

$$P(V_1 / S_{12}) = \frac{0,4 \cdot (1 - 0,05) \cdot 0,08}{0,4 \cdot (1 - 0,05) \cdot 0,08 + 0,6 \cdot 0,05 \cdot (1 - 0,08)} = 0,524$$

и

$$P(V_2 / S_{12}) = \frac{0,6 \cdot 0,05 \cdot (1 - 0,08)}{0,4 \cdot (1 - 0,05) \cdot 0,08 + 0,6 \cdot 0,05 \cdot (1 - 0,08)} = 0,476,$$

т. е. $P(V_1 / S_{12}) > P(V_2 / S_{12})$. Значит, в этом случае объект следует отнести уже к классу V_1 .

Отметим, что рассматриваемая схема принятия коллективного решения основывается на знании весьма ограниченных вероятностных характеристик, которые при решении практических задач, в частности задач медицинской и технической диагностики, легко могут оцениваться, учитывая предыдущий опыт.

При этом не требуется знать, на основании какой информации принимаются частные решения и как именно эти решения принимаются — на основе формальных или эвристических алгоритмов.

Наряду с этим мы сделали одно важное допущение о том, что решения тестов независимы, а вероятность ошибки каждого теста не зависит от класса, т. е. $P(A_i) = P(A_i / V_1) = P(A_i / V_2)$. Естественно, что такое допущение должно быть обосновано.

Для того чтобы продемонстрировать практическую возможность таких допущений, рассмотрим один из формальных algo-

ТАБЛИЦА 5.4

Условные распределения признаков в первом классе

x_1	x_2		
	$x_2 = x_2^-$	$x_2 = x_2^+$	$P(x_1/V_1)$
$x_1 = x_1^-$	0,076	0,874	0,95
$x_1 = x_1^+$	0,004	0,046	0,05
$P(x_2/V_1)$	0,08	0,92	—

ТАБЛИЦА 5.5

Условные распределения признаков во втором классе

x_1	x_2		
	$x_2 = x_2^-$	$x_2 = x_2^+$	$P(x_1/V_2)$
$x_1 = x_1^-$	0,046	0,004	0,05
$x_1 = x_1^+$	0,874	0,076	0,95
$P(x_2/V_2)$	0,92	0,08	—

ритмов принятия частных решений двумя независимыми тестами.

Предположим, что тесты классифицируют объект с априорными вероятностями $P(V_1) = 0,3$, $P(V_2) = 0,7$ по бинарным признакам, условные распределения которых заданы табл. 5.4 и 5.5.

Будем считать, что частные решения тестов принимаются по правилу максимума апостериорных вероятностей. Для определения вероятности ошибочных решений теста A_1 найдем апостериорные вероятности классов при возможных значениях признака x_1 :

$$\begin{aligned}
 P(V_1 / x_1^-) &= \frac{P(V_1)p(x_1^- / V_1)}{P(V_1)p(x_1^- / V_1) + P(V_2)p(x_1^- / V_2)} = \\
 &= \frac{0,3 \cdot 0,95}{0,3 \cdot 0,95 + 0,7 \cdot 0,05} = 0,891,
 \end{aligned}$$

$$P(V_2 / x_1^-) = \frac{P(V_2)p(x_1^- / V_1)}{P(V_1)p(x_1^- / V_1) + P(V_2)p(x_1^- / V_2)} =$$

$$= \frac{0,7 \cdot 0,05}{0,3 \cdot 0,95 + 0,7 \cdot 0,05} = 0,109,$$

$$\begin{aligned} P(V_1 / x_1^+) &= \frac{P(V_1)p(x_1^+ / V_1)}{P(V_1)p(x_1^+ / V_1) + P(V_2)p(x_1^+ / V_2)} = \\ &= \frac{0,3 \cdot 0,05}{0,3 \cdot 0,05 + 0,7 \cdot 0,95} = 0,022, \end{aligned}$$

$$\begin{aligned} P(V_2 / x_1^+) &= \frac{P(V_2)p(x_1^+ / V_2)}{P(V_1)p(x_1^+ / V_1) + P(V_2)p(x_1^+ / V_2)} = \\ &= \frac{0,7 \cdot 0,95}{0,3 \cdot 0,05 + 0,7 \cdot 0,95} = 0,978. \end{aligned}$$

Поскольку $P(V_1 / x_1^-) > P(V_2 / x_1^-)$, а $P(V_1 / x_1^+) < P(V_2 / x_1^+)$, то тест A_1 классифицирует объект по признаку x_1 следующим образом:

$$\delta_1 = \begin{cases} 1, & \text{если } x_1 = x_1^-, \\ 2, & \text{если } x_1 = x_1^+. \end{cases} \quad (5.55)$$

Очевидно, что $P(x_1 = x_1^+ / V_1) = P(x_1 = x_1^- / V_2) = 0,05$, а значит, вероятность ошибки первого теста не зависит от класса, т. е. $P(A_1) = P(A_1 / V_1) = P(A_1 / V_2) = 0,05$.

Аналогичным образом находим

$$\begin{aligned} P(V_1 / x_2^-) &= \frac{P(V_1)p(x_2^- / V_1)}{P(V_1)p(x_2^- / V_1) + P(V_2)p(x_2^- / V_2)} = \\ &= \frac{0,3 \cdot 0,08}{0,3 \cdot 0,08 + 0,7 \cdot 0,92} = 0,036, \end{aligned}$$

$$\begin{aligned} P(V_2 / x_2^-) &= \frac{P(V_2)p(x_2^- / V_2)}{P(V_1)p(x_2^- / V_1) + P(V_2)p(x_2^- / V_2)} = \\ &= \frac{0,7 \cdot 0,92}{0,3 \cdot 0,08 + 0,7 \cdot 0,92} = 0,964, \end{aligned}$$

$$\begin{aligned}
P(V_1 / x_2^+) &= \frac{P(V_1)p(x_2^+ / V_1)}{P(V_1)p(x_2^+ / V_1) + P(V_2)p(x_2^+ / V_2)} = \\
&= \frac{0,3 \cdot 0,92}{0,3 \cdot 0,92 + 0,7 \cdot 0,08} = 0,831, \\
P(V_2 / x_2^-) &= \frac{P(V_2)p(x_2^- / V_2)}{P(V_1)p(x_2^- / V_1) + P(V_2)p(x_2^- / V_2)} = \\
&= \frac{0,7 \cdot 0,08}{0,3 \cdot 0,92 + 0,7 \cdot 0,08} = 0,169.
\end{aligned}$$

Поскольку $P(V_1 / x_2^-) < P(V_2 / x_2^-)$, а $P(V_1 / x_2^+) > P(V_2 / x_2^+)$, тест A_2 классифицирует объект по признаку x_2 следующим образом:

$$\delta_2 = \begin{cases} 1, & \text{если } x_2 = x_2^+, \\ 2, & \text{если } x_2 = x_2^-. \end{cases} \quad (5.56)$$

При этом вероятность ошибки теста A_2 составляет $P(A_2) = 0,08$, причем эта вероятность также не зависит от класса, т. е. $P(A_2) = P(A_2 / V_1) = P(A_2 / V_2)$.

Как видно из табл. 5.4 и 5.5, признаки x_1 и x_2 статистически независимы в обоих классах, поскольку для любых их значений справедливо соотношение $p(x_1, x_2 / V_k) = p(x_1 / V_k) p(x_2 / V_k)$ при $k = 1, 2$. Отсюда непосредственно следует, что и решения (5.55), (5.56) тестов будут независимыми.

Предположим, что в момент принятия решений признаки получили следующие значения: $x_1 = x_1^-$, а $x_2 = x_2^-$. В этом случае, согласно (5.55), (5.56), $\delta_1 = 1$ и $\delta_2 = 2$, т. е. требуется принять окончательное решение в условиях противоречивой ситуации S_{12} . В этой ситуации, как было показано выше, окончательное решение следует принимать в пользу V_2 .

Покажем, что такое решение совпадает с оптимальным решением, основанным на результатах измерения совокупности двух признаков по формальному правилу максимума апостериорных вероятностей классов $P(V_1 / x_1, x_2)$ и $P(V_2 / x_1, x_2)$. По формуле Байеса определим эти вероятности при $x_1 = x_1^-$ и $x_2 = x_2^-$:

$$\begin{aligned}
 P(V_1 / x_1^- x_2^-) &= \frac{P(V_1)p(x_1^-, x_2^- / V_1)}{P(V_1)p(x_1^-, x_2^- / V_1) + P(V_2)p(x_1^-, x_2^- / V_2)} = \\
 &= \frac{0,3 \cdot 0,076}{0,3 \cdot 0,076 + 0,7 \cdot 0,046} = 0,415
 \end{aligned} \quad (5.57)$$

и

$$\begin{aligned}
 P(V_2 / x_1^- x_2^-) &= \frac{P(V_2)p(x_1^-, x_2^- / V_2)}{P(V_1)p(x_1^-, x_2^- / V_1) + P(V_2)p(x_1^-, x_2^- / V_2)} = \\
 &= \frac{0,7 \cdot 0,046}{0,3 \cdot 0,076 + 0,7 \cdot 0,046} = 0,585.
 \end{aligned} \quad (5.58)$$

Сравнение (5.53), (5.54) с (5.57), (5.58) позволяет заключить, что $P(V_k / x_1^-, x_2^-) = P(V_k / S_{12})$ для $k = 1, 2$. Несложно убедиться в том, что аналогичные равенства имеют место и при других возможных значениях признаков.

Следовательно, если тесты принимают независимые решения, причем $P(A_i / V_1) = P(A_i / V_2)$ при $i = 1, 2$, то предлагаемая схема эквивалентна оптимальной, обеспечивающей минимум средней вероятности ошибочной классификации по совокупности двух независимых признаков в классах.

В условиях данного примера средняя вероятность ошибочных решений, принимаемых по результатам классификации двух тестов, составляет 0,0406. Отметим, что это значение меньше вероятности ошибки частных решений каждого из тестов.

Предложенное правило легко может быть модифицировано для принятия коллективного решения по совокупности независимых тестов с известными операционными характеристиками.

Пусть требуется отнести обследуемого пациента к одному из двух классов: V_1 (больной), V_2 (здоровый), на основании информации, полученной от двух независимых тестов A_1, A_2 .

При этом считаются известными преваленс заболевания P , а также операционные характеристики тестов — чувствительность $S_E^{(i)}$ и специфичность $S_P^{(i)}$, $i = 1, 2$.

Тогда в ситуации S_{12} , когда A_1 признал пациента больным, а A_2 — здоровым, окончательный диагноз, обеспечивающий минимум средней вероятности ошибочного решения, должен приниматься согласно схеме

$$\begin{aligned} &\text{болен, если } S_E^{(1)}(1 - S_E^{(2)}) > \lambda_0[1 - S_P^{(1)}] S_P^{(2)}, \\ &\text{здоров, если } S_E^{(1)}(1 - S_E^{(2)}) < \lambda_0[1 - S_P^{(1)}] S_P^{(2)}, \end{aligned} \quad (5.59)$$

где $\lambda_0 = (1 - P) / P$ — параметр, определяемый преваленсом P заболевания в тестируемой группе.

5.3.3. Обучаемая система принятия коллективного решения

Обобщение предложенной схемы для общего случая позволяет не только формировать коллективные решения совокупности диагностических тестов, но и группы независимых экспертов.

Пусть N экспертов $A_1, \dots, A_N$ независимо один от другого принимают решения δ_j о состоянии объекта, который находится в одном из M состояний $V_1, \dots, V_M$, в виде индикаторных переменных (5.42).

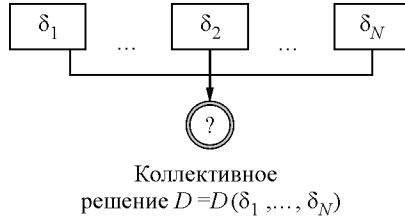
Понятно, что в общем случае число возможных комбинаций частных решений равно M^N . При этом только в M случаях решения будут согласованными (когда все эксперты примут решения в пользу одного класса), а в остальных случаях — решения противоречивы.

Предположим, что на основании предыдущего опыта получены оценки условных вероятностей ошибок $P(A_1 / V_k), \dots, P(A_N / V_k)$, $k = 1, \dots, M$, которые допускают эксперты, а также оценки вероятностей $P(V_k)$.

Построим формальную схему интеграции частных решений группы независимых экспертов, позволяющую принимать коллективное решение $D = D(\delta_1, \dots, \delta_N)$ о текущем состоянии объекта (рис. 5.10).

Пусть в результате обследования объекта получена некоторая комбинация S частных решений $\delta_1, \dots, \delta_N$ в форме (5.42). Обозначим I_m — множества номеров экспертов, принявших решение в пользу m -го класса ($m = 1, \dots, M$). Очевидно, что $I_i \cap I_j = \emptyset$ для любых $i, j = 1, \dots, M$ и $I_1 \cup \dots \cup I_M = \{1, \dots, N\}$.

Для минимизации средней вероятности ошибки коллективного решения $D = D(\delta_1, \dots, \delta_N)$ на множестве Θ возможных комбинаций частных решений будем для каждой фиксированной комбинации $S \in \Theta$ принимать окончательное решение в пользу

ЧАСТНЫЕ РЕШЕНИЯ N ЭКСПЕРТОВ**Рис. 5.10.** Задача формирования коллективного решения

того из классов, который имеет наибольшую апостериорную вероятность:

$$D(S) = \arg \max_{1 \leq m \leq M} P(V_m / S), \quad (5.60)$$

где

$$P(V_m / S) = \frac{P(V_m)P(S / V_m)}{P(S)}.$$

По определению $P(S / V_m)$ — вероятность того, что в ситуации, когда объект находится в состоянии V_m , эксперты, номера которых принадлежат множеству I_m , приняли правильные решения, а остальные ошиблись. Поскольку предполагается,

что решения экспертов независимы, то по формуле произведения вероятностей

$$P(S / V_m) = \prod_{j \in I_m} [1 - P(A_j / V_m)] \prod_{j \notin I_m} P(A_j / V_m). \quad (5.61)$$

На основании (5.60) с учетом (5.61) заключаем, что в условиях противоречивых частных решений окончательное (коллективное) решение следует принимать по схеме

$$D(S) = \arg \max_{1 \leq m \leq M} P(V_m) \prod_{j \in I_m} [1 - P(A_j / V_m)] \prod_{j \notin I_m} P(A_j / V_m). \quad (5.62)$$

Для иллюстрации схемы формирования коллективного решения по правилу (5.62) рассмотрим модельный пример.

Пусть некоторый объект находится в одном из трех классов состояний, образующих полную группу случайных событий с априорными вероятностями $P(V_1) = 0,7$, $P(V_2) = 0,08$ и $P(V_3) = 0,22$. Состояние объекта оценивается пятью независимыми экспертами. Вероятности ошибок экспертов и возможная комбинация принятых ими частных решений представлены в табл. 5.6.

Видно, что в данном случае частные решения экспертов противоречивы, причем $I_1 = \{1\}$, $I_2 = \{3, 4\}$, $I_3 = \{2, 5\}$.

Для принятия коллективного решения вычислим следующие величины:

$$P(V_1) \prod_{j \in I_1} [1 - P(A_j / V_1)] \prod_{j \notin I_1} P(A_j / V_1) = 4,03 \cdot 10^{-8},$$

$$P(V_2) \prod_{j \in I_2} [1 - P(A_j / V_2)] \prod_{j \notin I_2} P(A_j / V_2) = 1,12 \cdot 10^{-6},$$

$$P(V_3) \prod_{j \in I_3} [1 - P(A_j / V_3)] \prod_{j \notin I_3} P(A_j / V_3) = 3,73 \cdot 10^{-6}.$$

Поскольку третья из найденных величин максимальна, то на основании правила (5.62) принимаем окончательное решение в пользу класса V_3 . Отметим, что этот класс не имеет самую высокую априорную вероятность.

Вполне понятно, что при решении практических задач чаще всего не известны точные значения вероятностных характеристик, фигурирующих в правиле (5.62). Однако при достаточном объеме наблюдений вероятности $P(V_k)$ и $P(A_i / V_k)$ могут быть оценены соответствующими частотами:

$$\hat{P}(V_k) = G_k / G, \quad (5.63)$$

$$\hat{P}(A_i / V_k) = E_{ki} / G_k, \quad (5.64)$$

где G_k — число появлений k -го класса ($k = 1, \dots, M$) в выборке из G наблюдений, а E_{ki} — число ошибочных решений i -го эксперта ($i = 1, \dots, N$) при анализе ситуаций, когда объект Z находился в k -м состоянии.

Рассмотрим схему оценки частот (5.63), (5.64), которая удобна для практического применения и может быть положена в основу системы поддержки принятия коллективного решения.

ТАБЛИЦА 5.6

Частные решения и «квалификации» экспертов

Эксперт	Вероятности ошибок			Частные решения
	$P(A_i / V_1)$	$P(A_i / V_2)$	$P(A_i / V_3)$	
A_1	0,04	0,01	0,03	$\delta_1 = 1$
A_2	0,01	0,03	0,02	$\delta_2 = 3$
A_3	0,03	0,05	0,01	$\delta_3 = 2$
A_4	0,02	0,02	0,06	$\delta_4 = 2$
A_5	0,01	0,05	0,04	$\delta_5 = 3$

Предположим, что для каждого из G наблюдений известна точная принадлежность объекта одному из возможных классов, выраженная в виде указаний «учителя» $y[1], y[2], \dots, y[G]$, где $y[n] \in \{1, \dots, M\}$. Запишем частоту появления k -го класса, оцененную согласно (5.63), в виде

$$\hat{P}(V_k) = \frac{\sum_{n=1}^G \chi_k[n]}{G}, \quad (5.65)$$

где

$$\chi_k[n] = \begin{cases} 1, & \text{если } y[n] = k, \\ 0, & \text{если } y[n] \neq k. \end{cases}$$

Поскольку правую часть (5.65) можно выразить в виде суммы двух слагаемых

$$\frac{\sum_{n=1}^G \chi_k[n]}{G} = \frac{G-1}{G} \frac{\sum_{n=1}^{G-1} \chi_k[n]}{G-1} + \frac{\chi_k[G]}{G},$$

в первом из которых фигурирует оценка частоты появления k -го класса, вычисленная по $G-1$ наблюдениям, то после очевидных преобразований получим

$$\hat{P}_G(V_k) = \hat{P}_{G-1}(V_k) - \frac{1}{G} (\hat{P}_{G-1}(V_k) - \chi_k[G]). \quad (5.66)$$

Для оценки вероятностей ошибок экспертов рассмотрим последовательность $y_k[1], \dots, y_k[G_k]$ указаний «учителя», которые удовлетворяют условию $y_k[n] = k$. Очевидно, что величина E_{ki} , фигурирующая в правой части (5.64), может быть записана в виде суммы

$$E_{ki} = \sum_{n=1}^{G_k} \eta_{ki}[n],$$

где $\eta_{ki}[n]$ — штрафная функция, выраженная в форме

$$\eta_{ki}[n] = \begin{cases} 0, & \text{если } \delta_i[n] = k; \\ 1, & \text{если } \delta_i[n] \neq k. \end{cases}$$

Тогда оценка вероятности ошибки i -го эксперта при появлении k -го класса также может быть найдена по рекуррентной

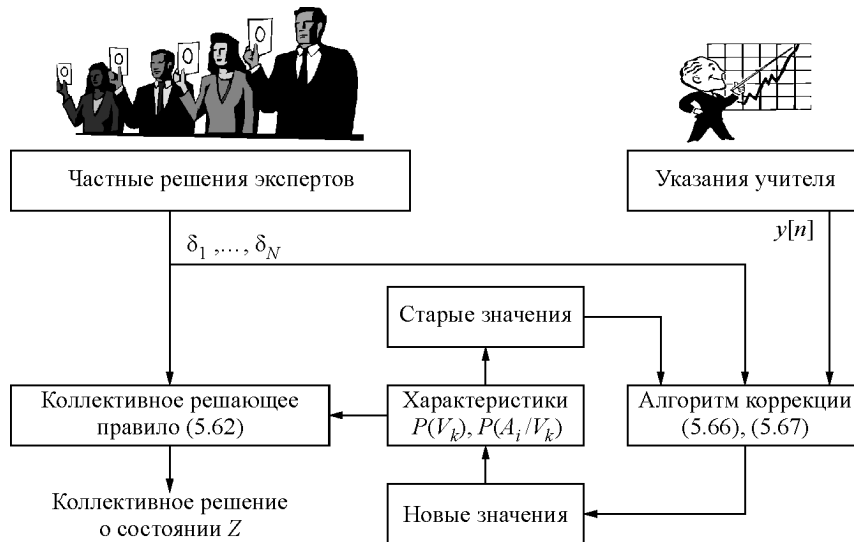


Рис. 5.11. Структура системы формирования коллективного решения

формуле

$$\hat{P}_{G_k}(A_i/V_k) = \hat{P}_{G_{k-1}}(A_i/V_k) - \frac{1}{G_k}(\hat{P}_{G_{k-1}}(A_i/V_k) - \eta_{ki}[G_k]). \quad (5.67)$$

Из выражений (5.66), (5.67) видно стремление к нулю величины поправки при неограниченном росте числа наблюдений, что, естественно, согласуется с предельной теоремой Бернулли о сходимости по вероятности частоты случайного события к его вероятности.

На основе предложенного подхода легко может быть реализована система поддержки принятия коллективных решений (рис. 5.11).

Коллективное решение реализуется на основе частных решений $\delta_1, \dots, \delta_N$ группы независимых экспертов $A_1, \dots, A_N$ по правилу (5.62), в котором используются текущие значения оценок вероятностных характеристик $P(V_k)$ и $P(A_i/V_k)$, $i = 1, \dots, N$, $k = 1, \dots, M$.

В базе данных (БД) системы фиксируются объемы наблюдений $G_1, \dots, G_M$ выборки $G = G_1 + \dots + G_M$, по которым были оценены указанные вероятностные характеристики для соответствующих классов.

Как только после принятия коллективного решения появляется возможность проверить истинное состояние объекта, дополнительная информация вводится в систему в виде указания «учителя» $y[n]$. С помощью рекуррентных формул (5.66) и (5.67) эта информация позволяет скорректировать текущие оценки $P(V_k)$ и $P(A_i/V_k)$, $i = 1, \dots, N$, $k = 1, \dots, M$.

Предложенная архитектура системы, совмещающая коллективное решение с возможностью периодической коррекции характеристик решающего правила, может быть использована в различных областях. Например, эту систему можно рекомендовать для диагностики заболеваний сердца, основанной на объединении методов электрокардиографии и магнитокардиографии. В данном случае указаниями учителя служат имеющиеся в распоряжении медиков результаты референтного метода (коронарографии), который по медицинским показаниям назначают некоторым из обследуемых пациентов.

5.4. Оценка полезности признаков при диагональной матрице потерь

5.4.1. Общее условие бесполезности релевантных признаков

Как уже отмечалось, при диагональной матрице потерь (5.2), когда $L_{11} = L_{22} = 0$, а $L_{12} = L_{21} = 1$, средний риск (5.3) сводится к средней вероятности ошибки

$$P(E) = \sum_{x \in X} p(x) \min \{P(V_1/x), P(V_2/x)\}, \quad (5.68)$$

где $p(x) = P(V_1)p(x/V_1) + P(V_2)p(x/V_2)$ — безусловное распределение признака x .

Согласно определению М. Бен-Бассата [256], диагностический признак x считается нерелевантным (*irrelevant feature*), если апостериорная вероятность k -го класса сохраняется неизменной при всех возможных значениях x , т. е.

$$P(V_k/x) = P(V_k) \quad \forall x \in X, \quad (5.69)$$

где X — носитель распределения $p(x)$.

Легко показать, что при числе классов $M = 2$ условие (5.69) эквивалентно тому, что признак x имеет тождественно совпа-

дающие условные распределения в классах, т. е.

$$p(x / V_1) \equiv p(x / V_2). \quad (5.70)$$

Совершенно очевидно, что при выполнении условия (5.70) признак x оказывается бесполезным, т. е. средняя вероятность ошибочных решений $P(E)$, принимаемых по результатам измерения значений этого признака, совпадает с априорной вероятностью ошибки $P_0 = \min \{P(V_1), P(V_2)\}$.

Покажем [215], что в общем случае обратное утверждение неверно: релевантный признак, имеющий различные распределения в классах

$$p(x / V_1) \neq p(x / V_2), \quad (5.71)$$

может оказаться бесполезным.

Для того чтобы формально определить условие бесполезности релевантного признака и продемонстрировать возможность выполнения такого условия как для дискретных, так и для непрерывных распределений, докажем следующую лемму.

Лемма 5.1. Пусть $P(V_1) \neq P(V_2)$. Тогда признак x является бесполезным, т. е. $P(E) = P_0$ в том и только в том случае, когда

$$[P(V_1) - P(V_2)][P(V_1 / x) - P(V_2 / x)] > 0 \quad \forall x \in X, \quad (5.72)$$

иными словами, для каждого возможного значения $x \in X$ знак разности апостериорных вероятностей

$$\Delta(x) = P(V_1 / x) - P(V_2 / x)$$

совпадает со знаком разности априорных вероятностей

$$\Delta_0 = P(V_1) - P(V_2).$$

Доказательство. Предположим, что неравенство (5.72) выполняется, например, $P(V_1) > P(V_2)$ и $P(V_1 / x) > P(V_2 / x)$ для всех возможных $x \in X$. Это значит, что

$$\min \{P(V_1), P(V_2)\} = P(V_2)$$

и

$$\min \{P(V_1 / x), P(V_2 / x)\} = P(V_2 / x) \quad \forall x \in X.$$

Следовательно, если выполняется условие (5.72), то в соответствии с (5.68) имеем

$$P(E) = \sum_{x \in X} p(x)P(V_2 / x) = P(V_2) = \min \{P(V_1), P(V_2)\} = P_0.$$

Предположим теперь, что условие (5.71) не выполняется, например $P(V_1) > P(V_2)$, но существует подмножество $X^* \subset X$ значений x , таких, что

$$P(V_1 / x) < P(V_2 / x) \quad \forall x \in X^*.$$

Тогда в соответствии с (5.68) имеем

$$P(E) = \sum_{x \in X \setminus X^*} p(x)P(V_2 / x) + \sum_{x \in X^*} p(x)P(V_1 / x). \quad (5.73)$$

Заменяя в правой части (5.73) $P(V_1 / x)$ на $P(V_2 / x)$ с учетом того, что $P(V_1 / x) < P(V_2 / x) \quad \forall x \in X^*$, заключаем, что если $X^* \neq \emptyset$, то

$$P(E) < \sum_{x \in X} p(x)P(V_2 / x),$$

т. е. $P(E) < P_0$. Лемма доказана.

Отметим, что условия леммы 5.1 вовсе не предполагают выполнения условия (5.70). Отсюда как раз и следует, что релевантный признак может быть бесполезным.

Продemonстрируем такую возможность на примерах дискретных и непрерывных распределений диагностических признаков.

5.4.2. Признак-симптом, принимающий два значения

Пусть x — бинарный признак (симптом), принимающий только два возможных значения $x = x^+$ и $x = x^-$ с условными вероятностями $p(x^+ / V_k)$, $p(x^- / V_k)$, $k = 1, 2$. Очевидно, что $p(x^- / V_k) = 1 - p(x^+ / V_k)$.

Утверждение 5.5. Если априорные вероятности классов не равны, т. е. $P(V_1) \neq P(V_2)$, то для любой заданной условной вероятности $p(x^+ / V_1)$ может быть указана такая условная вероят-

ность $p(x^+ / V_2)$, что

$$p(x^+ / V_1) \neq p(x^+ / V_2), \quad (5.74)$$

но при этом выполняется равенство $P(E) = P_0$, т. е. x — релевантный, но бесполезный признак.

Доказательство. Предположим вначале, что $P(V_2) < P(V_1)$. Пусть Ω — множество возможных значений $p(x^+ / V_1)$ и $p(x^+ / V_2)$, которые удовлетворяют неравенствам

$$p(x^+ / V_1) > \lambda_0 p(x^+ / V_2), \quad (5.75)$$

$$p(x^+ / V_1) < \lambda_0 p(x^+ / V_2) + (1 - \lambda_0), \quad (5.76)$$

где $\lambda_0 = P(V_2) / P(V_1)$ — отношение априорных вероятностей классов.

Поскольку $\lambda_0 < 1$ при $P(V_2) < P(V_1)$, то для заданного значения вероятности $p(x^+ / V_1)$ всегда можно указать соответствующее значение вероятности $p(x^+ / V_2) \neq p(x^+ / V_1)$, при котором неравенства (5.75) и (5.76) выполняются совместно, т. е. $\Omega \neq \emptyset$.

По формуле Байеса с учетом того, что $p(x^- / V_k) = 1 - p(x^+ / V_k)$ для $k = 1, 2$, из системы неравенств (5.75) и (5.76) следует

$$p(V_1 / x^+) > p(V_2 / x^+) \text{ и } p(V_1 / x^-) > p(V_2 / x^-).$$

Но это свидетельствует о том, что выполняется условие (5.70), а значит, в соответствии с леммой 5.1 выполняется равенство $P(E) = P_0$.

Доказательство утверждения 5.5 для случая $P(V_2) > P(V_1)$ выполняется аналогичным образом, но в этом случае вместо (5.75), (5.76) рассматриваются неравенства

$$p(x^+ / V_1) < \lambda_0 p(x^+ / V_2), \quad (5.77)$$

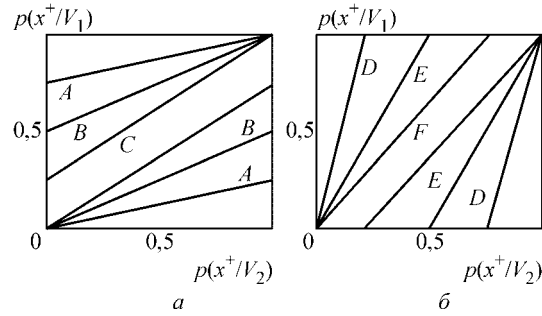


Рис. 5.12. Области бесполезности бинарного признака:

$\lambda_0 = 0,25$ (A); $\lambda_0 = 0,5$ (B); $\lambda_0 = 0,75$ (C); $\lambda_0 = 4,0$ (D); $\lambda_0 = 2,0$ (E); $\lambda_0 = 1,33$ (F)

$$p(x^+ / V_1) > \lambda_0 p(x^+ / V_2) + (1 - \lambda_0), \quad (5.78)$$

которые определяют множество $\Omega \neq \emptyset$ при $\lambda_0 > 1$.

Заметим, что $\Omega = \emptyset$ при $\lambda_0 = 1$, поскольку в этом случае неравенства (5.75), (5.76), так же как и неравенства (5.77), (5.78), не могут выполняться совместно. Этот факт позволяет сформулировать такое следствие.

Следствие 5.1. Если априорные вероятности классов равны, т. е. $P(V_1) = P(V_2)$, то при выполнении условия (5.74) релевантный бинарный признак заведомо полезен в смысле условия $P(E) < P_0$.

На рис. 5.12 показаны границы областей Ω бесполезности релевантного признака при различных значениях λ_0 . Как видно, область Ω уменьшается по мере приближения λ_0 к единице, и $\Omega = \emptyset$, когда λ_0 равно 1.

Рассмотрим модельный пример. Пусть $P(V_1) = 0,8$; $P(V_2) = 0,2$. Тогда $\lambda_0 = 0,25$. Предположим, что $p(x^+ / V_1) = 0,7$. Выберем соответствующее значение $p(x^+ / V_2) = 0,2$, при котором точка с координатами $p(x^+ / V_2) = 0,2$ и $p(x^+ / V_1) = 0,7$ принадлежит области A (рис. 5.12, a). Заметим, что при этом бинарный признак x имеет различные распределения в классах: $p(x^+ / V_1) \neq p(x^+ / V_2)$.

Для того чтобы убедиться в бесполезности признака, вычислим по формуле Байеса апостериорные вероятности классов при различных значениях признака:

$$P(V_1 / x^+) = 0,933; \quad P(V_2 / x^+) = 0,067; \quad P(V_1 / x^-) = 0,6;$$

$$P(V_2 / x^-) = 0,4.$$

Легко видеть, что

$$P(V_1 / x^+) > P(V_2 / x^+) \text{ и } P(V_1 / x^-) > P(V_2 / x^-),$$

т. е. при любом значении признака оптимальным является решение в пользу класса V_1 , совпадающее с априорным решением. Отсюда следует, что $P(E) = P_0 = 0,2$.

Полученные результаты позволяют сформулировать утверждение, определяющее необходимые и достаточные условия полезности признака-симптома.

Утверждение 5.6. Бинарный признак x , принимающий два возможных значения x^+ и x^- , обеспечивает диагностику состояний V_1 и V_2 со средней вероятностью ошибок меньше априорной ошибки в том и только в том случае, когда выполняется одно из следующих условий:

- а) $p(x^+ / V_1) \neq p(x^+ / V_2)$, если $\lambda_0 = 1$;
- б) $p(x^+ / V_1) < \lambda_0 p(x^+ / V_2)$ или $p(x^+ / V_1) > \lambda_0 p(x^+ / V_2) + (1 - \lambda_0)$, если $\lambda_0 < 1$;
- в) $p(x^+ / V_1) > \lambda_0 p(x^+ / V_2)$ или $p(x^+ / V_1) < \lambda_0 p(x^+ / V_2) + (1 - \lambda_0)$, если $\lambda_0 > 1$, где $\lambda_0 = P(V_2) / P(V_1)$ — отношение априорных вероятностей классов.

5.4.3. Непрерывный признак, нормально распределенный в классах

Покажем теперь, что условия леммы 5.1 могут выполняться не только для признаков, имеющих конечное число градаций, но и для непрерывных признаков.

Утверждение 5.6. Пусть

$$p(x / V_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left[-\frac{(x - M_1)^2}{2\sigma_1^2} \right], \quad (5.79)$$

$$p(x / V_2) = \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left[-\frac{(x - M_2)^2}{2\sigma_2^2} \right], \quad (5.80)$$

т. е. признак x имеет нормальные распределения в классах с математическими ожиданиями M_1, M_2 и среднеквадратическими отклонениями σ_1, σ_2 . Пусть априорные вероятности классов удовлетворяют условию $P(V_1) > P(V_2)$.

Тогда, если выполняются условия

$$\sigma_1 > \sigma_2 \quad (5.81)$$

и

$$\frac{\sigma_1}{\sigma_2} \exp \left[-\frac{(M_1 - M_2)^2}{2(\sigma_1^2 - \sigma_2^2)} \right] > \lambda_0, \quad (5.82)$$

где $\lambda_0 = P(V_2) / P(V_1)$ — отношение априорных вероятностей классов, то $P(E) = P_0$.

Доказательство. Отметим, что в силу (5.79) признак x имеет различные распределения в классах.

В соответствии с леммой 5.1 при $P(V_1) > P(V_2)$ признак x будет бесполезным в том и только в том случае, когда

$$p(x / V_1) > \lambda_0 p(x / V_2) \quad \forall x \in X. \quad (5.83)$$

Подстановка распределений из (5.79), (5.80) в (5.83) дает

$$\exp \left[\frac{(x - M_2)^2}{2\sigma_2^2} - \frac{(x - M_1)^2}{2\sigma_1^2} \right] > \lambda_0 \frac{\sigma_1}{\sigma_2} \quad \forall x \in X. \quad (5.84)$$

Логарифмируя левую и правую части (5.84), получим соотношение

$$\frac{(x - M_2)^2}{2\sigma_2^2} - \frac{(x - M_1)^2}{2\sigma_1^2} > \ln \left[\lambda_0 \frac{\sigma_1}{\sigma_2} \right] \quad \forall x \in X,$$

которое после очевидных преобразований можно представить в виде

$$[\sigma_1^2 - \sigma_2^2] x^2 + 2 (M_1 \sigma_2^2 - M_2 \sigma_1^2) x + \left[M_2^2 \sigma_1^2 - M_1^2 \sigma_2^2 - 2 \sigma_1^2 \sigma_2^2 \ln \left(\lambda_0 \frac{\sigma_1}{\sigma_2} \right) \right] > 0 \quad \forall x \in X. \quad (5.85)$$

Таким образом, в рассматриваемом случае условие бесполезности признака x сводится к условию положительности квадратного трехчлена, фигурирующего в левой части (5.85), при любом $x \in X$.

Понятно, что это возможно в том и только в том случае, когда параметры распределений $p(x/V_1)$ и $p(x/V_2)$ удовлетворяют условиям

$$\sigma_1^2 - \sigma_2^2 > 0, \quad (5.86)$$

$$4 (M_1 \sigma_2^2 - M_2 \sigma_1^2)^2 - 4 (\sigma_1^2 - \sigma_2^2) \left[M_2^2 \sigma_1^2 - M_1^2 \sigma_2^2 - 2 \sigma_1^2 \sigma_2^2 \ln \left(\lambda_0 \frac{\sigma_1}{\sigma_2} \right) \right] < 0. \quad (5.87)$$

Из (5.86) немедленно следует (5.81), а условие (5.87) после очевидных упрощений сводится к неравенству

$$\ln \left(\lambda_0 \frac{\sigma_2}{\sigma_1} \right) < - \frac{(M_1 - M_2)^2}{2(\sigma_1^2 - \sigma_2^2)}, \quad (5.88)$$

которое, с учетом того, что $\exp(\cdot) > 0$, эквивалентно условию (5.82).

Из условий утверждения 5.6 вытекает следствие.

Следствие 5.2. При выполнении условия $P(V_1) > P(V_2)$ релевантный признак, имеющий нормальные распределения в классах с одинаковыми математическими ожиданиями, бесполезен, если отношение среднеквадратических отклонений превышает отношение соответствующих априорных вероятностей классов, т. е.

$$\frac{\sigma_2}{\sigma_1} > \frac{P(V_2)}{P(V_1)}.$$

Следствие прямо вытекает из (5.82), если положить $M_1 = M_2$.

Отметим также, что при $\sigma_1 = \sigma_2$ квадратный трехчлен (5.85) переходит в линейную функцию x , которая не может быть положительна для любого $x \in X$, если $M_1 \neq M_2$. Отсюда немедленно вытекает важное для практического использования следствие.

Следствие 5.3. Всякий диагностический признак, имеющий нормальные распределения в классах с различными математическими ожиданиями и равными дисперсиями, полезен в смысле строгого неравенства $P(E) < P_0$.

Однако при $\sigma_1 \neq \sigma_2$ нельзя однозначно сделать вывод о полезности признака, нормально распределенного в классах, даже если $M_1 \neq M_2$.

Полученные результаты еще раз свидетельствуют о необходимости использования предложенной в пункте 5.2.3 процедуры усиления традиционного ROC-анализа для оценки диагностической эффективности признаков.

Для практического использования представляют интерес также оценка полезности признака x_n , $1 \leq n \leq N$, в совокупности с другими признаками N -мерного описания $x^{(N)} = (x_1, \dots, x_N)$. Если потери от правильных решений допустимо принять равными нулю, а ошибочных — единице, то для оценки полезности признака достаточно убедиться в том, что исключение признака из описания $x^{(N)}$ увеличивает среднюю вероятность ошибки распознавания. Иными словами, средние вероятности ошибочных решений $P_N(E)$ и $P_{N-1}(E)$, принимаемых соответственно по исходному $x^{(N)}$ и сокращенному $x^{(N-1)}$ описаниям, удовлетворяют строгому неравенству

$$P_N(E) < P_{N-1}(E). \quad (5.89)$$

В работе [218] доказаны теоремы, из которых следует, что если многомерные распределения $p(x^{(N)} / V_1)$ и $p(x^{(N)} / V_2)$ нормальны, то для выполнения строгого неравенства (5.89) достаточно, чтобы:

а) признак x_n не был коррелирован ни с одним из других $N - 1$ признаков в одном из классов и коррелирован хотя бы с одним из признаков в другом классе;

б) математическое ожидание и дисперсия этого признака были одинаковыми в классах, т. е.

$$M\{x_n / V_1\} = M\{x_n / V_2\}, \quad (5.90)$$

$$D\{x_n / V_1\} = D\{x_n / V_2\}. \quad (5.91)$$

Отметим, что признак, удовлетворяющий условиям (5.90) и (5.91), имеет одинаковые распределения в классах, т. е. является нерелевантным по определению М. Бен-Бассата. Однако за счет корреляционной связи с другими признаками он оказывается полезным в описании $x^{(N)} = (x_1, \dots, x_N)$.

Рассмотрим теперь практическое применение теоретических результатов на примерах синтеза двух классов прикладных ИТ оценки состояния объектов различной физической природы.

Первый класс ориентирован на оценку **количественных** характеристик текущего состояния объекта по сигналу с локально сосредоточенными признаками. В качестве примера рассмотрим задачу оперативного контроля технологических параметров литейного процесса — химического состава и механических свойств металла. Исходной информацией служит сигнал, несущий сведения об изменении температуры пробы металла в процессе кристаллизации.

Второй класс информационных технологий, обеспечивающих оценку **качественных** показателей текущего состояния объектов, представим на примерах обработки физиологических сигналов с локально сосредоточенными признаками — электрокардиограмм (ЭКГ) и магнитокардиограмм (МКГ).

Покажем, что компьютерная обработка в фазовом пространстве координат искаженных сигналов различной физической природы позволяет не только повысить эффективность оценки полезного сигнала, но и упростить задачу выделения информативных фрагментов сигнала, на которых сосредоточены диагностические признаки.

Важно отметить, что все рассматриваемые технологии разрабатывались по обобщенной схеме формирования информационного продукта с использованием инструментальной системы СИДИГРАФ.

6.1. Информационная технология «ТЕРМОГРАФ»

6.1.1. Общая характеристика информационной технологии «ТЕРМОГРАФ»

Проблема получения оперативной и достоверной информации о ходе протекания технологического процесса по-прежнему актуальна в металлургическом и литейном производстве. Высокая температура и агрессивность среды препятствуют созданию датчиков непосредственного контроля, а данные химической лаборатории и механических испытаний приходят с большим запаздыванием и не могут использоваться для управления быстропротекающими процессами доводки металла по ходу плавки.

В свое время нами была впервые поставлена и решена задача построения компьютерной технологии для оценки содержания углерода в сталеплавильной ванне по термограмме кристаллизации пробы жидкого металла [221, 223, 282]. Эта технология использовала идеи статистического распознавания образов, которые были реализованы в различных модификациях цифровых анализаторов «УГЛЕРОД» [181].

Впоследствии оказалось, что термограмма кристаллизации пробы жидкого металла несет гораздо больше информации и позволяет косвенно оценить не только содержание углерода, но и целый комплекс иных технологических параметров.

Для практической реализации такой возможности создана новая информационная технология «ТЕРМОГРАФ»¹ [177].

Главная задача технологии — последовательное преобразование исходного «сырья» (термограмм кристаллизации пробы жидкого чугуна) в требуемый информационный «продукт» — количественную оценку семи параметров химического состава, а также прогноз механических свойств выплавляемого металла — прочности на растяжение и твердости (рис. 6.1, *а*). Структура компьютерной системы для реализации ИТ построена в виде трех взаимосвязанных подсистем, объединенных общей базой данных (рис. 6.1, *б*).

¹ Программная реализация технологии выполнена Т.П. Потаповой.

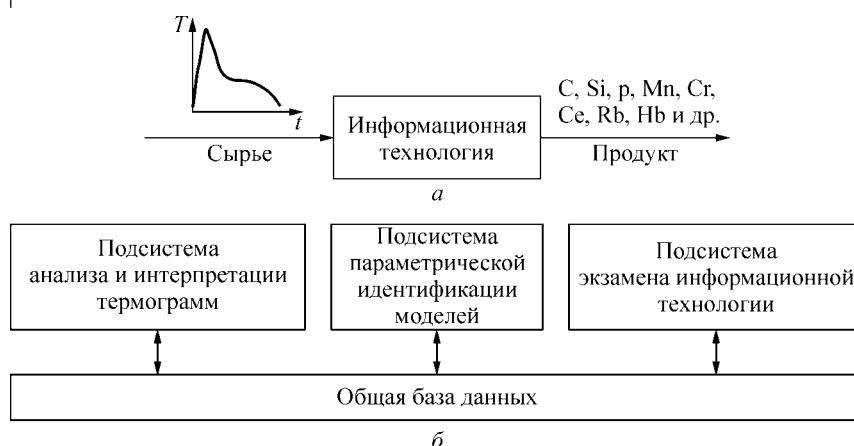


Рис. 6.1. Компьютерная система «ТЕРМОГРАФ»:

а — задача информационной технологии; б — упрощенная структура системы

Первая подсистема обеспечивает процесс обработки термограмм и реализацию математических моделей, построенных для конкретных условий эксплуатации ИТ. На вход подсистемы поступают сигналы $T(t)$, несущие информацию о температуре охлаждающейся пробы расплава. Сигнал формируется измерительной установкой, состоящей из огнеупорной пробницы, в которой происходит охлаждение и кристаллизация пробы расплава, и бесконтактного датчика температуры².

Вторая подсистема обеспечивает построение регрессионных моделей, связывающих технологические параметры расплава с локально сосредоточенными признаками термограмм, о которых пойдет речь далее.

Третья подсистема служит для оценки качества функционирования ИТ на основе сравнения результатов косвенной оценки технологических параметров с данными контрольных анализов.

6.1.2. Схема формирования информационного продукта

Подсистема анализа и интерпретации термограмм является главной подсистемой технологии. Рассмотрим кратко архитектуру этой подсистемы (рис. 6.2).

² Измерительная установка разработана НПО НИИТракторосельхозмаш (г. Москва).

Исходные данные, несущие информацию о температуре расплава в процессе охлаждения (термограмме $T(t)$), вводятся в подсистему непосредственно от измерительной установки с помощью аналого-цифрового преобразователя (АЦП) либо из предварительно созданных файлов, хранимых на диске. В таких файлах хранятся последовательности дискретных значений температуры $T_1, \dots, T_K$, $K = 400$, зафиксированных через равные промежутки времени $\Delta t = 1$ с.

Поскольку реальные термограммы (рис. 6.3, а) подвержены влиянию внутренних и внешних возмущений, первый этап обработки сводится к восстановлению полезного сигнала по искаженным реализациям.

Для фильтрации внешних помех используются вычислительные процедуры, описанные в гл. 4, которые обеспечивают подавление импульсных и ограниченных по уровню случайных помех (рис. 6.3, б).



Рис. 6.2. Архитектура подсистемы анализа и интерпретации термограмм

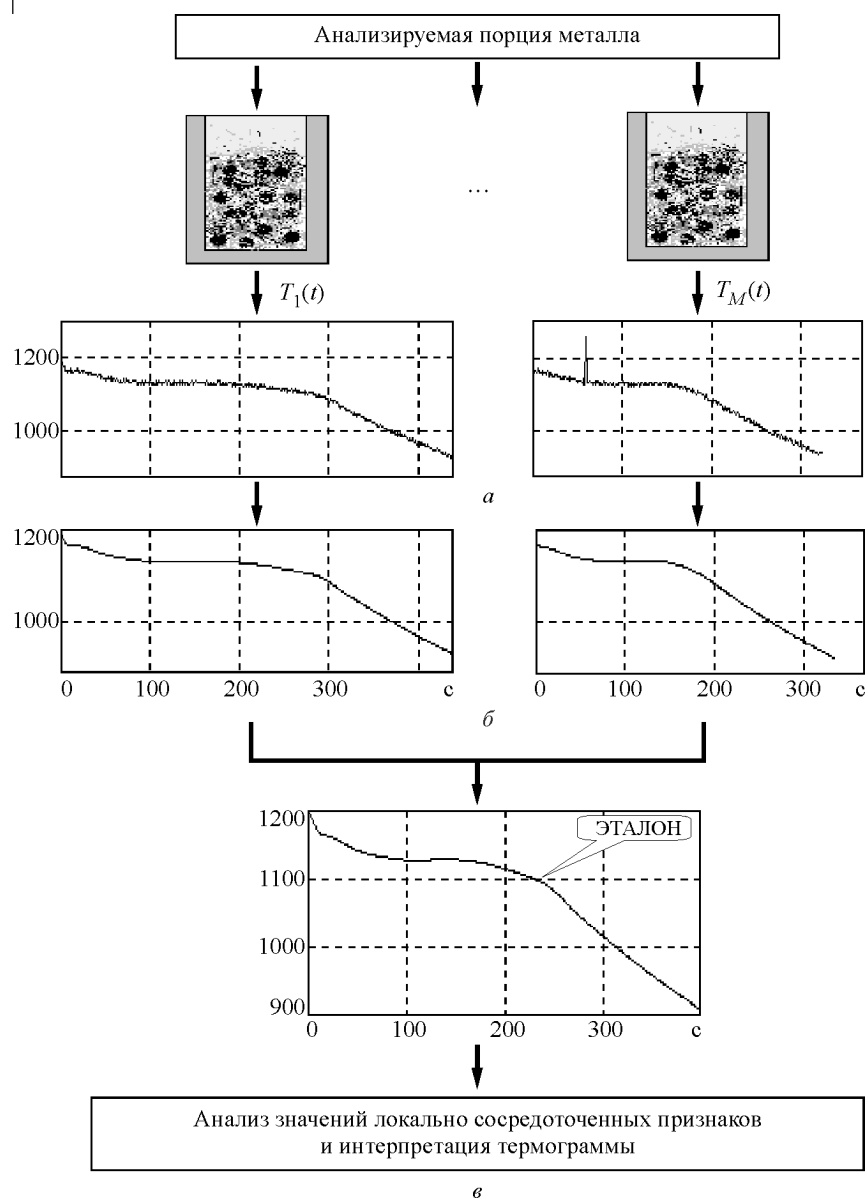


Рис. 6.3. Схема восстановления эталона термограммы:

a — термограммы, полученные при кристаллизации одной порции расплава; *б* — фильтрованные термограммы; *в* — эталонная термограмма, полученная после усреднения реализаций в фазовом пространстве

Для оценки эталонной термограммы применяется метод усреднения искаженных реализаций сигналов в фазовом пространстве координат. Используется совокупность $T_1(\cdot), \dots, T_M(\cdot)$ реализаций сигнала, которые регистрируются при фиксированном состоянии металла. Для этого одна порция расплава разливается в несколько пробниц и параллельно регистрируется процесс кристаллизации металла в пробах. В результате удается получить приемлемую оценку эталона термограммы, пригодную для дальнейшей интерпретации (рис. 6.3, в).

Диагностические признаки термограммы сосредоточены на ее фрагментах, отображающих отдельные стадии процесса кристаллизации (см. рис. 2.10). Оказалось, что компьютерная обработка наблюдений в фазовом пространстве позволяет не только повысить эффективность оценки полезного сигнала, но и упростить задачу выделения таких информативных фрагментов.

На рис. 6.4 показана термограмма кристаллизации пробы серого чугуна, полученная путем усреднения искаженных реализаций, первая производная усредненной термограммы и ее фазовая траектория. Легко видеть, что границы фрагментов (черные точки) более выразительны на графиках первой производной (рис. 6.4, б) и фазовой траектории (рис. 6.4, в), чем на самой термограмме (рис. 6.4, а).

Этот факт позволил существенно упростить и повысить эффективность процедур автоматического поиска границ информативных фрагментов термограммы.

Подсистема параметрической идентификации обеспечивает построение регрессионных моделей, связывающих технологические параметры чугуна с амплитудно-временными параметрами (признаками) наблюдаемой термограммы. Чтобы повысить точность оценки технологических параметров, модели строят для конкретных условий эксплуатации по конечному числу наблюдений с известными значениями технологических параметров.

Точки $A, \dots, J$ (рис. 6.4, з) характеризуют границы отдельных стадий процесса кристаллизации расплава — дендритной кристаллизации, кристаллизации эвтектики и т. д. В информационной технологии «ТЕРМОГРАФ» реализованы вычислительные процедуры, позволяющие определить для каждой из указанных характерных точек тройку величин $t_\mu, T_\mu, \dot{T}_\mu$, ($\mu = A, \dots, J$), представляющих собой момент времени t_μ , при котором наблюдается соответствующая точка, а также значение температуры T_μ и оценка ее первой производной $\dot{T}_\mu$ в этот же момент времени.

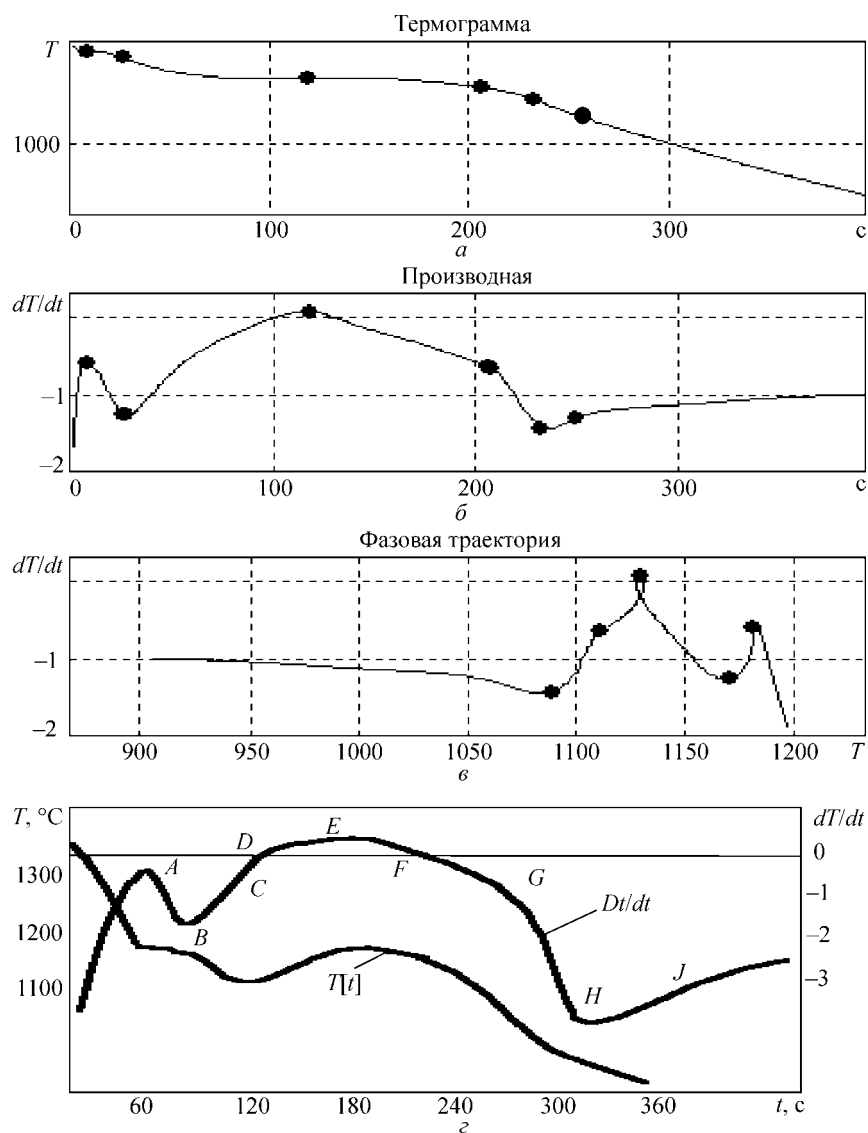


Рис. 6.4. Диагностические признаки термограммы:

a — термограмма; *б* — производная термограммы; *в* — фазовая траектория термограммы; *г* — границы информативных фрагментов

ТАБЛИЦА 6.1

Аргументы математических моделей

Признаки	Характерные точки термограммы								
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>J</i>
t_{μ} , с	22	43	101	159	178	209	267	286	307
T_{μ} , °C	1200	1186	1138	1116	1119	1125	1108	1083	1050
$\dot{T}_{\mu}$, °C/с	-0,20	-1,11	-0,59	0,00	0,27	0,00	-0,75	-2,02	-1,30

Тем самым осуществляется переход от исходного пространства наблюдений размерности $K_1 = 400$ к пространству аргументов математических моделей (потенциально полезным признакам) размерности $K_2 = 27$ (табл. 6.1).

Для определения технологических параметров по признакам термограммы используются модели в виде множественных уравнений регрессии:

$$s_j = K_j^T F_j(x), \quad (6.1)$$

где s_j — оценка j -го технологического параметра; $K_j = (K_{1j}, \dots, K_{Lj})$ — L -мерный вектор коэффициентов регрессии; $F_j(x) = (F_{1j}(x), \dots, F_{Lj}(x))$ — заданная вектор-функция; $x = (x_1, \dots, x_N)$ — N -мерный вектор, компоненты которого представляют собой значения указанных выше локально сосредоточенных признаков термограммы; T — знак транспонирования.

В общем случае задачу построения моделей (6.1) можно понимать и как задачу структурной идентификации, т. е. выбор наилучших базисных функций $F_{1j}(x), \dots, F_{Lj}(x)$, и как задачу параметрической идентификации, т. е. определение оптимальных значений параметров $K_{1j}, \dots, K_{Lj}$ моделей заданной структуры.

В реализованной версии ИТ решалась лишь задача параметрической идентификации. При этом использованы структуры математических моделей, предложенные в работе [307].

Первая задача подсистемы идентификации — формирование и накопление в БД значений регрессоров математических моделей, полученных в результате автоматической обработки обучающей выборки термограмм. Пользователь в режиме диалога имеет воз-

можность ввести в БД истинные значения технологических параметров соответствующих проб металла по данным химической лаборатории и лаборатории механических испытаний.

Статистическая обработка полученных данных сводится к построению уравнений множественных регрессий для каждого технологического параметра на основе использования МНК. Для решения систем нормальных уравнений применяется алгоритм обращения матриц методом исключения.

Модели, построенные по выборке $W \approx 100$ термограмм для условий Купянского литейного завода, имеют вид [207]:

содержание углерода

$$C = 9,84918 - 0,00413T_A - 0,00147T_D \quad (\sigma \approx 0,06); \quad (6.2)$$

содержание кремния

$$Si = -23,99104 + 0,00572T_D + 0,01762T_F + 31,4176 \frac{\dot{T}_F - \dot{T}_G}{t_F - t_G} \quad (\sigma \approx 0,18); \quad (6.3)$$

содержание хрома

$$\begin{aligned} Cr = & 2,6552 - 0,00572T_C + 0,00032(t_F - t_D) + 0,00109T_H - \\ & - 0,0258 \dot{T}_A + 0,004705 \frac{\dot{T}_C - \dot{T}_D}{t_C - t_D} - 3,07313 \frac{\dot{T}_F - \dot{T}_G}{t_F - t_G} \quad (\sigma \approx 0,0194); \quad (6.4) \end{aligned}$$

содержание марганца

$$\begin{aligned} Mn = & -1,1838 + 0,00186T_H + 0,00065(t_F - t_D) + \\ & + 0,09051\dot{T}_E + 0,95778 \frac{\dot{T}_A - \dot{T}_B}{t_A - t_B} \quad (\sigma \approx 0,076); \quad (6.5) \end{aligned}$$

содержание меди

$$\begin{aligned} Cu = & 0,32656 + 0,00088(T_A - T_F) + 0,00019(t_F - t_D) + \\ & + 0,01562T_B - 0,00061(t_E - t_A) \quad (\sigma \approx 0,04); \quad (6.6) \end{aligned}$$

содержание фосфора

$$\begin{aligned} P = & 0,38846 - 0,0006T_C + 0,00012(t_F - t_D) + 0,00034T_H - \\ & - 0,00234\dot{T}_A + 0,10638 \frac{\dot{T}_C - \dot{T}_D}{t_C - t_D} + 1,25 \frac{\dot{T}_F - \dot{T}_G}{t_F - t_G} \quad (\sigma \approx 0,0053); \quad (6.7) \end{aligned}$$

углеродный эквивалент

$$C_E = 9,67514 - 0,00758T_A + 0,00285T_D \quad (\sigma \approx 0,045); \quad (6.8)$$

коэффициент эвтектического насыщения

$$S_C = 2,56022 - 0,00199T_A + 0,00063T_D \quad (\sigma \approx 0,012); \quad (6.9)$$

прочность на растяжение

$$R_M = 30,97892 + 16,33124 \frac{\dot{T}_C - \dot{T}_D}{t_C - t_D} - 0,2572(T_A - T_F) + \\ + 0,00257(T_A - T_F)^2 + 0,66383\dot{T}_B \quad (\sigma \approx 1,95); \quad (6.10)$$

твёрдость

$$H_b = 226,48464 + 0,06006(T_A - T_F) + 12,43406\dot{T}_B - \\ - 14,93054\dot{T}_A \quad (\sigma \approx 13,33), \quad (6.11)$$

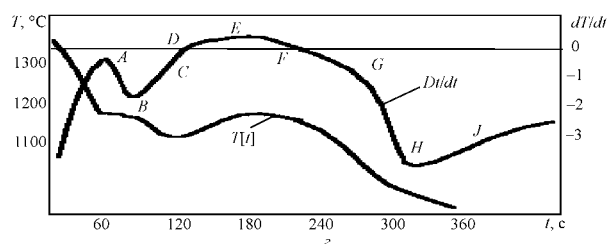
где σ — среднеквадратическое отклонение модельных и точных значений соответствующих параметров для наблюдений экзаменационной выборки, в которой было использовано 20 наблюдений, не участвовавших в построении моделей.

Результаты основных этапов обработки термограмм отображаются на экране монитора и принтере в виде графической и текстовой информации. Пользователю в удобной форме предоставляется только та информация, которая ему необходима для управления технологическим процессом выплавки металла (рис. 6.5).

Потребительские свойства ИТ «ТЕРМОГРАФ» главным образом определяет точность оценки технологических параметров. Поэтому периодически в процессе эксплуатации необходимо проводить сравнение результатов косвенной оценки технологических показателей с истинными значениями, полученными по результатам контрольных анализов. Для решения такой задачи в информационной технологии реализована подсистема экзамена, которая предоставляет пользователю наглядную информацию о точности оценки каждого j -го технологического параметра s_j в текстовом и графическом виде.

На вход подсистемы поступают результаты косвенной оценки технологических параметров, формируемые ИТ, которые со-

Замер № 5/15.09.89



Интегральные характеристики термограмм

Температура ликвидуса	1200,0 °C
Температура солидуса	1116,6 °C
Интервал кристаллизации	84,0 °C
Время кристаллизации	264 с

Технологические параметры

А. Химический состав расплава

Содержание углерода (C)	3,25 %
Содержание кремния (Si)	1,81 %
Содержание хрома (Cr)	0,24 %
Содержание марганца (Mn)	0,85 %
Содержание меди (Cu)	0,30 %
Содержание фосфора (P)	0,07 %
Углеродный эквивалент (C_E)	3,76 %
Коэффициент эвтектического насыщения (S_C)	0,87

Б. Механические свойства

Прочность на растяжение (R_M)	256,0 мПа
Твердость (H_b)	220,2 H_b

Рис. 6.5. Форма выходного документа ИТ «ТЕРМОГРАФ»

храняются в базе данных (БД). Структура БД позволяет после выполнения контрольных анализов ввести в соответствующие поля записей истинные значения технологических параметров (рис. 6.6).



Рис. 6.6. Архитектура подсистемы экзамена ИТ

На основании сформированных данных пользователю предоставляется информация в виде:

- таблиц расхождений $\Delta_j = \hat{s}_j - s_j$ результатов оценки каждого j -го технологического параметра с указанием номера замера, даты и времени измерения;
- когнитивного графического образа (скатерограмма отклонений по каждому из технологических параметров);
- интегральных оценок качества функционирования ИТ, в том числе статистических оценок расхождений результатов;
- оценки вероятности (процента) попадания точек экзаменационной выборки в задаваемый пользователем допуск.

Скатерограмма отклонений $\Delta_j = \hat{s}_j - s_j$ дает наглядное графическое представление об адекватности математических моделей результатам контрольных анализов.

На экране монитора формируется квадратное графическое окно с координатными осями $\hat{s}_j, s_j$, где $\hat{s}_j$ — численное значение j -го технологического параметра, вычисленное по соответствующей математической модели, а s_j — численное значение этого же параметра по результатам контрольного анализа.

Если экспериментальные точки симметрично располагаются относительно прямой $s_j = \hat{s}_j$, проходящей под углом 45° к коор-

динатным осям (диагонали графического окна), то это свидетельствует об адекватности математической модели, причем мерой адекватности служит степень разброса точек относительно указанной прямой.

В диалоге с системой пользователь может задать требуемое значение допуска Δ_j на отклонения результатов косвенной оценки данного технологического параметра от истинных значений, и в графическом окне автоматически выделяется область значений $|\hat{s}_j - s_j| \leq \Delta_j$, соответствующих заданному допуску с учетом принятого масштаба, а также вычисляется процент попадания точек экзаменационной выборки в эту область.

6.1.3. Аппаратная реализация упрощенных процедур обработки термограмм

Для построения «интеллектуального» датчика системы управления доменным процессом и литейным агрегатом можно несколько упростить задачу обработки термограмм, ограничившись контролем только трех основных технологических параметров — содержания углерода C, кремния Si и значения углеродного эквивалента C_E .

В первом приближении можно считать, что C_E описывается зависимостью $C_E = C + 1/3 \text{ Si}$. С учетом этого факта в литейном производстве был распространен метод оценки технологических показателей, использующий упрощенные модели вида [1]:

$$C_E = a_0 + a_1 T_L, \quad (6.12)$$

$$C = b_0 + b_1 (T_L - T_S), \quad (6.13)$$

$$\text{Si} = 3(C_E - C), \quad (6.14)$$

где T_L и T_S — температуры ликвидуса и солидуса расплава.

Отсюда следует, что для автоматической оценки C_E , C и Si достаточно определить всего два локально сосредоточенных параметра термограммы — температуры начала T_L и окончания T_S кристаллизации.

Для реализации такого подхода создана система «ЧУГУН», опытный образец которой апробирован еще в 1983 г. Он позво-

лил впервые в отечественной практике автоматизировать процесс косвенной оценки химического состава жидкого чугуна по сигналу, несущему информацию о температуре $T(t)$ при кристаллизации его пробы.

Ввод информации осуществляется посредством оригинального АЦП сигналов низкого уровня [182]. В отличие от традиционных подходов к построению АЦП сигналов низкого уровня, предусматривающих усиление преобразуемого сигнала $y(t)$ до стандартного уровня с помощью нормирующих усилителей, схема предложенного следящего АЦП основана на снижении уровня компенсирующего сигнала $y_k(t)$ до пределов, при которых возможна непосредственная компенсация преобразуемого сигнала $y(t)$.

Предлагаемый подход к преобразованию аналоговых сигналов в цифровой эквивалент имеет ряд существенных преимуществ по сравнению с традиционной схемой преобразования сигналов низкого уровня [217].

При традиционном построении следящего АЦП (рис. 6.7, а) сигнал $y_A(t)$, поступающий на вход компаратора, связан с преобразуемым аналоговым сигналом низкого уровня $y(t)$ соотношением

$$y_A(t) = K_A y(t), \quad (6.15)$$

где K_A — коэффициент передачи напряжения нормирующего преобразователя.

Компенсирующий аналоговый сигнал высокого уровня $y_k(t)$, поступающий в режиме следящего преобразования, связан с выходным цифровым кодом $N(t)$ соотношением

$$y_k(t) = K_{\text{ЦАП}} N(t), \quad (6.16)$$

где $K_{\text{ЦАП}}$ — коэффициент передачи цифро-аналогового преобразователя (ЦАП).

Из выражений (6.15) и (6.16) следует, что в установившемся режиме, когда $y_A(t) \approx y_k(t)$, цифровой эквивалент $N(t)$ равен

$$N(t) = K_A K_{\text{ЦАП}}^{-1} y(t). \quad (6.17)$$

Видно, что погрешность преобразования в цифровую форму сигнала низкого уровня $y(t)$ определяется как стабильностью ко-

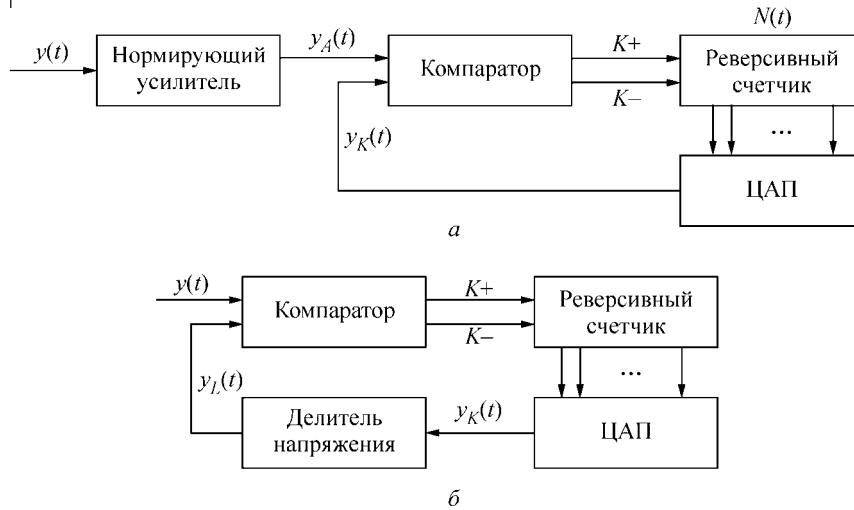


Рис. 6.7. Упрощенные схемы построения следящего АЦП сигналов низкого уровня:

a — традиционный подход; *б* — предлагаемый подход

эфициента $K_{\text{ЦАП}}$, так и стабильностью коэффициента передачи K_A нормирующего преобразователя.

Основное преимущество предлагаемой схемы построения АЦП (рис. 6.7, *б*) состоит в том, что цифровой код $N(t)$ на выходе преобразователя практически не зависит от коэффициента передачи K_A .

В самом деле, компенсирующий сигнал $y_L(t)$ низкого уровня, поступающий на вход компаратора, связан с сигналом $y_K(t)$ высокого уровня на выходе цифро-аналогового преобразователя соотношением

$$y_L(t) = y_K(t)K_D^{-1}, \quad (6.18)$$

где K_D — коэффициент передачи пассивного делителя напряжения, обеспечивающего снижение уровня компенсирующего сигнала до требуемых пределов.

Отсюда, с учетом (6.16), следует, что в установившемся режиме, когда происходит компенсация входного сигнала на низком уровне, т. е. $y_L(t) \approx y(t)$, выходной код $N(t)$ определяется

соотношением

$$N(t) = K_D K_{\text{ЦАП}}^{-1} y(t). \quad (6.19)$$

Сравнение выражений (6.17) и (6.19) показывает, что погрешность преобразования по предлагаемой схеме, в отличие от традиционной, не зависит от коэффициента передачи K_A нормирующего преобразователя, а зависит от стабильности коэффициента передачи K_D делителя напряжения. Поскольку же делитель представляет собой пассивную схему (в отличие от усилителя, построенного на активных элементах), то при прочих равных условиях предлагаемый подход позволяет получать более высокую точность преобразования, нежели традиционный.

К тому же погрешность преобразования по предлагаемой схеме в меньшей степени подвержена влиянию изменений сопротивления внешней электрической цепи и внутреннего сопротивления источника сигнала, поскольку величина тока, протекающего во внешней цепи, определяется не уровнем сигнала, а разностью величин преобразуемого и компенсирующего сигналов, которая в установившемся режиме не превышает шаг квантования. Это преимущество особенно актуально при работе с источниками сигнала с высоким внутренним сопротивлением, а также при преобразовании сигналов, передаваемых от удаленных источников.

Испытания преобразователя в условиях доменного цеха Оздского металлургического завода (Венгерская Республика) подтвердили его эффективность: погрешность преобразования не превышала 4 мкВ в диапазоне изменения сигнала 0—4 мВ.

Использование следящего АЦП для преобразования сигнала $T(t)$ в цифровую форму, который формирует не только параллельный код $N(T)$, но и унитарный (число-импульсный) код K_+, K_- , позволяет реализовать упрощенные процедуры обработки сигнала $T(t)$ аппаратным способом.

Основная идея, положенная в основу построения аппаратной обработки $T(t)$, в том числе процедур автоматического обнаружения на фоне случайных помех $h(t) \leq h_0$ фрагментов сигнала, соответствующих температурам T_L и T_S , состоит во взаимодействии счетчиков импульсов. Один из счетчиков переполняется, когда локальные приращения сигнала превышают заданный по-

рог по уровню h_0 , а второй — контролирует случайные интервалы времени $\Delta_\mu = t_\mu - t_{\mu-1}$, $\mu = 1, 2, \dots$, между моментами переполнения первого счетчика. Анализ случайных интервалов $\Delta_\mu = t_\mu - t_{\mu-1}$ позволяет в реальном масштабе времени обнаружить фрагменты сигнала заданной скорости и оценить их продолжительность во времени.

Варианты построения устройств, обеспечивающих обработку сигналов по такой схеме, описаны в работах [6, 7, 199—201, 232].

6.1.4. Модифицированная процедура параметрической идентификации моделей

Для градуировки системы «ЧУГУН» использовалась модифицированная процедура метода наименьших квадратов (МНК), позволяющая определить оптимальные значения параметров a_0 , a_1, b_0, b_1 линейных уравнений регрессии (6.12), (6.13) по показаниям предварительно не настроенной системы.

Продемонстрируем идею модификации МНК на примере определения параметров регрессионной модели (6.12) для оценки значения углеродного эквивалента C_E .

Традиционная процедура параметрической идентификации предполагает обработку экспериментальной выборки, представляющей собой совокупность наблюдений регрессора $T_L[1], \dots, T_L[G]$ и соответствующих им значений отклика $C_E[1], \dots, C_E[G]$. При этом оптимальные (по МНК) значения параметров a_0, a_1 определяются известными соотношениями

$$a_0 = \frac{\hat{M}\{C_E T_L^2\} - \hat{M}\{T_L\}\hat{M}\{T_L\}}{\hat{M}\{T_L^2\} - \hat{M}\{T_L\}\hat{M}\{T_L\}}, \quad (6.20)$$

$$a_1 = \frac{\hat{M}\{C_E T_L\} - \hat{M}\{C_E\}\hat{M}\{T_L\}}{\hat{M}\{T_L^2\} - \hat{M}\{T_L\}\hat{M}\{T_L\}}, \quad (6.21)$$

где $\hat{M}\{\eta\}$ — среднее значение величины η , найденное по G наблюдениям:

$$\hat{M}\{\eta\} = \frac{1}{G} \sum_{i=1}^G \eta[i]. \quad (6.22)$$

В системе «ЧУГУН» цифровой код температур $T_L[i]$, образуемый в процессе аппаратной обработки $T(t)$, представляет собой всего лишь промежуточную переменную, используемую для вычисления конечного результата $C_E^*[i]$ по формуле

$$C_E^*[i] = a_0^* + a_1^* T_L[i], \quad (6.23)$$

где a_0^*, a_1^* — настраиваемые параметры системы.

Очевидно, что минимум среднего квадрата отклонений показаний системы $C_E^*[i]$ от результатов контрольных анализов $C_E[i]$ будет обеспечен лишь в том случае, когда настраиваемые параметры a_0^*, a_1^* равны оптимальным значениям a_0, a_1 , определяемым соотношениями (6.20), (6.21), т. е.

$$a_0^* = a_0, \quad a_1^* = a_1. \quad (6.24)$$

В условиях отсутствия информации о температурах $T_L[i]$, с одной стороны, невозможно непосредственно воспользоваться формулами (6.20), (6.21) для вычисления оптимальных значений параметров, с другой — не имеет смысла снабжать систему «ЧУГУН» дополнительным средством отображения температур $T_L[i]$ только лишь для того, чтобы провести ее градуировку.

В связи с этим возникает вопрос: нельзя ли использовать показания самой системы, пусть даже вначале неверные, с тем чтобы провести ее оптимальную настройку.

Покажем, что такая возможность действительно существует [219]. Для этого предположим, что на этапе градуировки в системе установлены произвольные значения a_0^*, a_1^* , после чего проведено G измерений. Разумеется, если установленные значения a_0^*, a_1^* отличны от оптимальных значений a_0, a_1 , т. е. не выполняется условие (6.24), то полученные результаты $C_E^*[1], \dots, C_E^*[G]$ могут существенно отличаться от истинных результатов химического анализа $C_E[1], \dots, C_E[G]$.

Между тем, как видно из соотношения (6.23), на каждом i -м измерении результат $C_E^*[i]$ однозначно определяет величину $T_L[i]$, которая приводит к такому результату:

$$T_L[i] = \frac{C_E^*[i] - a_0^*}{a_1^*}, \quad i = 1, \dots, G. \quad (6.25)$$

На основании выражения (6.25) формулы (6.20), (6.21), определяющие оптимальные значения a_0 и a_1 , после очевидных преобразований могут быть модифицированы следующим образом:

$$a_0 = \frac{a_0^*(\hat{M}\{C_E C_E^*\} - \hat{M}\{C_E\}\hat{M}\{C_E^*\}) - \hat{M}\{C_E^*\}\hat{M}\{C_E C_E^*\} - \hat{M}\{C_E\}\hat{M}\{(C_E^*)^2\}}{\hat{M}\{(C_E^*)^2\} - \hat{M}\{C_E^*\}\hat{M}\{C_E^*\}}, \quad (6.26)$$

$$a_1 = \frac{a_1^*(\hat{M}\{C_E C_E^*\} - \hat{M}\{C_E\}\hat{M}\{C_E^*\})}{\hat{M}\{(C_E^*)^2\} - \hat{M}\{C_E^*\}\hat{M}\{C_E^*\}}, \quad (6.27)$$

где $\hat{M}\{\eta\}$ — среднее значение величины η , определяемое по формуле (6.22).

Следовательно, если на этапе градуировки фиксировать показания системы $C_E^*[1], \dots, C_E^*[G]$, полученные при произвольных значениях настраиваемых параметров a_0^*, a_1^* , и определить по результатам контрольного химического анализа истинные значения $C_E[1], \dots, C_E[G]$ в соответствующих пробах металла, то такой информации оказывается вполне достаточно для вычисления по формулам (6.26), (6.27) оптимальных значений параметров a_0, a_1 .

Аналогичным образом проводится параметрическая идентификация модели (6.13).

6.1.5. Результаты промышленных испытаний

Промышленные испытания системы «ЧУГУН» проводились на Московском чугунолитейном заводе «Станколит». В табл. 6.2 приведены абсолютные отклонения Δ_j прогнозных оценок от данных поверочного анализа (экзаменационная выборка включала в себя более 100 наблюдений). Точность оценки технологических параметров вполне приемлема для управления процессом выплавки литейного чугуна [1].

После успешных испытаний ИТ «ТЕРМОГРАФ» внедрена в НПО НИИТракторосельхозмаш (г. Москва). Система использовалась для оценки комплекса технологических параметров чугу-

ТАБЛИЦА 6.2

Результаты экзамена системы «ЧУГУН»

Параметр	Абсолютная погрешность, %	Процент случаев
C_E	<0,08	79
	>0,08	21
C	<0,05	87
	0,05—0,1	10
	>0,1	3
Si	<0,1	42
	0,1—0,15	21
	>0,15	37

на по локально сосредоточенным признакам термограммы в соответствии с регрессионными моделями (6.2)—(6.11).

Основное преимущество ИТ «ТЕРМОГРАФ» и «ЧУГУН» — оперативность получения информации, необходимой для управления процессом выплавки металла. Оценка технологических показателей осуществляется за 5—6 минут от момента отбора пробы, тогда как данные о химическом анализе проб металла в твердом виде поступают не ранее чем через 1—2 часа от момента отбора пробы, а результаты механических испытаний — через сутки. Поэтому разработанные системы могут исполнять роль «интеллектуальных» датчиков при управлении литейным и доменным процессами [216].

6.2. Информационная технология «FAZAWIN»

6.2.1. Общая характеристика информационной технологии «FAZAWIN»

При инфекционных воспалительных заболеваниях, в том числе ревматоидном артрите, возникают поражения сердечной мышцы. Это обусловлено прямым воздействием на миокард инфекционного агента и его токсинов, повреждающим действием антител, возникающих как ответ организма на инфекцию, а также расстройством нейрогуморальной регуляции деятельности сердца [143].

Следует отметить, что на ранних стадиях заболевания не наблюдается каких-либо заметных изменений ЭКГ. Иначе говоря, традиционный амплитудно-временной анализ ЭКГ больных ревматоидным артритом не достаточно чувствителен и специфичен.

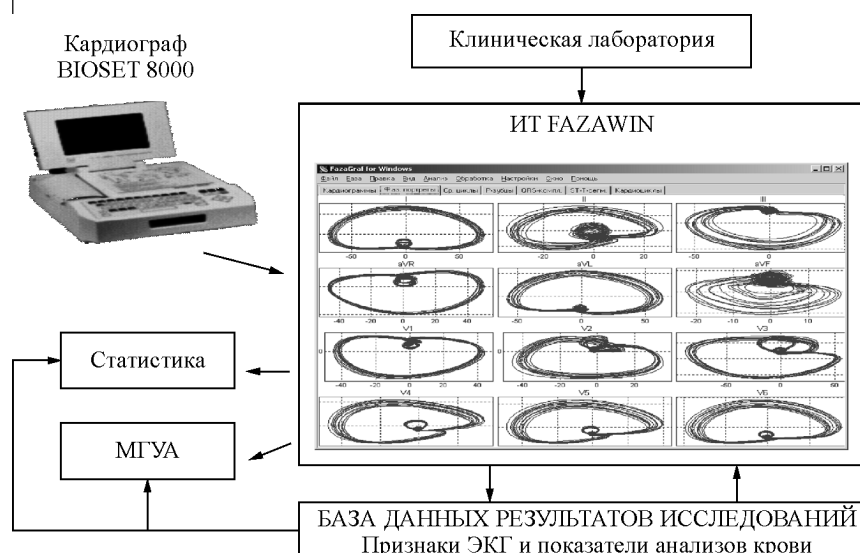


Рис. 6.8. Упрощенная структура компьютерной системы «FAZAWIN»

Для исследования возможностей повышения достоверности результатов диагностики поражений миокарда у больных ревматоидным артритом создана компьютерная система «FAZAWIN»³ (рис. 6.8), реализующая новую ИТ обработки ЭКГ в фазовом пространстве координат [143].

Исходные ЭКГ регистрировались электрокардиографом BIOSET 8102 MI фирмы Höpman (Германия). Кардиограф, подключающийся к компьютеру с помощью стандартного интерфейса RS-232, позволяет оцифровывать одновременно до 12 отведений через каждые 2 мс с разрешением в 12 цифровых разрядов. Взаимодействие между кардиографом и компьютером осуществляется в соответствии со стандартом TC251.

Передача данных в компьютер осуществляется с применением байт-ориентированного протокола канального уровня. Этот протокол обеспечивает достоверный транспорт данных за счет защиты блоков контрольными циклическими кодами (CRC) и повторными запросами в случае обнаружения ошибки.

³ Разработка проводилась в рамках международного контракта с фирмой LUM GmbH (Берлин) по программе «TRANSFORM». Реализация основных программных модулей системы выполнена Н.В. Прогонным.

База данных обеспечивала хранение исходных ЭКГ, результатов анализов крови, а также информации о пациентах. В основе БД лежат взаимосвязанные таблицы в формате Paradox 5,0, структура которых обеспечивает хранение всей необходимой информации.

На первом этапе компьютерной обработки подавлялись гармонические, случайные и импульсные помехи, а также устранялся дрейф изоэлектрической линии при использовании каскадного фильтра на основе базовых вычислительных процедур инструментальной системы «СИДИГРАФ».

На втором этапе проводилась оценка эталонных циклов ЭКГ на основе усреднения искаженных фазовых траекторий и определение амплитудно-временных параметров зубца P , комплекса QRS и сегмента $S-T$.

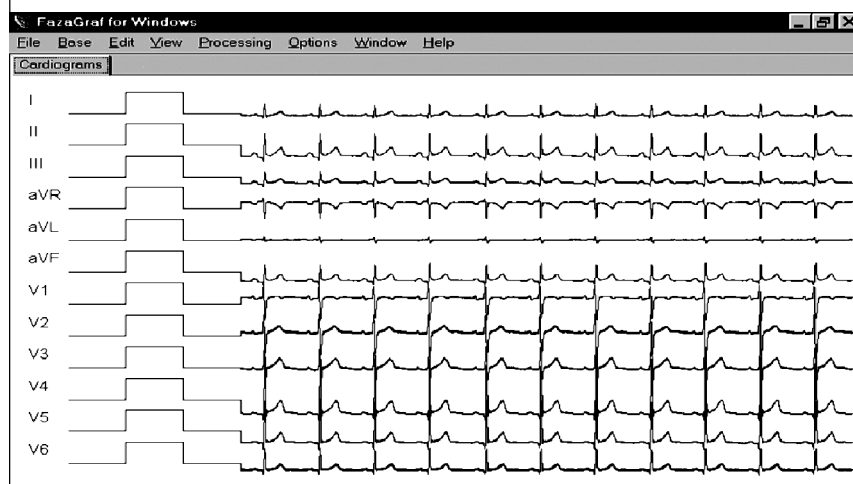
Результаты выполнения отдельных стадий обработки представлены в графическом виде на отдельных закладках программы (рис. 6.9—6.12).

6.2.2. Диагностические признаки ЭКГ в фазовом пространстве

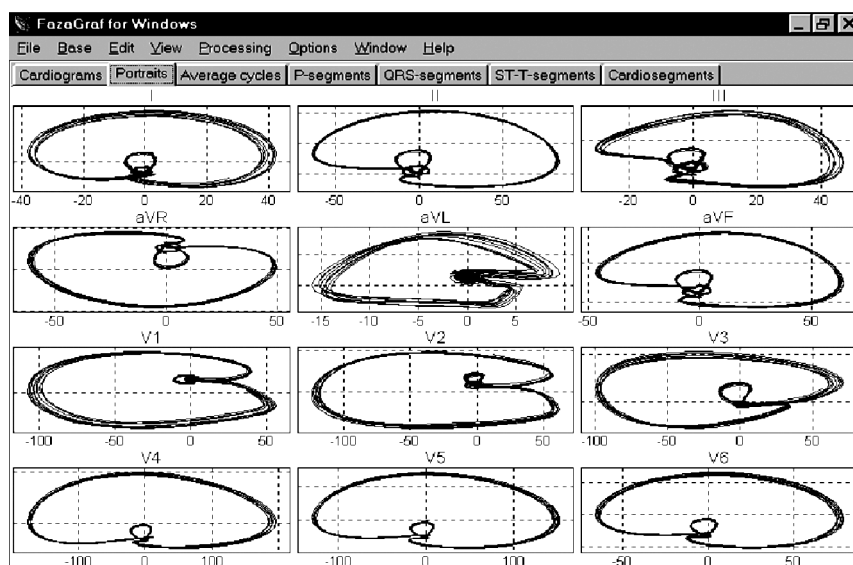
Понятно, что после восстановления 12 эталонных циклов ЭКГ (см. рис. 6.12) можно прямо воспользоваться любым известным алгоритмом диагностики, в частности Ганновским алгоритмом [389] интерпретации ЭКГ по традиционным амплитудно-временным признакам, сосредоточенным на локальных фрагментах эталонных циклов.

При создании новой ИТ обработки ЭКГ была поставлена задача — попытаться расширить состав традиционных для кардиологии амплитудно-временных признаков ЭКГ, введя дополнительные признаки ЭКГ в фазовом пространстве [203]. С этой целью в ИТ «FAZAWIN» реализованы процедуры, позволяющие измерять по четыре дополнительных признака ЭКГ в каждом из 12 отведений (рис. 6.13), а именно:

- параметры $\sigma[i]$, $i = 1, \dots, 12$, рассеивания точек фазовых траекторий;
- углы $\alpha[i]$, $i = 1, \dots, 12$, ориентации усредненной фазовой траектории;
- параметры $\beta_p[i]$, $i = 1, \dots, 12$, симметрии относительно оси $\dot{y} = 0$ фрагментов фазовых траекторий, соответствующих зубцам P ;



а

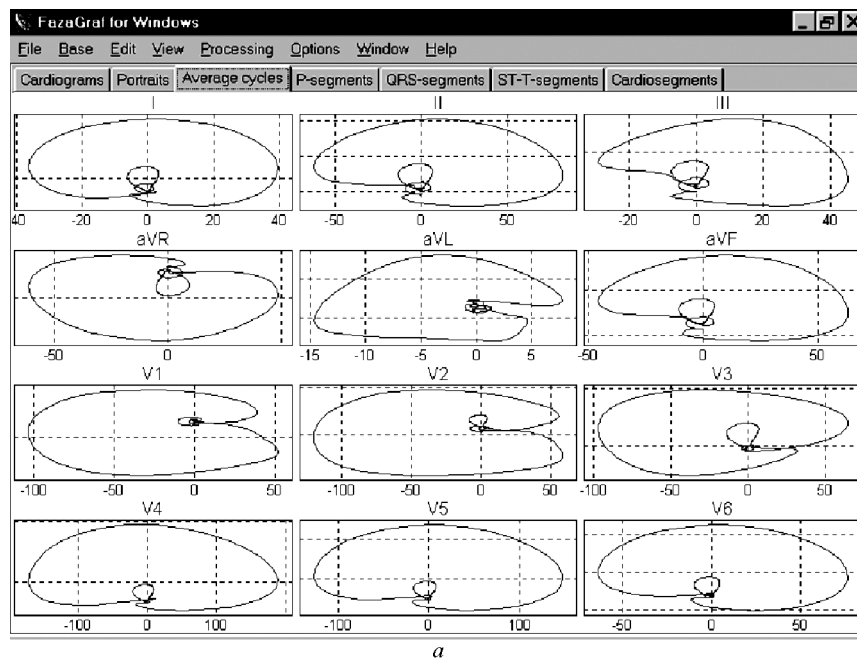


б

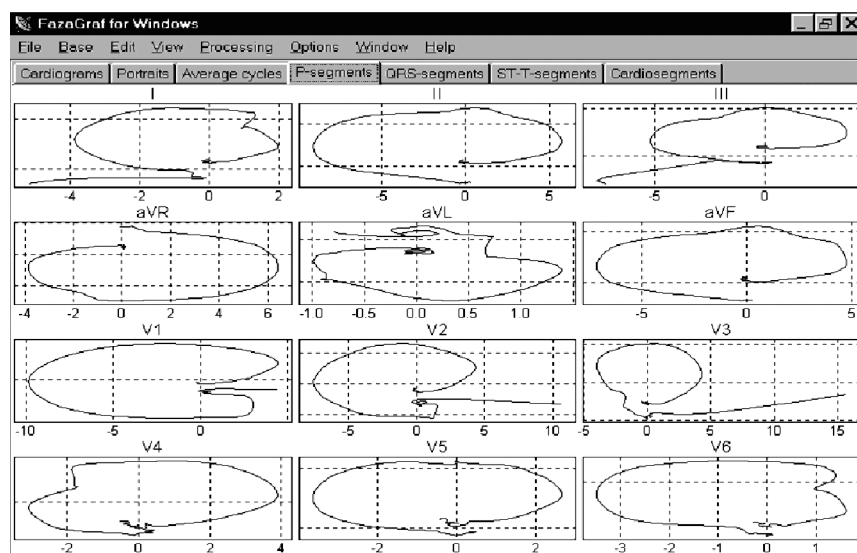
Рис. 6.9. Исходные ЭКГ в 12 отведениях:

а — во временной области; б — в фазовом пространстве

- параметры $\beta_T[i], i = 1, \dots, 12$, симметрии относительно оси $\dot{y} = 0$ фрагментов фазовых траекторий, соответствующих зубцам T .



а



б

Рис. 6.10. Усредненные фазовые траектории (а); фрагменты зубцов Р усредненных фазовых траекторий (б)

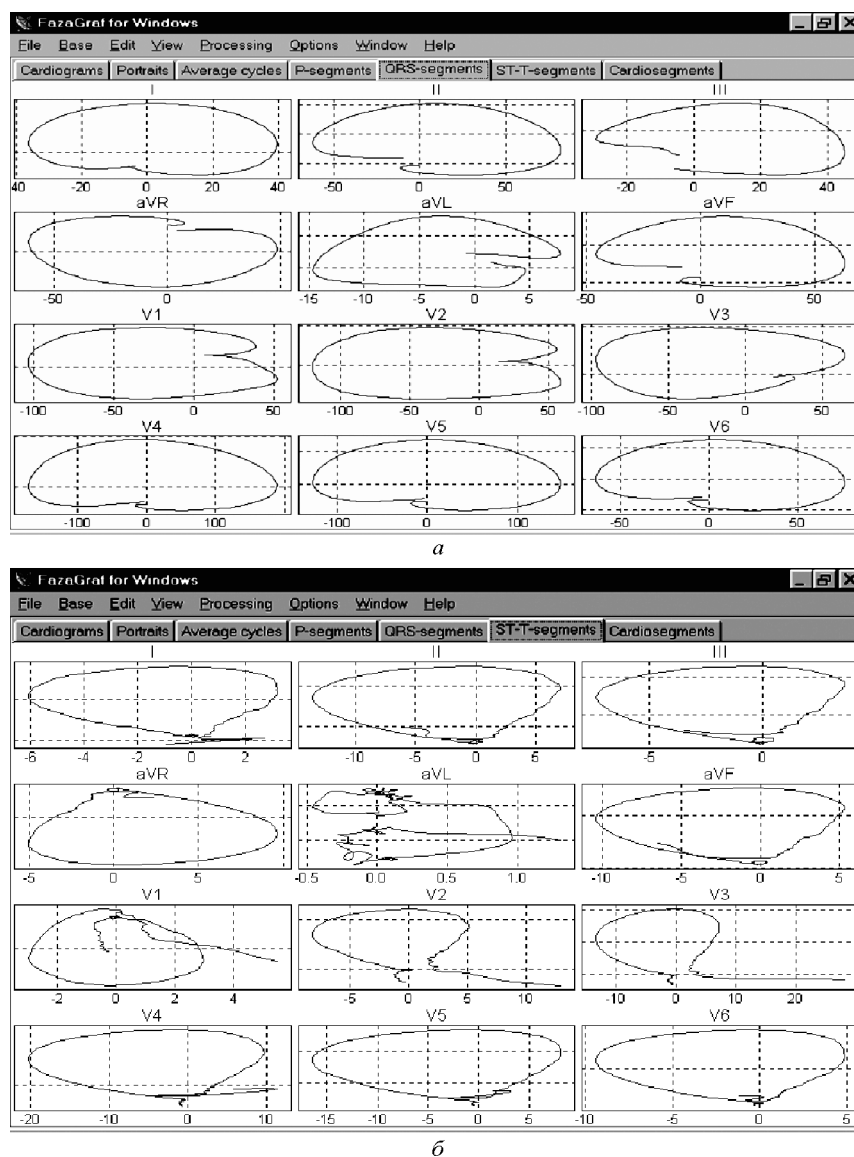


Рис. 6.11. Фрагменты комплексов QRS (а) и сегментов $S-T$ (б) усредненных фазовых траекторий

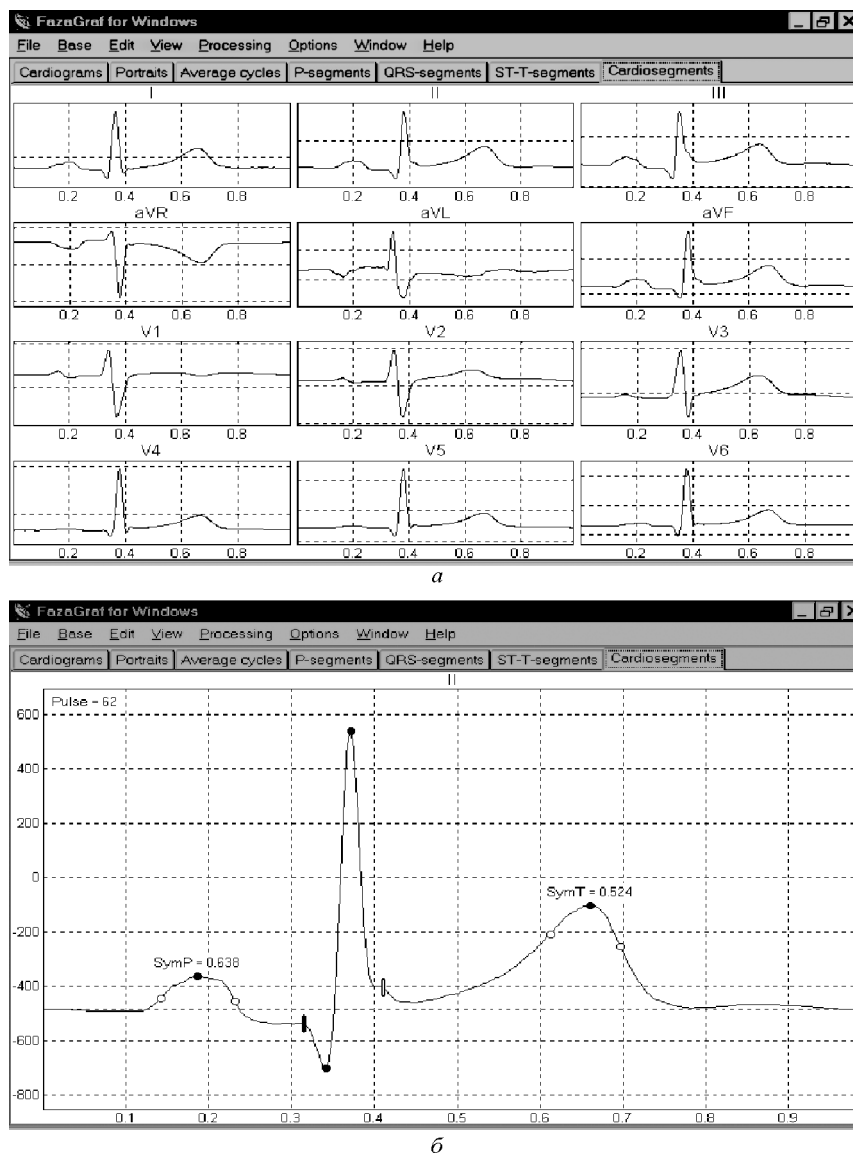


Рис. 6.12. Эталонные циклы ЭКГ в 12 отведениях (а), восстановленные по фазовым траекториям, и результат выделения их фрагментов с локально сосредоточенных признаков (б)

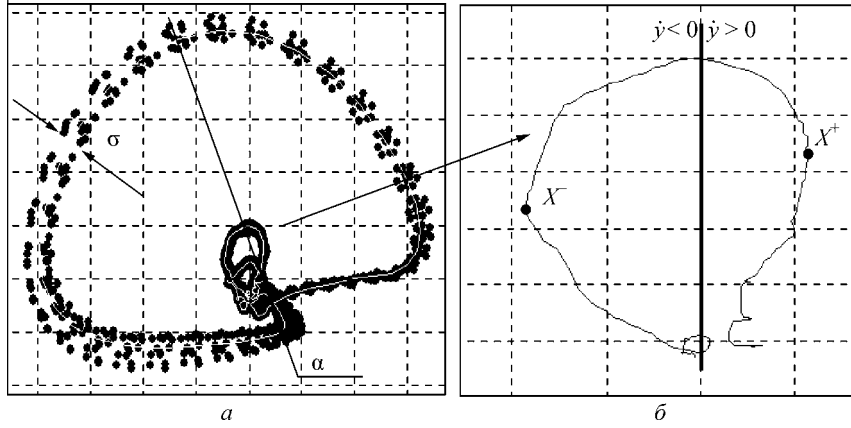


Рис. 6.13. Дополнительные признаки ЭКГ в фазовом пространстве:
 а — фазовые траектории; б — фрагмент усредненной фазовой траектории

Приемлемую оценку параметров $\sigma[\cdot]$ в каждом из 12 отведений удалось получать по элементам матрицы $D = \|R_H(Q_i, Q_j)\|$ хаусдорфовых расстояний (3.3), используя формулу

$$\sigma = \frac{1}{M-1} \sum_{m=1}^{M-1} R_H(Q_0, Q_m). \quad (6.28)$$

Для оценки параметров α и β проводится сегментация усредненной фазовой траектории, которая основана на несложной процедуре поиска точки $z_R = (y_R^*, \dot{y}_R^*)$ «самопересечения» внешнего витка траектории, соответствующей *QRS*-комплексу.

Угол ориентации α определяется по направлению прямых, проходящих через точку z_R и центр «тяжести» усредненной фазовой траектории, т. е. точку z_C с фазовыми координатами

$$y_C = 1 / K_0 \sum_{k=1}^{K_0} \hat{y}^*[k]; \quad \dot{y}_C = 1 / K_0 \sum_{k=1}^{K_0} \hat{\dot{y}}^*[k]. \quad (6.29)$$

Параметры β_P и β_T определяются отношением абсолютных значений $\dot{y}_{X^+}^*$ и $\dot{y}_{X^-}^*$ в точках X^+ и X^- соответствующего фрагмента усредненной фазовой траектории (см. рис. 6.13, б).

Таким образом, дополнительно измерялось 48 параметров фазовых траекторий ЭКГ, значения которых сохранялись для каж-

дого измерения в таблице БД (рис. 6.14, а). Там же сохранялись результаты анализов крови испытуемого (развернутый общий анализ, коагулограмма, печеночные пробы и ревмопробы) (рис. 6.14, б).

Предварительная обработка показала, что для фиксированного измерения значения первичных признаков β_p , β_T и σ были практически одинаковыми в каждом из 12 отведений, а их воз-

Parameters of processing

Parameters of complex processing:

	Disp	Alph	Ve1P	Ve2P	SymP	Ve1T	Ve2T	SymT	AmpP	AmpT	InOT	InOS
I	0.0186	75.62	1.989	3.848	0.517	3.254	6.091	0.534	53	174	0.34	0.08
II	0.0172	55.09	5.775	9.055	0.638	7.090	13.518	0.524	121	382	0.36	0.11
III	0.0219	44.14	3.781	5.201	0.727	3.723	7.360	0.504	83	212	0.33	0.09
aVR	0.0175	241.39	3.858	6.460	0.597	5.048	9.701	0.520	-83	-277	0.36	0.11
aVL	0.0483	148.64	0.992	1.399	0.709	0.287	0.966	0.297	-23	-21	0.29	0.11
aVF	0.0163	54.27	4.795	7.165	0.669	5.409	10.411	0.520	100	296	0.36	0.11
V1	0.0110	201.72	4.512	9.891	0.456	2.970	2.999	0.990	77	-36	0.37	0.12
V2	0.0103	191.07	4.418	7.647	0.576	5.211	8.155	0.639	72	303	0.32	0.11
V3	0.0155	145.44	4.310	3.964	1.086	7.221	13.451	0.537	74	485	0.32	0.09
V4	0.0096	101.52	3.934	3.083	1.276	9.843	20.261	0.486	66	620	0.36	0.10
V5	0.0097	94.21	2.681	3.637	0.737	8.176	16.399	0.499	47	479	0.36	0.11
V6	0.0130	71.93	1.644	3.483	0.472	4.916	9.372	0.525	50	272	0.36	0.11

Number of rows = 12

Close

а

Blood test

New Delete << >>

Erdman A. G.

Date of testing 30.10.98 (DD.MM.YYYY):

General blood test		Coagulograms		Rheumatic tests	
1. Red blood cells, x10e12	3.9	1. Prothrimbin index, %	78	1. Mucoprotein, un. opt. pl.	0.23
2. Haemoglobin, g/l	100	2. Recalcification time, sec.	125	2. Total protein, g/l	70
3. Color index	0.9	3. Fibrin, mg	20	3. Serum albumins, %	48.8
4. Reticulocytes, %		4. Fibrinogen, mg/l	4500	4. Serum globulins, %	51.2
5. White blood cells, x10e9	6.8	5. Beta-naphthal test	0	5. Alpha-1 globulin, %	5
6. Basophils, %		6. Ethanol test	0	6. Alpha-2 globulin, %	11.4
7. Eosinophils, %	1			7. Beta-globulin, %	12.4
8. Band neutrophils, %	3	Hepatic tests		8. Gamma-globulins, %	22.4
9. Segmented neutrophils, %	47	1. Total bilirubin, micromole/l	14.2	9. Antistreptolysine-O, MU/ml	
10. Lymphocytes, %	48	2. Thymol turbidity test, un.	4.3	10. C-reactive protein	2
11. Monocytes, %	1	3. Phernal test	0		
12. Erythrocyte sedimentation rate, mm/g	27	4. Alanine transaminase	0.25		
		5. Asparagine transaminase	0.2		

Close

б

Рис. 6.14. Интерфейсные окна системы «FAZAWIN»:

а — для отображения первичных признаков ЭКГ в фазовом пространстве; б — для ввода результатов анализа крови

можные отклонения от средних значений, скорее всего, можно связывать со случайными факторами, а не с физиологическим состоянием пациента.

Для упрощения дальнейших исследований и повышения надежности измерения диагностических признаков предложено перейти от первичных признаков к вторичным, интегрально оценивающим все 12 отведений.

С этой целью для каждого испытуемого регистрировались не менее 10 ЭКГ и по ним определялись средние значения параметров σ , β_P , β_T во всех 12 отведениях. Для исключения «диких» значений при усреднении отбрасывали наибольшее и наименьшее значения, в результате этого повышалась надежность измерения указанных вторичных признаков.

При формировании вторичных признаков использовали не сам угол α ориентации усредненной фазовой траектории, а его отклонение $\Delta\alpha_i$ ($i = 2, \dots, 12$) по отношению к первому стандартному отведению. Таким образом, вместо 48 первичных признаков при дальнейших статистических исследованиях использовалось 14 дополнительных признаков, характеризующих все 12 отведений ЭКГ в фазовом пространстве, в том числе:

- σ^I — интегральный показатель разброса фазовых траекторий;
- β_P^I — интегральный параметр симметрии P -зубцов;
- β_T^I — интегральный параметр симметрии T -зубцов;
- $\alpha_{II}, \alpha_{III}, \alpha_{AVR}, \alpha_{AVL}, \dots, \alpha_{V5}, \alpha_{V6}$ — относительные углы ориентации усредненных траекторий соответствующих отведений.

Специальная функция сканирования базы данных, реализованная в ИТ FAZAWIN, позволяет сформировать таблицу, в которой сохраняются указанные вторичные признаки для каждого испытуемого. В эту же таблицу вводятся формируемые электрокардиографом BIOSET 8102 MI значения традиционных признаков ЭКГ во временной области.

6.2.3. Результаты клинических исследований

Клинические испытания ИТ «FAZAWIN» проводились в отделении ревматологии Национального научного центра «Институт кардиологии им. акад. Н.Д. Стражеско». Главная цель испытаний — оценка диагностических возможностей разработанной

технологии для анализа и интерпретации ЭКГ больных ревматоидным артритом, а также изучение взаимосвязи показателей ЭКГ с результатами анализов крови, отражающими активность воспалительного процесса.

В клинических испытаниях участвовали 43 человека, в том числе 18 мужчин и 25 женщин, составивших 3 группы.

В первую группу включены 16 больных (6 мужчин и 10 женщин) хроническим ревматизмом, сочетанным с ишемической болезнью сердца (ИБС).

Во вторую — 14 больных (7 мужчин и 7 женщин) ревматоидным артритом. Все больные находились на стационарном лечении. Клинический диагноз устанавливался на основе комплексного физикального, инструментального и лабораторного обследования.

В третью группу испытуемых (контрольная группа) были включены 13 добровольцев (5 мужчин и 8 женщин) без кардиологической патологии в анамнезе, с нормальными результатами физикального обследования и ЭКГ.

Статистические исследования выполнялись на основе обработки данных, полученной в результате сканирования БД для трех групп испытуемых (более 1000 записей ЭКГ). В таблице были сведены значения признаков ЭКГ во временной области (21 признак), интегральных признаков ЭКГ в фазовом пространстве (14 признаков) и соответствующие этим ЭКГ показатели анализов крови испытуемых (33 показателя).

Статистическая обработка проводилась с использованием t -теста (критерия Стьюдента), корреляционного, регрессионного и дискриминантного анализов (применялась программа STATISTICA 5,0), а также внешней программы, реализующей комбинаторный алгоритм МГУА [184].

Результаты испытаний подробно описаны в работе [143]. Поэтому дадим лишь краткую оценку наиболее интересным результатам.

Установлено, что с надежностью вывода $P > 0,99$ у больных ревматоидным артритом, ЭКГ которых были признаны нормальными при их традиционном анализе во временной области, наблюдались неслучайные изменения средних значений признаков ЭКГ в фазовом пространстве:

- уменьшение интегрального показателя σ^I разброса фазовых траекторий (0,06 и 0,11 соответственно);

- увеличение относительного угла ориентации $\Delta\alpha_{V_1}$ усредненной фазовой траектории в первом грудном отведении (165,2 и 123,3 соответственно);
- увеличение относительного угла ориентации $\Delta\alpha_{V_3}$ усредненной фазовой траектории в третьем грудном отведении (101,9 и 63,8 соответственно);
- уменьшение относительного угла ориентации $\Delta\alpha_{V_5}$ усредненной фазовой траектории в третьем грудном отведении (97,3 и 184,9 соответственно).

Статистическая обработка ЭКГ испытуемых контрольной группы показала наличие корреляционной связи с коэффициентом корреляции $r = 0,94$ между относительными углами $\Delta\alpha_{AVR}$ и $\Delta\alpha_{V_1}$ ориентации усредненных фазовых траекторий в отведениях AVR и V_1 . На ЭКГ испытуемых всех групп наблюдалась отрицательная корреляционная связь с коэффициентом корреляции $r = -0,73$ между максимальным фронтальным вектором QRS -комплекса (в град.) и $\Delta\alpha_{V_1}$.

С надежностью вывода $P > 0,95$ установлена также парная корреляция между признаками ЭКГ в фазовом пространстве и показателями крови, отражающими активность воспалительного процесса, в том числе:

- относительными углами $\Delta\alpha_{V_2}, \Delta\alpha_{V_3}$ ориентации усредненной фазовой траектории во втором и третьем грудном отведении с лейкоцитами ($\times 10^9$);
- интегральным параметром β_P^I симметрии зубца P и фибриногеном (мг/л);
- интегральным параметром β_T^I симметрии зубца T и палочкоядерными нейтрофилами.

Для более детального изучения статистической зависимости между предложенными признаками ЭКГ в фазовом пространстве и показателями крови был привлечен комбинаторный алгоритм МГУА. В качестве регрессоров использовались 33 показателя венозной крови испытуемых (табл. 6.3).

В результате были построены модели оптимальной сложности, связывающие признаки ЭКГ в фазовом пространстве (отклики) с указанными регрессорами [207]. В частности, установлено, что существуют надежные статистические связи между относительными углами ориентации усредненной фазовой траек-

ТАБЛИЦА 6.3

Потенциальные регрессоры

№	Обозначение	Наименование
1	<i>ERY</i>	Эритроциты, $\times 10^{12}$
2	<i>HGB</i>	Гемоглобин, г/л
3	<i>COL</i>	Цветной показатель
4	<i>RET</i>	Ретикулоциты, %
5	<i>LEU</i>	Лейкоциты, $\times 10^9$
6	<i>BAS</i>	Базофилы, %
7	<i>EOS</i>	Эозинофилы, %
8	<i>PLN</i>	Палочкоядерные нейтрофилы, %
9	<i>SGN</i>	Сегментоядерные нейтрофилы, %
10	<i>LYM</i>	Лимфоциты, %
11	<i>MON</i>	Моноциты, %
12	<i>SOE</i>	Скорость оседания эритроцитов, мм/г
13	<i>PTI</i>	Протромбиновый индекс, %
14	<i>REC</i>	Время рекальцификации, с
15	<i>FIB</i>	Фибрин, мг
16	<i>FBG</i>	Фибриноген, мг/л
17	<i>BNA</i>	Бета-нафтоловый тест
18	<i>ETH</i>	Етаноловый тест
19	<i>BIL</i>	Билирубин общий, л
20	<i>TIM</i>	Тимоловая проба
21	<i>FRM</i>	Фермоловая проба
22	<i>ALT</i>	Аланиновая трансфераза
23	<i>AST</i>	Аспаргиновая трансфераза
24	<i>MUK</i>	Мукопротеины
25	<i>BEL</i>	Общий белок, г/л
26	<i>ALB</i>	Альбумины, %
27	<i>GLB</i>	Глобулины, %
28	<i>A1G</i>	Альфа-1 глобулин, %
29	<i>A2G</i>	Альфа-2 глобулин, %
30	<i>BGL</i>	Бета-глобулин, %
31	<i>GGL</i>	Гамма-глобулин, %
32	<i>ASL</i>	Антистрептомицин-О
33	<i>CRP</i>	С-реактивный протеин

тории ЭКГ во втором стандартном (отклик $\Delta\alpha_2$) и в пятом грудном (отклик $\Delta\alpha_{V5}$) отведениях и параметрами крови. Эти модели имеют вид

$$\begin{aligned} \Delta\alpha_2 = & -4,75984LEU - 36,7901PLN - 2,43534LYM + 1,3082REC + \\ & + 0,02732FBG - 122,317FRM - 3,24185GLB - 0,75813A1G + \\ & + 16,497BGL + 13,836GGL, \quad r_M \approx 0,779, \quad n = 85; \quad (6.30) \end{aligned}$$

$$\begin{aligned}
\Delta\alpha_{V5} = & 156,687ERY - 13,4807LEU + 1,58091REC + \\
& + 23,0974FIB - 0,0707351FBG - 109,307FRM - \\
& - 610,457ALT - 0,23831BEL + 0,746503AIG + \\
& + 10,7829GGL + 0,122102ASL, \quad r_M \approx 0,777, \quad n = 87, \quad (6.31)
\end{aligned}$$

где r_M — коэффициент множественной корреляции; n — число наблюдений.

Таким образом, подтверждена гипотеза о взаимосвязи признаков ЭКГ в фазовом пространстве и параметров крови, отражающих активность воспалительных заболеваний.

Оказалось, что совокупность признаков ЭКГ во временной области и в фазовом пространстве позволяет распознавать представителей трех исследуемых групп даже в тех случаях, когда традиционный метод анализа признаков во временной области признавал ЭКГ неизменными. Показано, что диагностические решения могут приниматься на основе измерения двух комплексных показателей, связанных с признаками ЭКГ в виде линейных функций

$$\begin{aligned}
R_1 = & -26,08\sigma^I + 3,5318\beta_P^I + 8,3044\beta_T^I - 0,0047\Delta\alpha_3 + \\
& + 0,0252\Delta\alpha_{V1} + 0,00177\Delta\alpha_{V2} + 0,0094\Delta\alpha_{V4} + 0,0009\Delta\alpha_{V6} - \\
& - 66,297F_V + 0,0210P_\Sigma - 12,5815, \quad (6.32)
\end{aligned}$$

$$\begin{aligned}
R_2 = & -7,33449\sigma^I + 6,0665\beta_P^I - 1,42512\beta_T^I - \\
& - 0,00422\Delta\alpha_3 - 0,00707\Delta\alpha_{V1} + 0,00076\Delta\alpha_{V2} + 0,00382\Delta\alpha_{V4} + \\
& + 0,00775\Delta\alpha_{V6} + 19,62837F_V - 0,00832P_\Sigma - 1,09178, \quad (6.33)
\end{aligned}$$

где σ^I — интегральный показатель разброса фазовых траекторий; β_P^I, β_T^I — интегральные показатели симметрии фрагментов усредненных фазовых траекторий, соответствующих зубцам P и T ; $\Delta\alpha_3, \Delta\alpha_{V1}, \Delta\alpha_{V2}, \Delta\alpha_{V4}, \Delta\alpha_{V6}$ — относительные углы ориентации усредненных фазовых траекторий в третьем стандартном, первом, втором, четвертом и шестом отведениях; F_V — фронтальный вектор волны P ; P_Σ — признак, характеризующий суммарную «площадь» волны P во всех 12 отведениях.

ТАБЛИЦА 6.4

Решающее правило для классификации испытуемых трех групп

Группа	Правило классификации
Ревматоидный артрит (РА)	$R_1 > 0,25$
Ревматизм и/или ИБС	$R_1 < 0,25$ и $R_2 > 1,5$
Контрольная группа (КГ)	$R_1 < 0,25$ и $R_2 < 1,5$

Решение принимается по пороговому правилу, приведенному в табл. 6.4.

При оценке решающего правила по выборке, состоящей из 44 наблюдений, не участвующих в обучении, была допущена только 1 ошибка: здоровый испытуемый был отнесен к группе больных ревматоидным артритом (рис. 6.15).

Результаты экспериментов показали, что возможности ЭКГ диагностики далеко еще не исчерпаны: применение ИТ, реализующей предложенный метод обработки ЭКГ в фазовом пространстве, позволяет повысить чувствительность и специфичность этого метода при диагностике больных ревматоидным артритом.

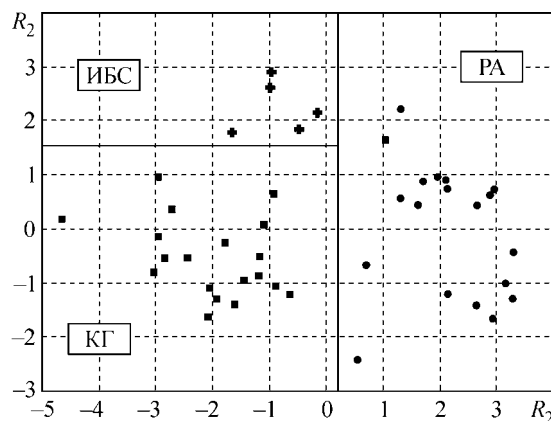


Рис. 6.15. Результаты оценки ИТ «FAZAWIN»:

РА — группа больных ревматоидным артритом; ИБС — группа больных хроническим ревматизмом, сочетанным с ишемической болезнью сердца; КГ — здоровые добровольцы (контрольная группа)

6.3. Информационная технология «ФАЗАГРАФ»

6.3.1. Задача информационной технологии

На современном этапе развития промышленного производства все большее значение приобретает контроль функционального состояния операторов, осуществляющих эксплуатацию оборудования. Эта проблема не менее актуальна, чем проблема контроля самого оборудования, поскольку иногда ошибочные действия оператора, особенно в нештатных ситуациях, приводят к крупномасштабным авариям, последствия которых хорошо известны.

Следует отметить, что указанная проблема не может быть в полной мере разрешена лишь за счет регулярных диспансерных обследований операторов, проводимых в условиях поликлиники или лечебно-оздоровительного центра предприятия. Основной целью плановой диспансеризации является выявление ранней или скрытой стадии болезни обследуемого и принятие решения о пригодности к выполнению профессиональных обязанностей.

В то же время трудовая деятельность операторов ряда профессий (водителей транспортных средств, летчиков, авиадиспетчеров, операторов химических и металлургических агрегатов, атомных электростанций и т. п.) часто происходит в необычных (экстремальных) условиях среды и сопряжена со стрессовыми ситуациями. Даже если на очередном профессиональном медицинском осмотре такие операторы признаны практически здоровыми и допущены к выполнению профессиональных обязанностей, то в процессе трудовой деятельности необходимо осуществлять постоянный контроль их текущего функционального состояния.

Известно, что ритм сердца — универсальная реакция организма на любое воздействие со стороны внешней и внутренней среды. Он содержит в себе информацию о функциональном состоянии всех звеньев регулирования жизнедеятельности человека как в норме, так и при различных патологиях [15]. Компьютерный анализ вариабельности сердечного ритма позволяет получить ценную информацию о состоянии регуляторных систем организма оператора.

Однако для определения функционального состояния сердца, как основного системообразующего органа, только данных анализа вариабельности не достаточно. Поэтому для повыше-

ния достоверности заключения о функциональном состоянии оператора предлагается дополнить анализ вариабельности сердечного ритма упрощенным морфологическим анализом ЭКГ испытуемого.

6.3.2. Базовые гипотезы, положенные в основу ИТ «ФАЗАГРАФ»

Фазовая траектория ЭКГ внешне напоминает известную в математике кривую — кардиоиду, название которой вызвано сходством ее графика с формой сердца человека (см. рис. 6.9, б).

Естественно, автор далек от мысли приписывать данное совпадение «интеллектуальным» свойствам когнитивной компьютерной графики, а лишь хотел бы присоединиться к мысли, высказанной академиком Д.А. Поспеловым [153] о том, что когнитивный графический образ, демонстрирующий некую тайну Природы, — это всегда интересно, красиво и ... *неожиданно!*

Что же неожиданного несет порождаемая компьютером фазовая траектория ЭКГ, которую смело можно причислить к когнитивному графическому образу?

Для ответа на этот вопрос проводились многочисленные наблюдения, на основе которых удалось обнаружить ряд интересных свойств этих образов [180, 267, 276, 278].

Под действием физических и/или эмоциональных нагрузок происходит «деформация» когнитивного образа, которая чаще всего выражается в увеличении величины разброса точек фазовых траекторий относительно среднего цикла. После снятия нагрузок по истечении определенного промежутка времени, зависящего от функционального состояния испытуемого, происходит восстановление исходного когнитивного образа (рис. 6.16). Именно этот факт послужил основанием для введения параметра σ в расширенное пространство признаков ЭКГ, о котором было изложено выше.

Наблюдения также показали, что под действием физической нагрузки у слабо тренированных людей происходит увеличение показателя β_T симметрии фрагмента усредненной фазовой траектории, в то время как у «спортивных» — такие изменения чаще всего несущественны (рис. 6.17).

Эти предварительные наблюдения и стали предпосылками разработки новой ИТ «ФАЗАГРАФ» для диагностики функционального состояния операторов [148, 149, 212].

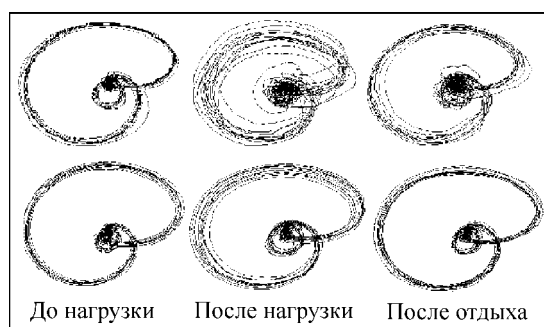


Рис. 6.16. Динамика показателя σ под действием нагрузки: сверху — нетренированный испытуемый; внизу — тренированный

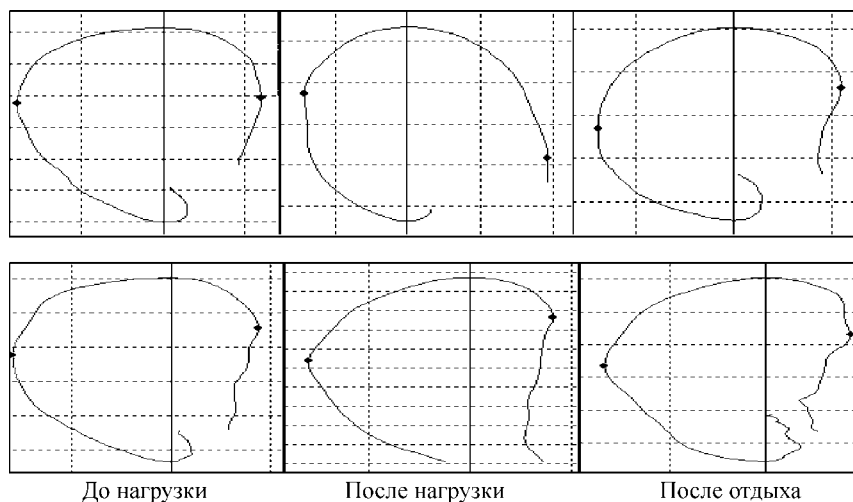


Рис. 6.17. Динамика показателя β_T под действием нагрузки: сверху — нетренированный испытуемый; внизу — тренированный

6.3.3. Общая характеристика информационной технологии

При построении ИТ была поставлена задача упрощения процедуры регистрации ЭКГ с тем, чтобы ИТ могла использоваться непосредственно на рабочем месте оператора.

С этой целью предложено осуществлять обработку ЭКГ только первого стандартного отведения (правая рука — левая рука), которое интегрально отражает поведение электрического

вектора сердца. Для регистрации сигнала используется специальный микропроцессорный сенсор ⁴, который позволяет регистрировать ЭКГ в первом стандартном отведении при прикосновении двух пальцев к миниатюрным электродам и вводить пакеты оцифрованных данных в персональный компьютер по протоколу стандартных интерфейсов — проводному USB либо беспроводному (инфракрасному) IrDA (рис. 6.18).

Дальнейшая обработка сигнала производится с помощью компьютерной программы ⁵, реализующей предложенный метод анализа ЭКГ в фазовом пространстве координат [203, 281, 283]. Оценка функционального состояния оператора проводится:

- путем анализа диагностических признаков ЭКГ в состоянии покоя;
- сравнением значений диагностических признаков ЭКГ до и сразу после нагрузки;
- сравнением значений диагностических признаков текущей ЭКГ и ЭКГ, хранимой в архиве базы данных.

Предусмотрено тестирование оператора на основе трех типов нагрузок: недозированной физической нагрузки (например, пробы Руфье), физической нагрузки в виде дозированной велоэргометрической пробы (мощностью 50—75 Вт) и психоэмоциональной нагрузки в условиях дефицита времени.



Рис. 6.18. Внешний вид микропроцессорного регистратора ЭКГ с пальцевыми электродами

⁴ Эта и последующие версии микропроцессорных регистраторов ЭКГ разработаны Международным научно-учебным центром информационных технологий и систем НАН Украины и МОН Украины совместно с АОЗТ «Сольвейг» (г. Киев) в рамках ГНТП «Образный компьютер».

⁵ Реализация основных программных модулей ИТ «ФАЗАГРАФ» и ее последующих модификаций выполнена Н.А. Семергеем.

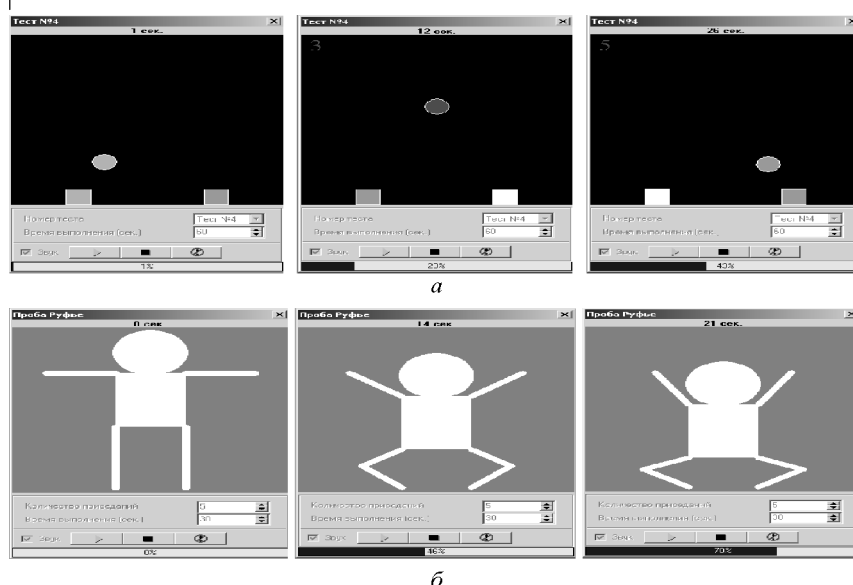


Рис. 6.19. Интерфейсы подсистемы управления нагрузкой: *а* — психоэмоциональной; *б* — физической

Для имитации психоэмоциональной нагрузки, воздействующей на оператора в условиях реальной трудовой деятельности, предложено использовать так называемые стресс-тесты, стимулирующие простые и сложные зрительно-моторные реакции оператора на предъявляемые компьютером графические стимулы (рис. 6.19, *а*).

В процессе выполнения теста при правильных действиях оператора темп предъявления компьютером стимулов автоматически ускоряется, а при ошибочных — замедляется, что создает стрессовую ситуацию для испытуемого. При этом фиксируется общее число неправильных действий оператора и производится оценка его реакций по отношению к лидеру. Тест завершается по истечении заданного промежутка времени либо по достижении заданного числа предъявлений стимулов.

В подсистеме управления нагрузкой реализован ряд тестов разной степени сложности.

Тест 1. На экране монитора через случайные промежутки времени появляются графические образы двух цветов, которые испытуемый должен направить в корзину соответствующего цвета, используя клавиши управления курсором.

Тест 2. На экране монитора через случайные промежутки времени появляются графические фигуры трех цветов, две из которых представляют собой сигналы возбуждающего типа, а фигура третьего цвета — сигнал тормозного типа. Испытуемый должен направить стимулы возбуждающего типа в корзину соответствующего цвета, а на стимул тормозного типа не должен реагировать.

Тест 3. Дополнительной психоэмоциональной нагрузкой к тесту 2 является то, что в случайные моменты времени меняются местами цвета левой и правой корзин.

Тест 4. Дополнительной психоэмоциональной нагрузкой к тесту 3 является то, что в некоторые случайные моменты времени случайным образом меняются цвета графических стимулов и корзин, причем в данном случае к сигналу тормозного типа испытуемый должен отнести тот графический стимул, цвет которого не совпадает с цветами обеих корзин.

Для управления недозированной физической нагрузкой в виде пробы Руфье в интерфейсе системы реализовано специальное окно, в котором отображается требуемый темп приседаний в виде анимационной картинке, сопровождаемой звуковым сигналом (рис. 6.19, б). Время нагрузки и ее темп регулируются в зависимости от пола, возраста и физической подготовки испытуемого.

6.3.4. Схема формирования информационного продукта

Последовательность стадий формирования информационного продукта (рис. 6.20) полностью соответствуют рассмотренной в гл. 2 обобщенной модели построения информационной ИТ обработки сигнала с локально сосредоточенными признаками и предусматривает три основных этапа обработки: выделение полезного сигнала по искаженным реализациям; анализ диагностических признаков; интерпретация сигнала (оценка функционального состояния оператора) по результатам измерения признаков.

Поскольку регистрация ЭКГ проводилась непосредственно на рабочем месте, предварительная обработка исходного сигнала потребовала использования каскадного фильтра (см. гл. 4). Оценка эталонного цикла в фазовом пространстве и во временной области реализуется на основе алгоритмов, описанных в гл. 3, и предусматривает:

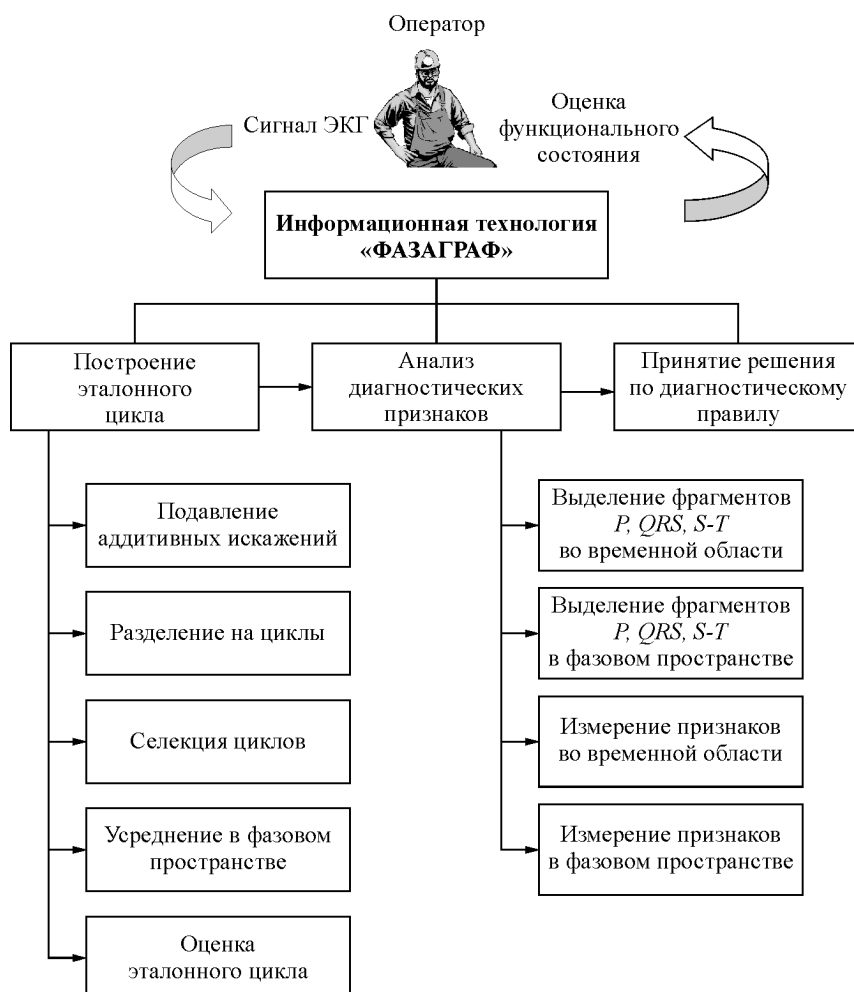


Рис. 6.20. Упрощенная схема формирования информационного продукта

- разделение предварительно отфильтрованного сигнала на отдельные циклы по профилю $r_0^2[k]$ расстояний между нормированными фазовыми координатами $y^*[k], \dot{y}^*[k]$ каждой k -й точки и базовой точки k_0 , определяемой соотношением (3.12);

- выделение доминантного (типичного) цикла наблюдаемого сигнала по опорной траектории Q_0 , найденной по матрице хаусдорфовых расстояний в соответствии с соотношением (3.4);
- селекцию нетипичных циклов;
- оценку среднего цикла ЭКГ во временной области и в фазовом пространстве.

6.3.5. Оценка эффективности ИТ «ФАЗАГРАФ»

для решения задачи скрининга ишемической болезни сердца

Диагностика ишемической болезни сердца (ИБС) — одна из основных задач современной кардиологии. Сердечно-сосудистыми заболеваниями страдают три четверти населения Украины, и в 62,5 % случаев они становятся причиной смерти.

Основной причиной смерти от сердечно-сосудистой патологии является ишемическая болезнь сердца. Данные недавних эпидемиологических исследований свидетельствуют о резком повышении встречаемости ИБС в индустриально развитых странах. В Украине с 1995 по 2005 гг. заболеваемость ИБС выросла вдвое (с 10 тыс. до более чем 20 тыс. на 100 тыс. населения).

По наблюдениям клиницистов, приблизительно в половине случаев госпитализация по поводу инфаркта была первым в жизни контактом больного с кардиологом. Это объясняется как бессимптомным течением ИБС, так и частой недооценкой больными эпизодов дискомфорта в грудной клетке. В этой связи разработка новых методов для объективного скрининга ИБС является актуальной и своевременной задачей.

Разработка ИТ «ФАЗАГРАФ» и ее последующих версий потребовала проведения исследований, направленных на оценку эффективности диагностических признаков. С этой целью в систему включен модуль, обеспечивающий построение гистограмм условных распределений признаков (рис. 6.21), а также модуль усиленного *ROC*-анализа (рис. 6.22), в котором реализована процедура, описанная в гл. 5. Такие инструментальные средства позволили в единой информационной среде оценивать полезность предварительно отобранных признаков и обосновать эффективность диагностического теста с точки зрения уменьшения априорного риска.

Клинические испытания проводились в отделении ишемических болезней сердца Национального научного центра «Институт кардиологии им. акад. Н.Д. Стражеско» АМН Украины (Ки-

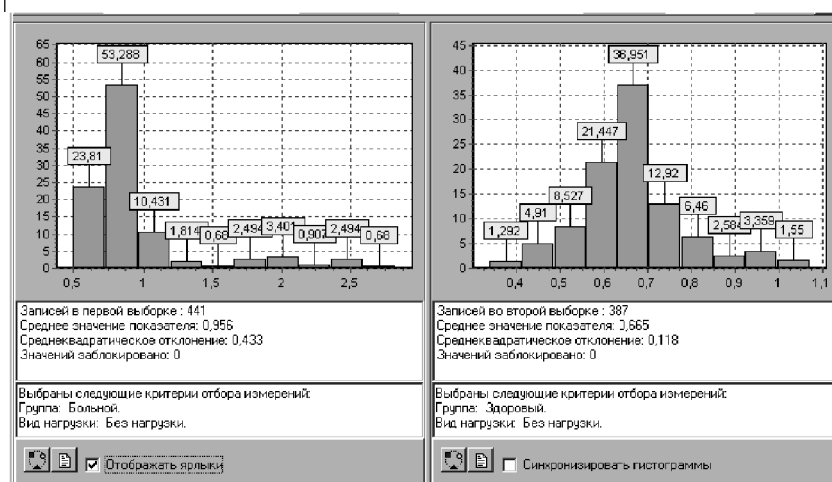


Рис. 6.21. Интерфейс модуля построения гистограмм

ев). Кроме того, использовались данные испытаний в четырех клиниках Германии — кардиологической клинике университета Дуйсбург-Эссен (Essen University Hospital), католическом госпитале «Филлипусстифт» (Katholical Hospital «Phillpusstift» (Essen)), Центре сердца земли Северный Рейн-Вестфалия (Heart and Diabetes Center of North Rhein-Weastfalia (Bad-Oeynhausen)), Германском центре сердца (German Heart Center (Berlin)) [283].

Выборка наблюдений составила записи ЭКГ 441 верифицированного больного ишемической болезнью сердца и 387 записей ЭКГ здоровых добровольцев, включенных в контрольную группу (КГ).

На рис. 6.21 представлены гистограммы условных распределений значений признака β_T в группе больных ИБС (слева) и здоровых добровольцев (справа). Видно, что среднее значение параметра β_T существенно различалось в группах и составило $0,956 \pm 0,43$ и $0,665 \pm 0,12$ соответственно. Проверка полученного результата по t -критерию Стьюдента показала, что с высокой вероятностью ($P > 0,999$) гипотеза о случайном различии средних значений может быть не принята.

Анализ ROC-кривой (см. рис. 6.22), которая отображает зависимость чувствительности S_E теста по признаку β_T от величи-

ны $1 - S_p$, где S_p — специфичность теста, показывает, что кривая достаточно далека от диагонали, причем площадь под ROC -кривой составляет 0,832 ед.

Как отмечалось в гл. 5, большая площадь под ROC -кривой свидетельствует лишь о выполнении необходимых условий полезности теста. Поэтому в разработанный модуль усиленного ROC -анализа введена дополнительная процедура, обеспечивающая автоматическое построение области полезности теста, удовлетворяющей достаточным условиям (5.33) и (5.34).

Для проверки полезности теста достаточно ввести в модуль частоту $P_{\text{ИБС}}$ (преваленс) исследуемого заболевания и допустимое соотношение $\omega_{\text{ИБС}}$ потерь от ложноотрицательных и ложноположительных ошибок диагностики.

Согласно данным медицинской статистики [169], частота (преваленс) ИБС среди лиц старше 40 лет в крупных городах Европы составляет 4—6 %. На рис. 6.22 выделена область допустимых значений чувствительности и специфичности теста, построенная при $P_{\text{ИБС}} = 6\%$ и соотношении потерь ложноотрицательных и ложноположительных ошибок $\omega_{\text{ИБС}} = 5$.

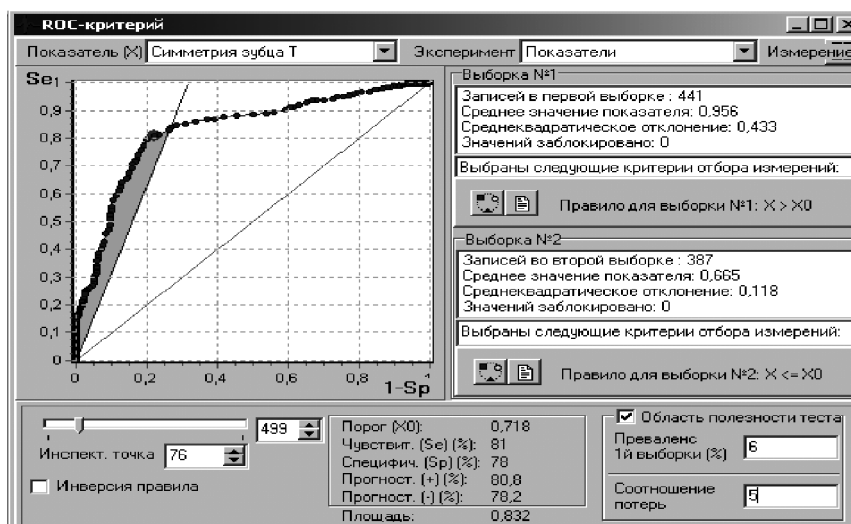


Рис. 6.22. Интерфейс модуля построения ROC -кривой

Принятие решений по пороговому правилу:

$$\begin{aligned} &\text{ИБС,} && \text{если } \beta_T > \beta_0, \\ &\text{Норма,} && \text{если } \beta_T < \beta_0, \end{aligned} \quad (6.34)$$

где $\beta_0 \approx 0,72$, обеспечивает чувствительность $S_E = 81\%$ и специфичность $S_P = 78\%$. Отметим (см. рис. 6.22), что тест с такими операционными характеристиками попадает в область полезных тестов. Следовательно, правило (6.34) может быть рекомендовано для скрининга ИБС при массовых профилактических обследованиях населения.

Помимо чувствительности и специфичности, в медицинской диагностике используют так называемые показатели прогностичности (*predictive value*) положительного $P(+)$ и отрицательного $P(-)$ результатов, которые оценивают соответственно апостериорные вероятности наличия заболевания при положительном результате тестирования и отсутствия заболевания при отрицательном результате.

Из формулы Байеса следует, что, в принятых нами обозначениях, эти показатели оцениваются соотношениями

$$P(+) = \frac{1}{1 + \frac{1-P}{P} \frac{1-S_P}{S_E}}, \quad P(-) = \frac{1}{1 + \frac{P}{1-P} \frac{1-S_E}{S_P}}. \quad (6.35)$$

Для полноты картины в разработанном модуле усиленного ROC-анализа реализованы также процедуры вычисления $P(+)$ и $P(-)$ по этим соотношениям (см. рис. 6.22). Как видно из (6.35), при тестировании ИБС по правилу (6.34) обеспечивается $P(+)=80,8\%$ и $P(-)=78,2\%$, что вполне приемлемо для скрининга ИБС при массовых донозологических обследованиях населения.

Заметим также, что согласно условиям (5.40) тест, основанный на правиле (6.34), остается полезным в достаточно широком диапазоне допустимых значений соотношения потерь: $4,2 \leq \omega_{\text{ИБС}} \leq 64,1$.

6.3.6. Интерфейс пользователя

При создании системы мы стремились упростить не только алгоритмы обработки сигнала, но и по возможности интерфейс ко-

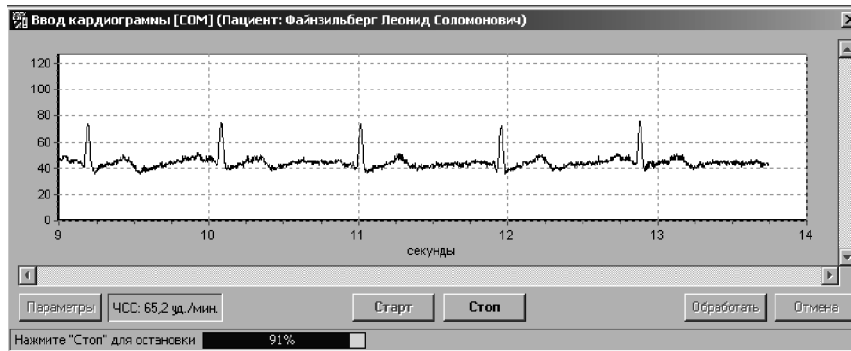


Рис. 6.23. Окно ввода ЭКГ

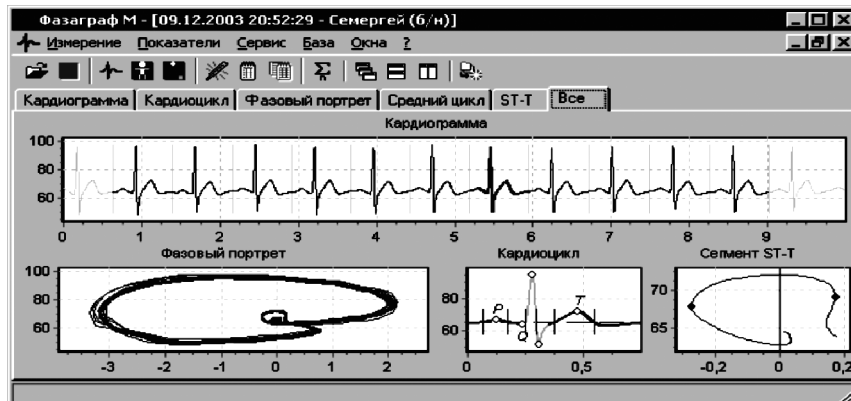


Рис. 6.24. Окно вывода графических результатов обработки

нечного пользователя, предоставив ему только необходимую информацию.

В штатном режиме интерфейс пользователя включает окно ввода ЭКГ с минимальным количеством настраиваемых параметров, в частности продолжительности измерения (рис. 6.23)

В окне вывода результатов обработки отображаются фильтрованная ЭКГ с метками разделения на циклы и выделенным опорным циклом, фазовая траектория ЭКГ, эталонный цикл во временной области с метками выделенных фрагментов (P , QRS , $ST-T$), а также в увеличенном масштабе фрагмент усредненной фазовой траектории ЭКГ, соответствующий зубцу T (рис. 6.24).



Рис. 6.25. Графический индикатор результатов обработки

Окончательный результат обработки ЭКГ отображается на специальном графическом индикаторе в виде градусника (рис. 6.25), который показывает текущее и среднее (по данным архива) значения признака β_T , полученные в результате обработки ЭКГ конкретного пациента. Шкала индикатора трехцветная. Зеленый цвет сигнализирует о принадлежности показателя зоне «норма» ($\beta_T < 0,7$), желтый — зоне «внимание» ($0,7 \leq \beta_T \leq 0,9$), а красный — зоне «опасность» ($\beta_T > 0,9$).

К тому же результат тестирования озвучивается соответствующим голосовым сообщением. В таком режиме обслуживание комплекса вполне доступно среднему медицинскому работнику.

Естественно, что врач может получать и более полную информацию о пациенте. Интерфейс комплекса имеет окна, в которых помимо основного признака β_T отображаются и другие признаки ЭКГ, в частности частота сердечных сокращений (ЧСС), вариабельность ЧСС, причем отклонение от нормы каждого из признаков оценивается отдельно и сигнализируется соответствующим цветом. Существует также окно результатов статистической обработки, в котором помимо текущих значений отображаются средние значения диагностических признаков, в том числе при различных состояниях испытуемого и внешней среды.

В системе предусмотрена также возможность удобной настройки основных алгоритмов обработки сигнала (рис. 6.26). Однако, в отличие от основных интерфейсных окон, доступ к управлению параметрами настройки ограничен паролем, что исключает вмешательство в эти важные функции системы со стороны неквалифицированных пользователей.

6.3.7. Практические результаты

Первый образец системы, реализующей ИТ «ФАЗАГРАФ», был использован по прямому назначению — для оценки текущего функционального состояния операторов химических процессов во время трудовой деятельности.

Информационная технология также применялась при выполнении биофизических экспериментов, проводимых Институтом земного магнетизма, ионосферы и распространения радиоволн РАН (Россия, г. Троицк). По данным, полученным в ре-

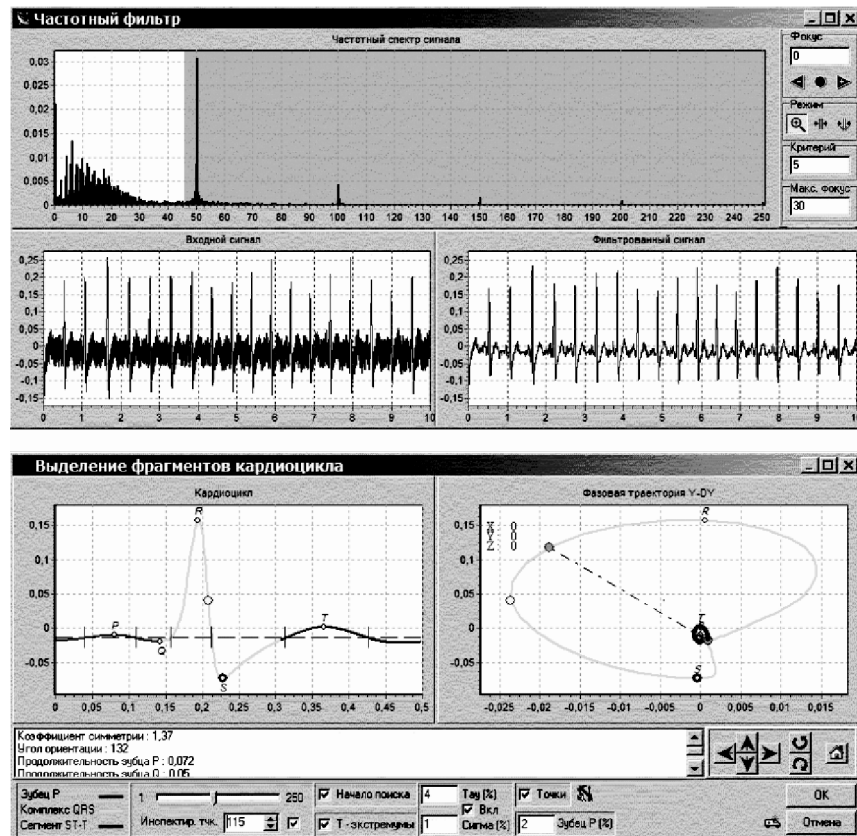


Рис. 6.26. Интерфейсы настройки вычислительных модулей:

а — частотно-избирательная фильтрация; б — выделение информативных фрагментов цикла ЭКГ

зультате длительного эксперимента с участием постоянной группы здоровых добровольцев, среднее значение параметра β_T симметрии фрагмента усредненной фазовой траектории изменялось более чем в 2 раза при высокой магнитной активности солнца [51]. Система успешно использовалась также при оценке функционального состояния спортсменов сборной Украины по стрельбе из лука и в других организациях.

6.4. Информационная технология «ИКАР»

6.4.1. Общая характеристика информационной технологии «ИКАР»

ИТ «ИКАР» [224, 283] является результатом дальнейшего развития информационной технологии «ФАЗАГРАФ». Основное преимущество ИТ «ИКАР» состоит в том, что в ней реализованы алгоритмы обработки сигнала в реальном масштабе времени. Это дает возможность проводить анализ сигналов, зарегистрированных в течение большого промежутка времени.

Обработка записей ЭКГ в реальном масштабе времени позволила не только проводить оценку функционального состояния сердца на основе морфологического анализа сигнала по комплексу диагностических признаков во временной области и в фазовом пространстве, но и оценивать баланс возбуждающего (симпатического) и тормозного (парасимпатического) отделов вегетативной нервной системы на основе анализа вариабельности сердечного ритма.

Известно, что текущая активность симпатического и парасимпатического отделов — результат реакции системы регуляции кровообращения на изменяющиеся условия окружающей среды. При этом адаптационные реакции организма индивидуальны. К тому же у одного и того же человека в разное время суток под действием физических и эмоциональных нагрузок могут происходить те или иные изменения баланса симпатического и парасимпатического отделов вегетативной нервной системы.

Анализом вариабельности сердечного ритма (ВСР) начали активно заниматься в начале 60-х годов прошлого века [15]. В настоящее время большинство исследователей пользуются предложенными в 1996 г. стандартами измерений, физиологической интерпретации ВСР и рекомендациями по клиническому использованию этого метода [300].

Анализ variability сердечного ритма основан на распознавании и измерении временных интервалов между R -зубцами ЭКГ (R — R -интервалов), построении динамических рядов этих кардиоинтервалов и на последующем анализе полученных числовых рядов во временной и частотной областях.

Статистический анализ variability сердечного ритма сводится к выделению нормальных R — R -интервалов (с исключением экстрасистол и артефактов), которые в пульсометрии принято называть NN -интервалами, и определению по этим интервалам таких показателей [111]:

- $SDNN$ (СКО) — среднее квадратическое отклонение NN -интервалов;
- $RMSSD$ — квадратный корень из суммы квадратов разности величин последовательных пар NN -интервалов;
- $NN50$ — количество пар последовательных NN -интервалов, различающихся более чем на 50 миллисекунд, полученное за весь период записи;
- $PNN50$ — процент $NN50$ от общего количества NN -интервалов;
- CV — коэффициент вариации, представляющий собой нормированную оценку СКО:

$$CV = \text{СКО} / M * 100,$$

где M — среднее значение NN -интервалов;

- D — СКО в квадрате, отражает суммарную мощность всех периодических и непериодических колебаний;
- Mo (мода) — наиболее часто встречающееся в данном динамическом ряде значение NN -интервала;
- AMo (амплитуда моды) — число NN -интервалов, соответствующих значению моды (в % к объему выборки);
- $MxDMn$ (вариационный размах) — вычисляется по разности максимального (Mx) и минимального (Mn) значений NN -интервалов (этот показатель при аритмиях может быть искажен);
- $ИН$ (стресс-индекс) — индекс напряжения регуляторных систем по Баевскому [15], определяемый по формуле

$$ИН = AMo / 2Mo * MxDMn;$$

Помимо статистических показателей при анализе ВСП также определяют следующие частотные компоненты спектра мощности:

- HF — спектр в частотном диапазоне 0,4—0,15 Гц;
- LF — спектр в частотном диапазоне 0,15—0,04 Гц.

ТАБЛИЦА 6.5

Физиологическая интерпретация показателей ВСП

№	Обозначение	Наименование показателя	Физиологическая интерпретация
1	<i>SDNN</i>	Стандартное отклонение полного массива кардиоинтервалов	Суммарный эффект вегетативной регуляции кровообращения
2	<i>RMSSD</i>	Квадратный корень из суммы разностей последовательного ряда кардиоинтервалов	Активность парасимпатического звена вегетативной регуляции
4	<i>PNN50</i>	Число пар кардиоинтервалов с разностью более 50 мс (в % к общему числу кардиоинтервалов в массиве)	Показатель степени преобладания парасимпатического звена регуляции над симпатическим (относительное значение)
5	<i>CV</i>	Коэффициент вариации полного массива кардиоинтервалов	Нормированный показатель суммарного эффекта регуляции
6	<i>MxDMn</i>	Разность между максимальным и минимальным значениями кардиоинтервалов	Максимальная амплитуда регуляторных влияний
7	<i>Mo</i>	Мода	Наиболее вероятный уровень функционирования сердечно-сосудистой системы
8	<i>AMo</i>	Амплитуда моды	Условный показатель активности симпатического звена регуляции
9	<i>SI</i>	Стресс-индекс (индекс напряжения регуляторных систем)	Степень напряжения регуляторных систем (степень преобладания активности центральных механизмов регуляции над автономными)
13	<i>HF, %</i>	Относительная мощность спектра высокочастотной компоненты ВСП	Относительный уровень активности парасимпатического звена регуляции
14	<i>LF, %</i>	Относительная мощность спектра низкочастотной компоненты ВСП	Относительный уровень активности вазомоторного центра

Эти характеристики используются далее для вычисления индекса вагосимпатического взаимодействия, представляющего собой отношение LF/HF .

Основные показатели ВСР и их физиологическая интерпретация [111] представлены в табл. 6.5.

На практике используются также геометрические методы визуализации ВСР: гистограмма нормальных R — R -интервалов с шагом 50 мс и скаттерограмма, которая представляет собой изображение последовательности наблюдаемых кардиоинтервалов в координатах $(R - R)_{i+1}$, $(R - R)_i$.

Следует отметить, что в литературе довольно часто содержатся противоречивые данные о значениях указанных выше показателей в норме и патологии. Поэтому приведем лишь некоторые сведения из многочисленных отечественных и зарубежных публикаций, посвященных анализу ВСР.

Нормальные значения СКО находятся в пределах 40—80 мс. Однако эти значения имеют возрастно-половые особенности, которые должны учитываться при оценке результатов исследования. При анализе коротких записей, как правило, увеличение СКО указывает на усиление автономной регуляции, т. е. усиление влияния дыхания на ритм сердца, что чаще всего наблюдается во сне. Уменьшение СКО связано с усилением симпатической регуляции, которая подавляет активность автономного контура.

Резкое снижение СКО обусловлено значительным напряжением регуляторных систем, когда в процесс регуляции включаются высшие уровни управления, что ведет к почти полному подавлению активности автономного контура.

Показатель активности парасимпатического звена вегетативной регуляции $RMSSD$ вычисляется по динамическому ряду разностей значений последовательных пар кардиоинтервалов. Чем выше значение $RMSSD$, тем активнее звено парасимпатической регуляции. В норме значения этого показателя находятся в пределах 20—50 мс.

Индекс напряжения регуляторных систем (ИН), характеризующий активность механизмов симпатической регуляции, в норме колеблется в пределах 80—150 усл. ед. Этот показатель чрезвычайно чувствителен к усилению тонуса симпатической нервной системы. Небольшая нагрузка (физическая или эмоциональная) увеличивает ИН в 1,5—2 раза, а значительная нагрузка — в 5—10 раз. У больных с постоянным напряжением регуляторных систем ИН в покое равен 400—600 усл. ед.

Дыхательная составляющая (HF) обычно составляет 15—25 % суммарной мощности спектра. Снижение этого показателя до 8—10 % указывает на смещение вегетативного баланса в сторону

преобладания симпатического отдела. Если же значение HF ниже 2—3 %, то можно говорить о резком преобладании симпатической активности. В этом случае существенно уменьшаются также показатели $RMSSD$ и $PNN50$. Мощность низкочастотной составляющей спектра (LF) обычно в норме колеблется от 15 до 35—40 %. Ее увеличение может указывать на снижение активности вазомоторного центра или на замедление барорефлекторной регуляции.

6.4.2. Интерфейс пользователя

Укрупненная схема последовательности процедур, реализующих алгоритм обработки ЭКГ в ИТ «ИКАР», показана на рис. 6.27.

Процесс ввода и обработки ЭКГ отображается в окнах программы, выполненных в виде трех закладок. В первой закладке (рис. 6.28) отображается исходный сигнал, фильтрованный сигнал, а также результат морфологического анализа сигнала (индикация ненадежных циклов и отображение характерных точек P , Q , R , S , T каждого цикла).

После обнаружения опорного цикла программа автоматически переходит на вторую закладку (рис. 6.29). На этой закладке в реальном масштабе времени отображается фазовая траектория сигнала и ее фрагмент, соответствующий периоду деполяризации, строится график, визуализирующий динамику изменения симметрии фрагмента деполяризации фазовой траектории (основной диагностический признак), а также график построенного среднего цикла во временной области.

Кроме того, на второй закладке отображаются числовые значения параметров, которые традиционно используются в кардиологической практике — продолжительности и амплитуды сегментов усредненного цикла.

И наконец, после завершения ввода сигнала на третьей закладке программы (рис. 6.30) отображаются результаты анализа variability сердечного ритма в виде ритмограммы, скатерограммы, гистограммы, спектрограммы и графика баланса мощности. На этой же закладке отображаются результаты автоматических вычислений показателей статистического и спектрального анализов ВСР.

В ИТ «ИКАР» реализовано два способа графического отображения результатов интерпретации ЭКГ. Первый способ, ориентированный на непрофессиональных пользователей, развивает

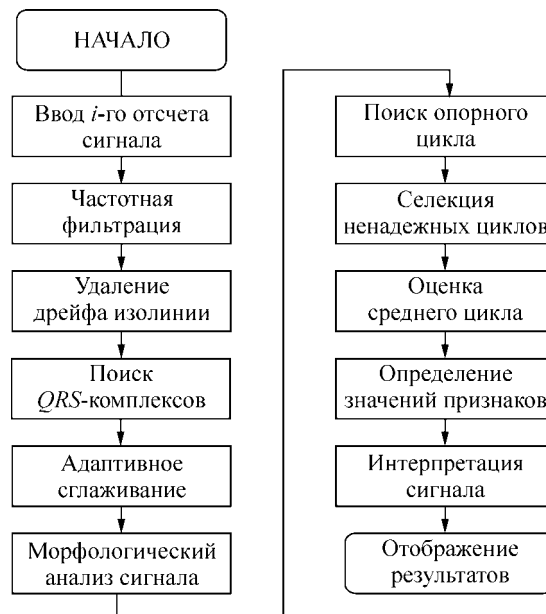


Рис. 6.27. Укрупненная схема обработки сигнала в реальном времени

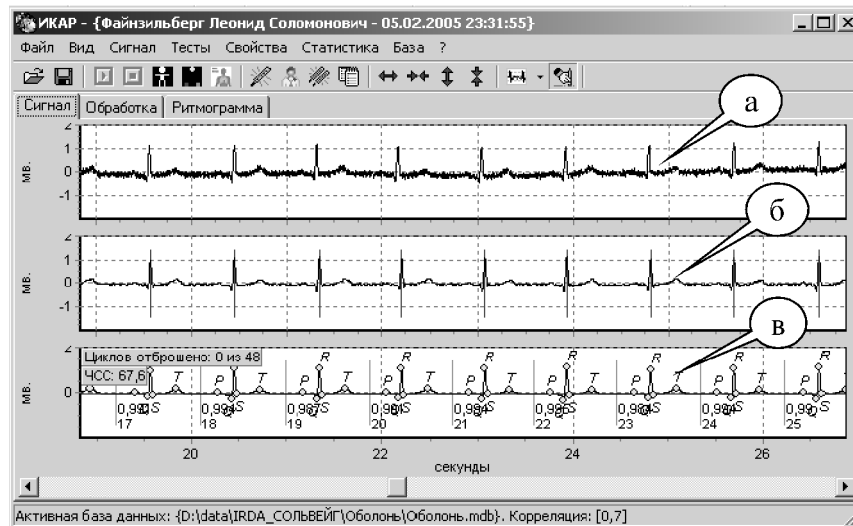


Рис. 6.28. Окно отображения ЭКГ:

a — исходный сигнал; *б* — результат фильтрации; *в* — результат морфологического анализа

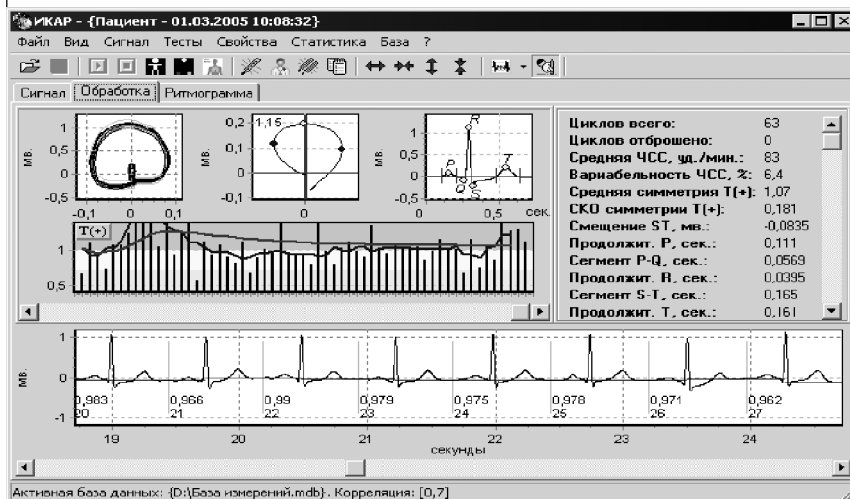


Рис. 6.29. Окно отображения результатов обработки сигнала во временной области и в фазовом пространстве

метод отображения результатов, реализованный в ИТ «ФАЗАГРАФ». Различие состоит в том, что помимо градусника, индицирующего текущее значение функционального состояния сердца (на основе анализа симметрии участка реполяризации), дополнитель-

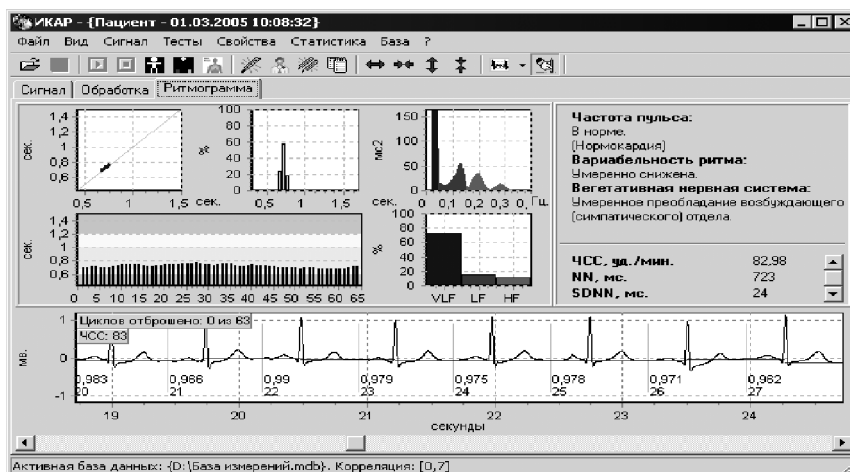


Рис. 6.30. Окно отображения результатов анализа ВСР



Рис. 6.31. Индикатор функционального состояния

но отображаются результаты анализа сердечного ритма (частота пульса, вариабельность сердечного ритма) и результаты анализа симпато-вагусного баланса (рис. 6.31).

Более подробную информацию о результатах обработки можно получить, используя второй способ отображения в виде когнитивного графического образа (рис. 6.32). Этот способ позволя-

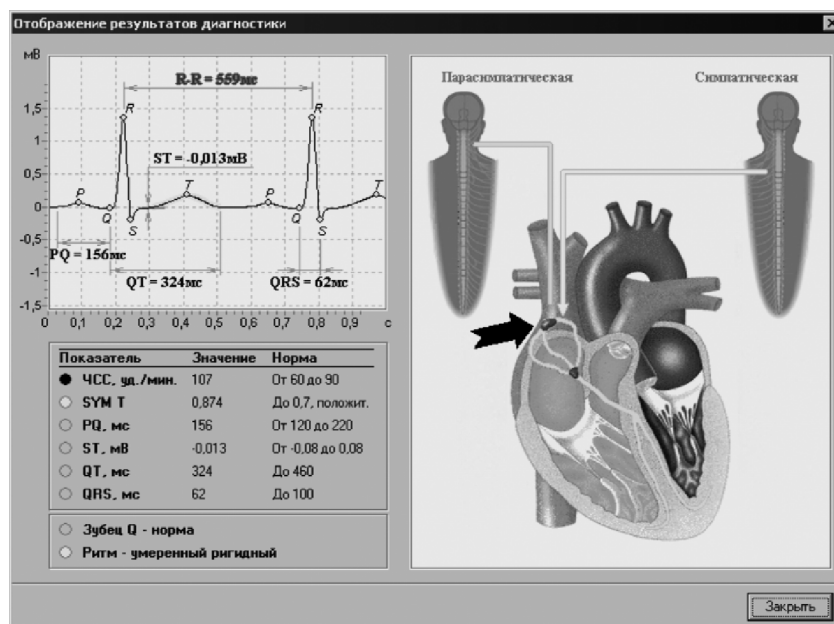


Рис. 6.32. Отображение результатов интерпретации сигнала в виде когнитивного графического образа

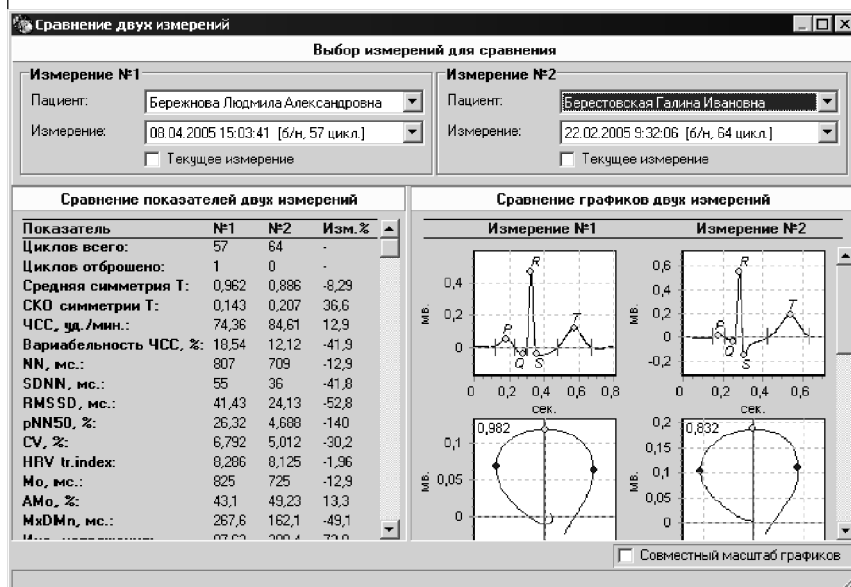


Рис. 6.33. Результат сравнения двух измерений

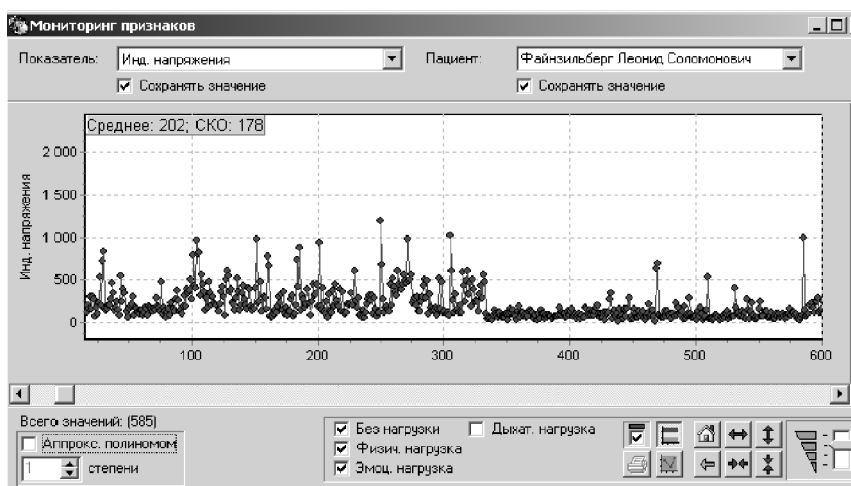


Рис. 6.34. Результат отображения динамики изменения показателя

ет на графических картинках сердца и отделов вегетативной системы «увидеть» обнаруженные отклонения от нормы, которые индицируются желтой или красной окраской соответствующих

участков изображения. Наводя мышку на такой участок, пользователь получает звуковой комментарий, а также подсказку о возможных причинах возникновения обнаруженного отклонения.

В ИТ «ИКАР» предусмотрен ряд удобных процедур, обеспечивающих анализ динамики изменения показателей функционального состояния пациента, в частности, программа отображения в специальном окне результатов сравнения двух ЭКГ, сохраненных в базе данных, или сравнения текущей ЭКГ с ранее сохраненной (рис. 6.33).

Кроме того, можно воспользоваться процедурой отображения динамики изменения любого выбранного показателя за весь период наблюдения конкретного пациента, обследования которого сохранены в базе данных (рис. 6.34).

ИТ «ИКАР» позволяет также проводить статистическую обработку накопленных данных и отображать результаты такой обработки в удобной форме.

6.4.3. Автономное микропроцессорное устройство «ИКАР-2006»

Некоторые из элементов предложенной информационной технологии удалось реализовать в автономном микропроцессорном устройстве «ИКАР-2006».

Устройство «ИКАР-2006» представляет собой малогабаритный прибор. Управление и основные функции устройства реализованы на микроконтроллере серии C8051FXXX фирмы Silicon Laboratories.

Устройство имеет автономный источник питания (стандартная литий-ионная аккумуляторная батарея) и внутренний жидкокристаллический индикатор (ЖКИ) с подсветкой (разрешение 128 × 64 пикселя). В специальный отсек устройства вставляется съемная флеш-карточка памяти, емкость которой позволяет записать не менее 180 час непрерывной записи ЭКГ.

Регистрация ЭКГ первого стандартного отведения осуществляется непосредственно с пальцев рук испытуемого (рис. 6.35). Предусмотрен также ввод сигнала из любого грудного отведения с помощью внешнего кабеля и одноразовых электродов.

Процесс ввода ЭКГ отображается на внутреннем экране устройства (рис. 6.36). С помощью пятикнопочного навигатора (джойстика), расположенного на передней панели устройства, пользова-

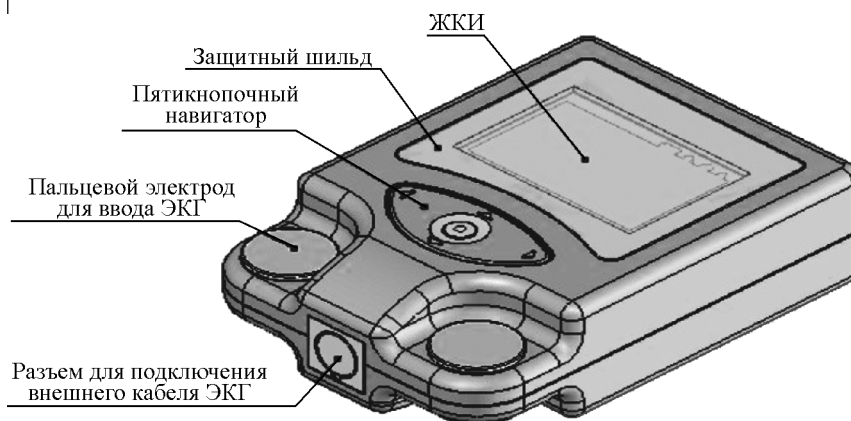


Рис. 6.35. Внешний вид микропроцессорного устройства «ИКАР-2006»



Рис. 6.36. Процесс регистрации ЭКГ с пальцев пользователя

тель может установить требуемый масштаб вывода графика ЭКГ по оси абсцисс и ординат.

Внутреннее программное обеспечение устройства позволяет автоматически вычислять: частоту сердечных сокращений (ЧСС), вариабельность ЧСС в %, продолжительность интервалов $P-Q$, комплекс QRS , смещение сегмента ST относительно изоэлектрической линии и другие традиционные признаки ЭКГ. Кроме того, оценивается значение оригинального признака β_T — симметрии участка реполяризации фазовой траектории.

Вычисленные значения диагностических признаков отображаются на индикаторе в виде градусника на внутреннем экране устройства (рис. 6.37).

В центральной части индикатора выводится название показателя, единицы измерения и рассчитанное числовое значение показателя (в приведенном примере ЧСС равна 75 уд./мин). В верхней части индикатора выводится одно из текстовых сообщений: «Норма», «Удовлетворительно» (несущественные отклонения от нормы) или «Внимание» (значительные отклонения вычисленного значения показателя от нормы).

Для лучшего зрительного восприятия в центре индикатора выводится мигающая метка, имитирующая ртутный столбик в градуснике, движущийся по шкале в соответствии с вычисленным значением показателя (см. рис. 6.37). Бесцветное поле шкалы символизирует нормальный диапазон значений показателя, серое поле — несущественный выход показателя за пределы нормы, черное — значительное отклонение показателя от нормы.

Внутреннее программное обеспечение устройства «ИКАР-2006» реализует иерархическую систему меню, с помощью которой пользователь имеет возможность выбрать и настроить нужный режим работы устройства. Один из таких режимов предусматривает автоматическую запись информации из грудных отведений по сигналам внутреннего таймера. При этом пользователь может задать требуемое число циклов автоматической записи, интервалы измерения и паузы между измерениями.

В устройстве реализован двусторонний обмен данными с персональным компьютером, в частности передача в реальном масштабе времени пакетов цифровых данных с помощью радиоканала ZigBee (2,4 ГГц, 1 МВт). Это дает возможность более глубоко анализировать и интерпретировать регистрируемые ЭКГ с помощью внешнего программного обеспечения ИТ «ИКАР».

Естественно, информация, сохраненная на флеш-карточке, также может быть введена в персональный компьютер для дальнейшей обработки.

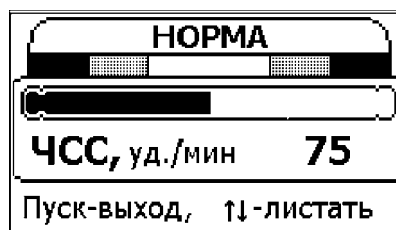


Рис. 6.37. Отображение результатов измерения (на примере ЧСС)

6.4.4. Результаты физиологических экспериментов и клинических испытаний

Для оценки эффективности ИТ «ИКАР» и экспериментальной проверки сформулированных гипотез о диагностической ценности признаков ЭКГ в фазовом пространстве проводились эксперименты на животных и медицинские испытания в клинических условиях ⁶.

Основная цель эксперимента на животных состояла в наблюдении за изменением признака β_T (симметрии участка реполяризации усредненной фазовой траектории) в условиях искусственной ишемии и реперфузии. Эксперименты проводились под руководством академика А.А. Мойбенко в Институте физиологии им. А.А. Богомольца НАН Украины.

Результаты экспериментов еще раз подтвердили гипотезу о полезности измерения β_T для выявления начальных признаков ишемической болезни сердца: как только животному проводилась манипуляция, вызывающая искусственную ишемию, значения признака β_T увеличивались, постепенно приближаясь к порогу «опасных» (рис. 6.38).

Клинические испытания ИТ «ИКАР» также подтвердили сформулированную гипотезу. Эксперименты проводились в Национальном научном центре «Институт кардиологии им. акад. Н.Д. Стражеско».

На рис. 6.39 показана динамика изменения признака β_T в процессе эксперимента с велоэргометрической пробой, в котором участвовал здоровый доброволец. Видно, что значения признака β_T увеличивались по мере увеличения нагрузки, а затем снижались в период отдыха (реституции).

Интересные результаты получены после статистической обработки экспериментальных данных испытаний ИТ «ИКАР» в инфарктном отделении клинической больницы г. Киева. Поэтому остановимся на этих результатах более подробно [236].

Целью экспериментов было изучение взаимосвязи между параметрами ЭКГ в фазовом пространстве и клинико-функциональными признаками динамики состояния больных с острым коронарным синдромом. Обследован 41 больной (29 мужчин и

⁶ Методическое руководство экспериментами осуществлял И.А. Чайковский.

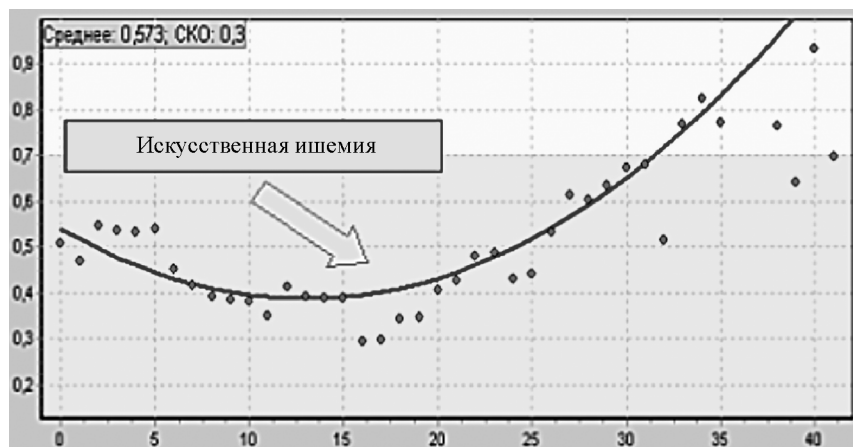


Рис. 6.38. Результаты эксперимента на собаке

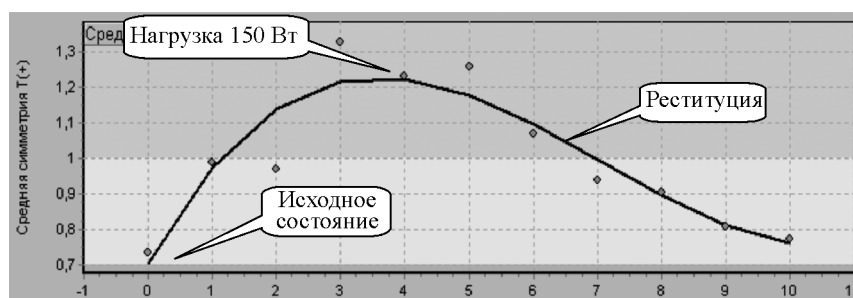


Рис. 6.39. Результаты эксперимента с велоэргометрической пробой

12 женщин), находящиеся в отделении кардиореанимации с различными признаками острого коронарного синдрома (табл. 6.6).

Результат лечения этих больных был следующим: 34 человека были переведены на этап реабилитации с явным клиническим улучшением (группа 1). У семи больных течение острого коронарного синдрома было неблагоприятным, в том числе 1 больной умер (группа 2).

Результаты статистической обработки данных, полученных в процессе эксперимента, представлены в табл. 6.7.

Оказалось, что в 88 % случаев у больных 1-й группы после лечения симметрия зубца *T* уменьшилась в среднем на 25 %, в то время как у больных 2-й группы (с неблагоприятным резуль-

ТАБЛИЦА 6.6

Основные характеристики обследованной группы

Количество обследованных	41
Средний возраст	62 ± 11
Больные с острым Q-инфарктом миокарда	18
Больные с острым неQ-инфарктом миокарда	9
Больные с инфарктом миокарда передней локализации	14
Больные с инфарктом миокарда задней локализации	13
Больные с нестабильной стенокардией	14

ТАБЛИЦА 6.7

Значения показателей ЭКГ до и после лечения

Показатель		1-я группа (n = 34)	2-я группа (n = 7)
β_T	До лечения	1,02 ± 0,57	0,87 ± 0,07
	После лечения	0,76 ± 0,43	1,09 ± 0,13
	Прирост	-0,25 ± 0,08	0,22 ± 0,05
ST, мВ	До лечения	-0,03 ± 0,12	0,01 ± 0,2
	После лечения	-0,005 ± 0,1	0,02 ± 0,17
	Прирост	0,024 ± 0,07	0,01 ± 0,09
T, мВ	До лечения	0,16 ± 0,19	0,12 ± 0,08
	После лечения	0,18 ± 0,14	0,07 ± 0,2
	Прирост	0,02 ± 0,01	-0,05 ± 0,02
R, мВ	До лечения	0,79 ± 0,2	0,81 ± 0,14
	После лечения	1,03 ± 0,31	0,98 ± 0,27
	Прирост	0,22 ± 0,10	0,17 ± 0,11
QT, с	До лечения	0,41 ± 0,05	0,4 ± 0,04
	После лечения	0,41 ± 0,06	0,41 ± 0,06
	Прирост	0 ± 0,03	0,01 ± 0,04
ЧСС, уд./мин	До лечения	63 ± 11	69 ± 16
	После лечения	65 ± 9	61 ± 13
	Прирост	2 ± 2	-7 ± 4

татом лечения) отмечается тенденция к увеличению симметрии волны T . Анализ возрастания симметрии зубца T выявил статистически достоверные различия между группами 1 и 2 ($P < 0,001$).

На рис. 6.40 представлена динамика изменений показателя β_T у больного Ч. с диагнозом «нестабильная стенокардия (инфаркт в анамнезе), кардиогенный шок, отек легких».

Как видно из рисунка, в первые дни лечения наблюдалось уменьшение значений β_T . Однако начиная с 5-х суток госпита-

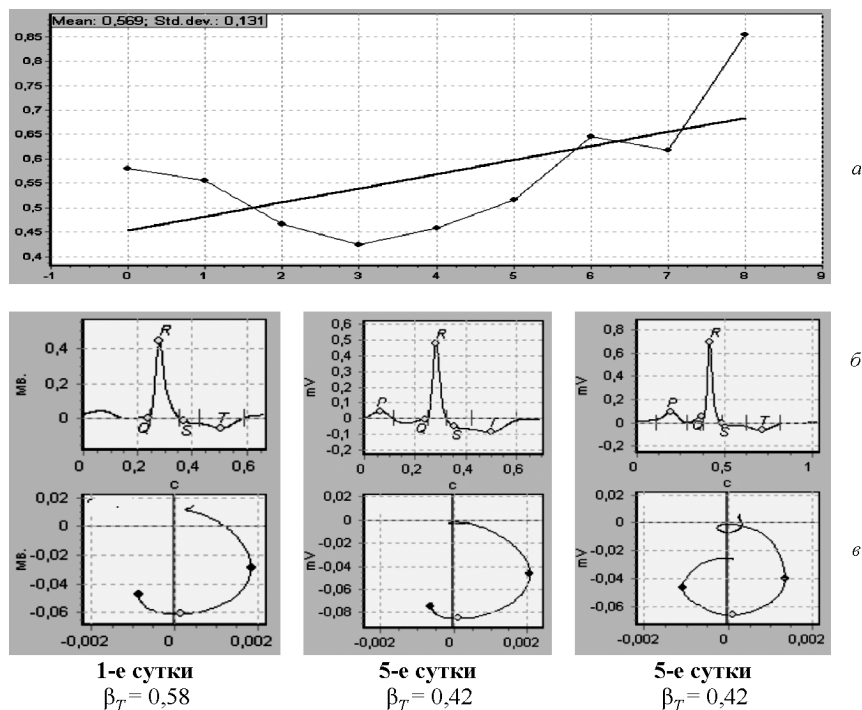


Рис. 6.40. Динамика изменений показателя β_T у больного Ч:

а — тренд показателя β_T ; б — усредненный ЭКГ цикл; в — фрагменты фазовой траектории зубца Т

лизации клиническое улучшение прекратилось. С 10-х суток госпитализации усилился болевой синдром, уменьшилась толерантность к физической нагрузке, усилилась одышка. При измерениях электрокардиограммы в фазовом пространстве было зарегистрировано парадоксальное увеличение показателя значений β_T . Через 14 суток после госпитализации больной умер. Важно отметить, что существенное увеличение параметра β_T , характеризующего симметрию зубца Т ЭКГ в фазовом пространстве, предшествовало ухудшению состояния больного Ч. В то же время, начиная с 3-х суток госпитализации при визуальном анализе ЭКГ в 12-ти отведениях этого больного не было отмечено изменений.

В процессе экспериментов установлена также корреляция средней степени между уровнем активности ферментов некроза

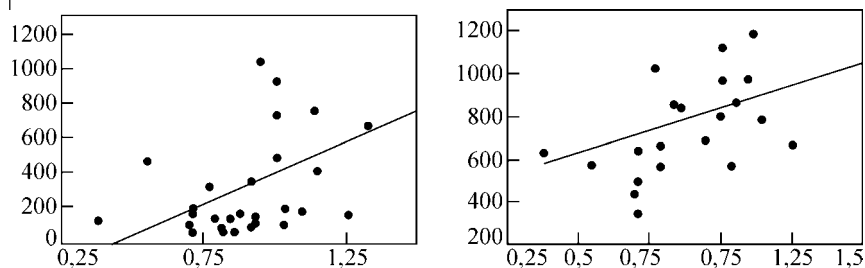


Рис. 6.41. Взаимосвязь величин КФК (слева) и ЛДГ (справа) с β_T

КФК и ЛДГ и значениями β_T ($r = 0,54$ и $r = 0,46$ соответственно) (рис. 6.41).

На основании полученных результатов можно сделать вывод, что предложенная информационная технология анализа ЭКГ в фазовом пространстве является не только диагностически эффективной для профилактической медицины и домашнего применения, но и может быть полезной для клинического применения в качестве удобного инструмента, позволяющего оптимизировать наблюдение больных с острым коронарным синдромом.

6.5. Информационная технология «MAGWIN»

6.5.1. Общая характеристика ИТ «MAGWIN»

В последнее время для исследования биоэлектрических процессов, происходящих в сердце человека, все большее признание получает метод магнитокардиографии (МКГ), основанный на измерении параметров магнитного поля над грудной клеткой пациента. Хотя МКГ и ЭКГ порождаются одними и теми же источниками, считается, что при регистрации магнитного поля сердца потенциально возможно получить более полную информацию о различных поражениях миокарда.

Информативность метода обусловлена его высокой чувствительностью к тангенциальным компонентам волны возбуждения сердца и меньшей (по сравнению с поверхностной ЭКГ) зависимостью результатов измерения от многослойной проводящей среды тела пациента [157].

Для регистрации МКГ используют измерительные системы, построенные на основе сверхпроводящих квантовых интерферометрических датчиков (СКВИД). К такому классу относится из-

мерительная система, разработанная сотрудниками Института кибернетики им. В.М. Глушкова НАН Украины под руководством чл.-кор. И.Д. Войтовича [367].

Следует отметить, что измерение слабых полей магнитного поля сердца — лишь одна из задач, решение которых необходимо для практической реализации метода МКГ. В отличие от ЭКГ, которая, вообще говоря, допускает визуальную интерпретацию, обработка МКГ невозможна без применения современных компьютерных средств.

Дело в том, что первичные данные, формируемые магнитометрическими измерительными системами, не могут быть непосредственно использованы для медицинского заключения. Связано это не только с наличием большого уровня шума, но и с тем, что магнитное поле является уже вторичным фактором, обусловленным протеканием электрических токов в сердце человека. Более того, сам характер распределения тока связан с источниками этого тока и наличием или отсутствием неоднородной проводимости ткани.

Для нахождения распределения плотности тока по известному магнитному полю необходимо решить обратную задачу теории стационарного поля, которая, как известно, не имеет однозначного решения. Поэтому задача решается при определенных предположениях о распределении тока на плоской поверхности (двумерные карты распределения тока), хотя в некоторых исследованиях используются и более сложные модели с трехмерным распределением.

Не менее сложной проблемой является построение эффективных вычислительных алгоритмов, позволяющих интерпретировать карты плотностей тока. К сожалению, знания, имеющиеся в предметной области, пока еще недостаточны для построения алгоритмов, основанных на точных («физических») моделях растекания тока в здоровых и патологических областях миокарда. Поэтому для построения алгоритмов диагностики не остается ничего иного, как вводить разумные гипотезы о качественных характеристиках структуры растекания тока в проводящей среде и проверять эти гипотезы на имеющемся клиническом материале.

Такой подход и реализован в ИТ «MAGWIN», разработанной под руководством автора по заказу фирмы SQUID AG (Германия) [273, 284, 349]. На рис. 6.42 показана структура программной системы ⁷, в которой реализована эта технология.

⁷ Программная реализация ИТ «MAGWIN» осуществлена Н.В. Прогонным и Д.В. Иващенко.

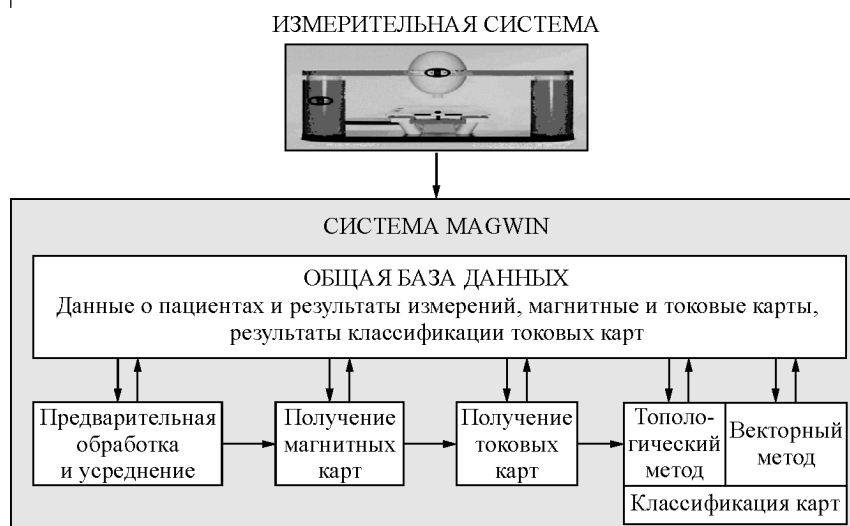


Рис. 6.42. Упрощенная структура системы «MAGWIN»

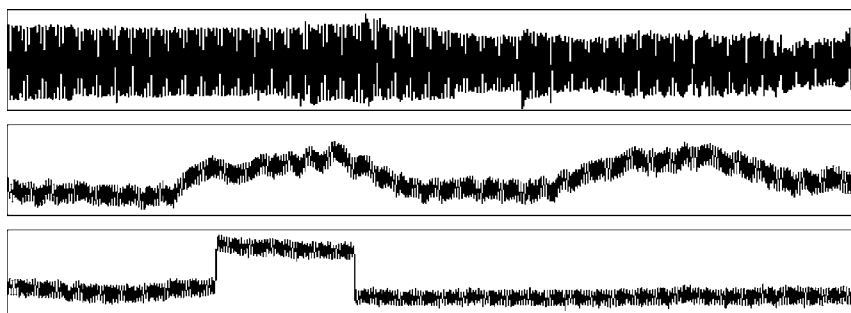


Рис. 6.43. Исходные МКГ, поступающие на вход ИТ «MAGWIN»

На вход ИТ поступают исходные МКГ, измеряемые с частотой квантования 100 Гц в 36 фиксированных точках над поверхностью грудной клетки пациента в течение 30—60 с каждая. Поскольку измерения производятся в реальных клинических условиях (без использования дорогостоящего экранированного помещения), исходные сигналы (сырые МКГ) существенно искажены различными помехами (рис. 6.43).

Для предварительной обработки сигналов использован каскадный фильтр подавления аддитивных искажений. Помимо про-

блемы подавления «традиционных» сетевых помех, в клиниках Германии пришлось столкнуться с дополнительной проблемой подавления гармонической помехи, возникающей на частоте $f_h \approx 16,7$ Гц.

Применение узкополосного заградительного фильтра, построенного на основе прямого и обратного ДПФ с привлечением предложенной поисковой процедуры уменьшения растекания спектра помехи, позволило успешно справиться и с этой задачей, несмотря на то, что частота помехи $f_h \approx 16,7$ Гц располагалась в области «полезных» частот МКГ.

Потребовалось также применить и ряд других процедур, рассмотренных в гл. 4, в частности, процедуру подавления резких скачков сигнала и дрейфа изолинии, а также адаптивного сглаживания, обеспечивающего уменьшение случайной шумовой компоненты. После фильтрации внешних аддитивных помех проводилась оценка эталонных циклов с использованием предложенного метода усреднения сигналов в фазовом пространстве координат. В результате удалось получить приемлемые оценки эталонных циклов МКГ в каждой из 36 точек измерения (рис. 6.44).

Подсистема получения магнитных карт реализует алгоритмы двумерной интерполяции по дискретным значениям усредненных циклов МКГ во всех пространственных точках измерения (рис. 6.45). Степень насыщенности серого цвета определяется значениями параметров магнитного поля сердца человека.

Интерфейс подсистемы позволяет пользователю выбрать для последующего анализа необходимый интервал в пределах цикла МКГ, в частности S – T -фрагмент.

Следует отметить, что без получения магнитных карт, которые являются результатом предварительной обработки и усреднения исходных сигналов, невозможно было бы получить и токовые карты, непосредственно используемые для диагностики.

При реализации системы использован модульный принцип построения основных компонент и универсальный интерфейс (4.4) взаимодействия основной программы с внешними вычислительными процедурами. Благодаря этому в системе удалось достаточно просто интегрировать не только собственные вычислительные модули, но и вычислительные процедуры, разработанные другими исполнителями.

В подсистеме получения токовых карт использован внешний программный модуль⁸, обеспечивающий решение обратной за-

⁸ Внешний программный модуль разработан С.Л. Романович.

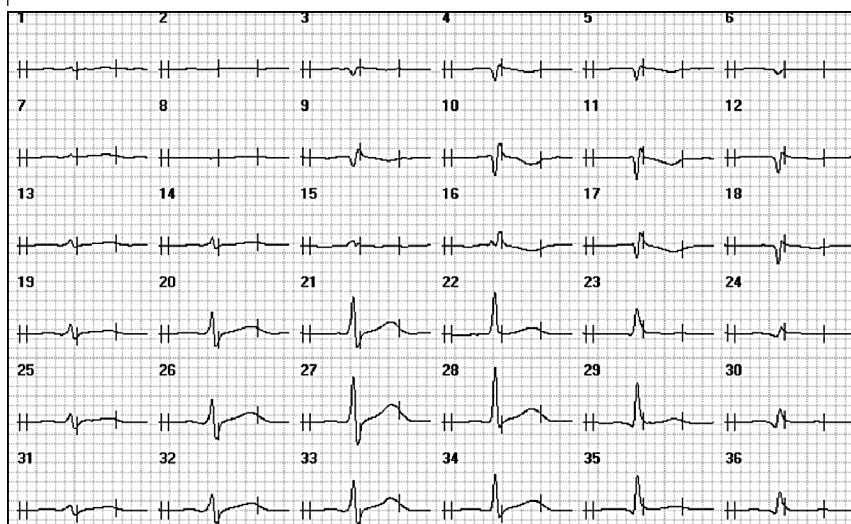


Рис. 6.44. Эталонные циклы МКГ, восстановленные в 36 точках измерения

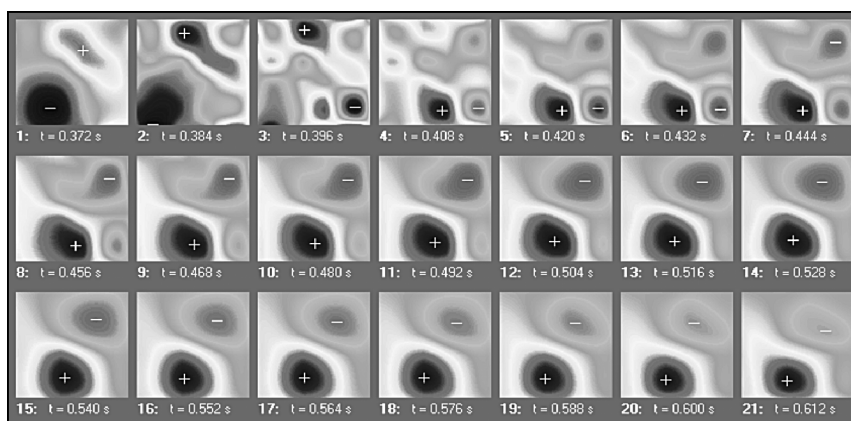


Рис. 6.45. Карты магнитного поля здорового пациента на $S-T$ -сегменте

дачи по методу С.С. Романовича [356]. Входными данными модуля являются последовательности матриц значений магнитного поля в заданном временном интервале, которые экспортируются основной программой ИТ «MAGWIN», а выходными — последовательность матриц значения плотностей тока на дискретном множестве точек плоскости сердца.

Полученные данные импортируются основной программой, сохраняются в общей БД и используются для последующей интерпретации.

6.5.2. Алгоритм интерпретации, основанный на классификации карт плотностей тока

Для построения алгоритмов диагностики проводились многочисленные исследования результатов обработки МКГ верифицированных пациентов с ишемической болезнью сердца (ИБС) и здоровых добровольцев. Такие исследования стали еще одним убедительным примером применения идей когнитивной компьютерной графики к формированию гипотез в мало изученной предметной области.

Исследования заключались в визуальной оценке врачами-кардиологами карт плотностей тока, «порождаемых» компьютером при регистрации МКГ представителей указанных групп. Изучение карт, соответствующих периоду реполяризации миокарда (фрагменту $S-T$), позволило сформировать предварительное представление о топологии карты здорового человека и больных пациентов⁹.

«Идеальной» считается карта распределения тока (рис. 6.46, *a*), которая имеет дипольную структуру с двумя симметричными вихрями тока, созданными в однородной электропроводной среде одним фронтом источников, направленным влево вниз по отношению к пациенту (ориентировочно в пределах углов от -10° до -80° , отсчитываемых против часовой стрелки от горизонтальной оси). Такие карты наблюдались у большинства здоровых пациентов.

Отметим, что карты пациентов, страдающих ИБС, демонстрируют картину распределения токов, которая не имеет дипольной структуры. В этом случае вихри тока деформируются, становятся несимметричными, их количество возрастает, причем области с относительно высокой плотностью тока часто образуют отдельные кластеры, ориентированные в различных направлениях (рис. 6.46, *б*).

Для автоматической интерпретации результатов МКГ исследований разработан алгоритм классификации карт плотностей тока, который формирует числовой показатель $\bar{C}_{ST} \in [0, 1]$, интегрально оценивающий отклонения карт от «идеальной» на участке реполяризации сердечного цикла [38, 284].

Упрощенная схема алгоритма классификации карт, реализованного в одном из модулей системы «MAGWIN», состоит в следующем.

⁹ Неформальное описание таких карт предоставил автору И.А. Чайковский.

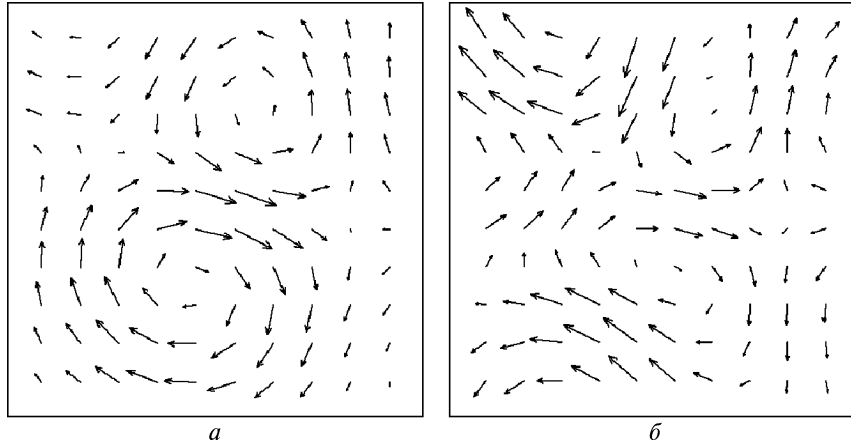


Рис. 6.46. Карты плотностей тока:

a — здорового человека; b — пациента, страдающего ИБС

На вход алгоритма поступает последовательность карт $\psi_1, \dots, \psi_K$ плотностей тока, построенных для S — T -фрагмента сердечного цикла. Каждая карта представляет собой $N \times N$ ($N = 10$) матрицу $\psi_k = \|\bar{v}_{ij}\|$ векторов $\bar{v}_{i,j} = (x_{ij}, y_{ij})$ плотностей тока, заданных своими проекциями x_{ij}, y_{ij} .

Обработка каждой k -й карты сводится к последовательности шагов.

Шаг 1. Преобразование координат $\bar{v}_{i,j} = (x_{ij}, y_{ij}) \Rightarrow \bar{v}_{i,j} = (l_{ij}, \alpha_{ij})$, где $l_{ij} \in [0, 1]$ — модуль; $\alpha_{ij} \in [0, 2\pi]$ — направление:

$$l_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2},$$

$$\alpha_{ij} = \begin{cases} \arctg(y_{ij} / x_{ij}), & \text{если } x_{ij} > 0, y_{ij} \geq 0, \\ \arctg(y_{ij} / x_{ij}) + 2\pi, & \text{если } x_{ij} > 0, y_{ij} < 0, \\ \arctg(y_{ij} / x_{ij}) + \pi, & \text{если } x_{ij} < 0, \\ \frac{\pi}{2}, & \text{если } x_{ij} = 0, y_{ij} > 0, \\ \frac{3\pi}{2}, & \text{если } x_{ij} = 0, y_{ij} < 0. \end{cases}$$

Шаг 2. Определение функций принадлежности:

$$b_{i,j} = \begin{cases} 1, & \text{если } G_1^* \leq \alpha_{ij} \leq G_2^*, \\ \mu, & \text{если } G_2^* < \alpha_{ij} < R_1^* \text{ или } R_2^* < \alpha_{ij} < G_1^*, \\ 0, & \text{если } \alpha_{ij} \leq R_2^* \text{ или } \alpha_{ij} \geq R_1^*, \end{cases}$$

где $0 < R_2^* < G_1^* < G_2^* < R_1^* < 2\pi$ — константы, ограничивающие «нормальные» и «патологические» направления векторов $\vec{v}_{i,j}$, а $\mu_{ij} \in [0, 1]$ — величина, определяемая соотношением

$$\mu_{i,j} = \begin{cases} |(a_{i,j} - R_2^*) / (G_1^* - R_2^*)|, & \text{если } R_2^* < a_{i,j} < G_1^*, \\ |(a_{i,j} - R_1^*) / (G_2^* - R_1^*)|, & \text{если } G_2^* < a_{i,j} < R_1^*. \end{cases}$$

Шаг 3. Выделение кластеров, в которых группируются векторы $\vec{v}_{ij}^0$, удовлетворяющие условию

$$b_{ij} l_{ij} \geq l_0, \quad (6.36)$$

где l_0 — константа, назначаемая пользователем (по умолчанию порог l_0 задается равным 80 % модуля наибольшего из наблюдаемых векторов). Считается, что векторы принадлежат одному кластеру, если они находятся в смежных столбцах или строках матрицы ψ_k .

Шаг 4. Определение параметра λ , характеризующего долю (процент) векторов $\vec{v}_{ij}^0$, удовлетворяющих условию (6.36), числа N_{cl} кластеров, в которых они сгруппированы, и при $N_{cl} \geq 2$ вычисление наибольшего расстояния $D_{\max}$ между кластерами (рис. 6.47). Под расстоянием между двумя кластерами понимается величина $d_{lm} = \max(d_i, d_j)$, где d_i, d_j — соответственно количество строк и столбцов матрицы ψ_k , разделяющих l -й и m -й кластеры.

Шаг 5. Вычисления значения индексной переменной $C_k \in [0, 1]$, характеризующего «близость» k -й карты к «идеальной». Для упрощения расчетов переменной C_k приписывается только пять возможных градаций:

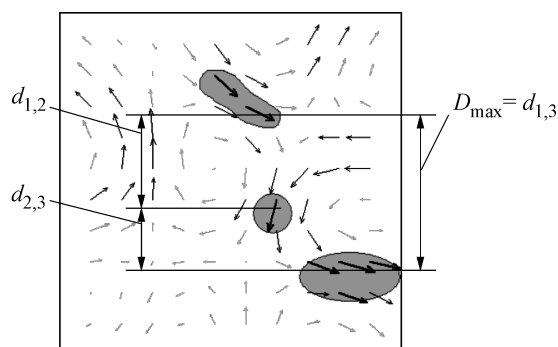


Рис. 6.47. Расстояние $D_{\max}$, вычисленное между тремя кластерами

$$C_k = \begin{cases} 0, & \text{если } \lambda \geq \lambda_0, N_{cl} = 1, \\ 0,25, & \text{если } \lambda \geq \lambda_0, N_{cl} = 2, D_{\max} = 1, \\ 0,5, & \text{если } \lambda \geq \lambda_0, N_{cl} > 2, D_{\max} \leq 2, \\ 0,75, & \text{если } \lambda \geq \lambda_0, N_{cl} > 2, D_{\max} > 2, \\ 1, & \text{если } \lambda < \lambda_0, \end{cases}$$

где λ_0 — заданный порог.

На конечном этапе обработки последовательности карт $\psi_1, \dots, \psi_K$ усредняются значения индексной переменной $\bar{C}_{ST} = \sum_{k=1}^K C_k$. Для повышения устойчивости оценки $\bar{C}_{ST}$ перед усреднением C_k удаляются наибольшее и наименьшее значения C_k , вычисленные для карт участка реполяризации (фрагмента $S-T$ сердечного цикла).

Интерфейс программного модуля, реализующего описанный выше алгоритм, содержит два окна. В первом (рабочем) окне (рис. 6.48) отображаются карты плотностей тока, построенные для выбранного интервала, причем с помощью органов управления пользователь имеет возможность последовательно просматривать эти карты и изменять параметры их изображения на экране. В этом же окне отображается результат автоматической интерпретации карт. Реализована также возможность подготовки отчета об МКГ исследовании пациента в виде твердой копии на принтере.

Для удобства настройки параметров, фигурирующих в описанном выше алгоритме, предусмотрено дополнительное окно

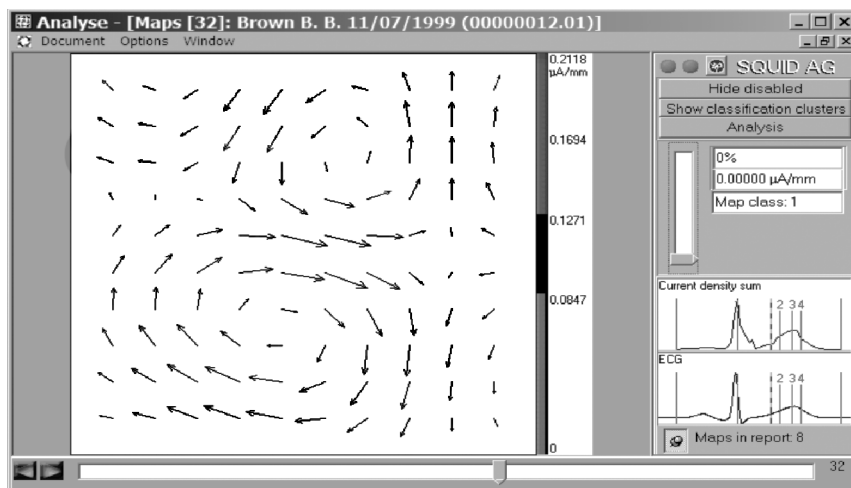


Рис. 6.48. Основное рабочее окно модуля интерпретации

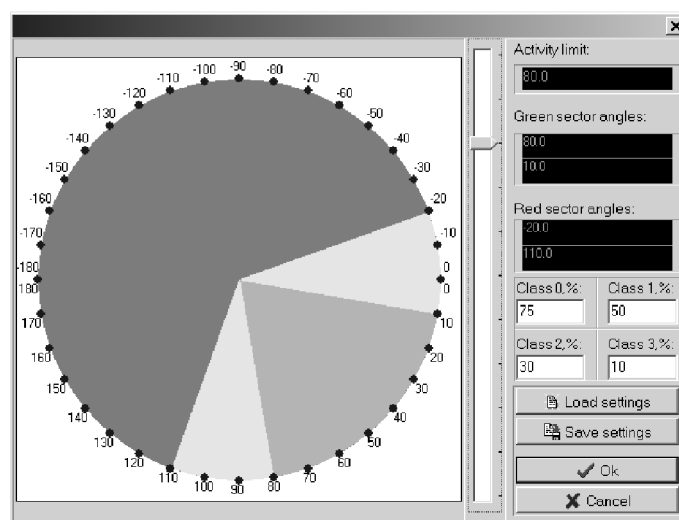


Рис. 6.49. Окно настройки параметров модуля интерпретации

настройки (рис. 6.49). Однако доступ к этому окну обеспечен по паролю, что исключает непреднамеренные действия неквалифицированных пользователей при работе с системой.

6.5.3. Результаты клинических испытаний

Основные медицинские испытания ИТ «MAGWIN» проводились в клиниках Германии [284] — кардиологической клинике университета Дуйсбург-Эссен, католическом госпитале «Филлипусстифт», Центре сердца земли Северный Рейн-Вестфалия и Германском центре сердца.

Измерения МКГ выполнялись с помощью четырехканального СКВИД-магнетометра второго порядка в неэкранированном помещении. Данные регистрировались в течение 30 секунд в каждой из 36 точек плоской квадратной сетки с шагом 4 см. Все последовательные шаги обработки и анализа данных, включая визуализацию и интерпретацию результатов, выполнялись с применением ИТ «MAGWIN».

Карты плотностей тока генерировались с промежутком 10 мс. В результате для анализа $S-T$ -интервала в каждом обследовании использовалось примерно 20 карт.

Клинический материал: 123 пациента с диагностированной ишемической болезнью сердца (ИБС), но с нормальной или неспецифически измененной ЭКГ покоя. Всем больным была проведена коронароангиография и выявлен стеноз $\geq 50\%$ как минимум в одной коронарной артерии. У всех отобранных больных на электрокардиограмме покоя в 12 отведениях отсутствовали: депрессия или элевация сегмента ST , а также негативизация зубца T более чем в двух отведениях; патологический зубец Q ; признаки полной блокады ножек пучка Гисса; фибриляция предсердий и желудочковые нарушения ритма высоких градаций. Кроме того, у всех отобранных больных по данным ЭхоКГ не были выявлены нарушения региональной и глобальной сократимости.

Контрольная группа состояла из 124 добровольцев без анамнестических признаков сердечно-сосудистой патологии, с нормальными результатами ЭКГ покоя, стресс-ЭКГ, ЭхоКГ покоя. Большая часть испытуемых приглашалась из местных отделений полиции и пожарной охраны.

Разработанный алгоритм интерпретации карт прошел испытания на этих клинических данных. Испытания показали, что с высокой достоверностью ($P > 0,999$) средние значения интегрального показателя $\overline{C}_{ST}$ были разными и составили $\overline{C}_{ST} = 0,6 \pm 0,2$ для пациентов с ИБС и $\overline{C}_{ST} = 0,28 \pm 0,18$ для здоровых

людей. Следует напомнить, что и первые, и вторые обследуемые имели нормальные или неспецифически измененные ЭКГ.

При диагностике использовано правило

$$\begin{aligned} \text{ИБС,} \quad & \text{если } \bar{C}_{ST} > C_0, \\ \text{Норма,} \quad & \text{если } \bar{C}_{ST} < C_0, \end{aligned} \quad (6.37)$$

где $C_0 \in [0,1] = \text{const}$ — пороговое значение, выбор которого определяет чувствительность S_E и специфичность S_P теста.

Анализ экспериментальной ROC-кривой показал, что тест, основанный на правиле (6.37) при $C_0 \approx 0,5$, обеспечивает чувствительность $S_E = 66\%$ и специфичность $S_P = 87\%$.

Поскольку, как уже отмечалось, согласно данным медицинской статистики, prevalence ИБС среди населения стран Европы составляет около 6%, то в силу условия (5.40) можно констатировать: разработанный тест полезен для выявления больных ИБС с неизменными ЭКГ в достаточно широком диапазоне соотношения потерь от ложноотрицательных и ложноположительных ошибок: $4,4 \leq \omega_{\text{ИБС}} \leq 39,6$.

6.6. Перспективные информационные технологии

6.6.1. Идентификация личности на основе уникальных особенностей фазового портрета ЭКГ

На современном этапе развития информационного общества все большее значение приобретают методы защиты от несанкционированного доступа к информации и различным техническим средствам. Существующие системы аутентификации пользователя, основанные на проверке пароля, индивидуального ключа либо иного идентифицирующего документа, имеют существенный недостаток: пароль можно забыть, потерять или подделать. Поэтому все больший интерес вызывают компьютерные биометрические системы аутентификации, основанные на уникальных биологических характеристиках человека.

Наибольшее распространение получили дактилоскопические системы, основанные на автоматическом распознавании отпечатков пальцев. Известны также методы защиты от несанкционированного доступа, основанные на автоматическом распознавании изображения лица и индивидуальных свойств речевого сигнала.

В работе [278] нами впервые была сформулирована гипотеза об индивидуальных особенностях фазового портрета электрокардиограммы. Проведенные к настоящему времени исследования подтвердили эту гипотезу и позволили создать экспериментальные образцы интеллектуальных информационных технологий, основанных на этом методе.

Проанализируем две постановки практически важных задач. Первую задачу (назовем ее задачей «идентификации») сформулируем следующим образом. Существует ограниченный список пользователей $V = \{V_1, \dots, V_U\}$. Требуется построить ИТ, которая на основе автоматической обработки сигнала об электрической активности сердца одного из пользователей $V_i \in V$ позволит указать его номер в списке.

Вторую задачу, которую назовем задачей «верификации» (аутентификации), сформулируем несколько иначе. Испытуемый сам указывает фамилию (login), определяющую его номер в списке $V = \{V_1, \dots, V_U\}$. Требуется построить ИТ, которая на основе индивидуальных особенностей сигнала об электрической активности сердца испытуемого («интеллектуального» пароля) подтвердит или опровергнет названный login. При этом допускается, что испытуемый может не принадлежать списку $V = \{V_1, \dots, V_U\}$.

Как видно из рис. 6.50, различие двумерных фазовых портретов ЭКГ более выразительны, чем различие одномерных ЭКГ во временной области, породивших эти портреты.

Даже при визуальной оценке легко увидеть, что фазовые портреты ЭКГ различных испытуемых (рис. 6.51), подобно отпечаткам пальцев, имеют ряд индивидуальных особенностей. При этом различия некоторых из них (например, первого и шестого) весьма существенны, в то время как иные (например, второй и четырнадцатый) демонстрируют менее выраженные отличия.

Однако эксперименты показали, что даже если двумерные фазовые портреты ЭКГ практически совпадают (рис. 6.52, слева), то при переходе в трехмерное фазовое пространство с координатами $y - \dot{y} - \ddot{y}$ удастся обнаружить некоторые различия якобы тождественных фазовых портретов при определенных углах проектирования (рис. 6.52, справа).

Исследования также показали, что фазовые портреты сохраняли свои индивидуальные особенности на протяжении достаточно большого периода наблюдений (рис. 6.53).

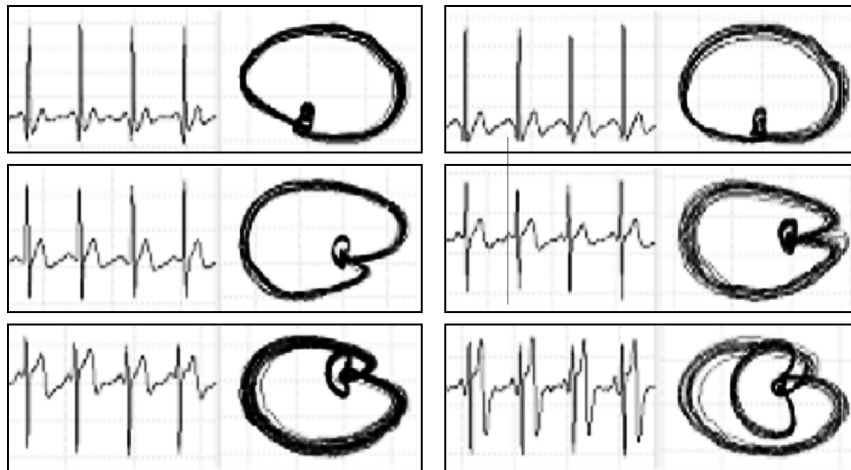


Рис. 6.50. Примеры реальных ЭКГ во временной области и фазовом пространстве

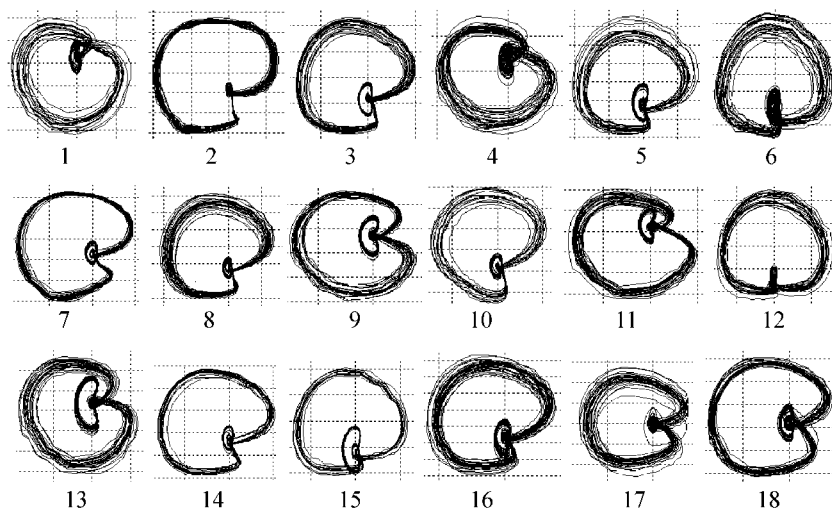


Рис. 6.51. Фазовые портреты ЭКГ 18 испытуемых

Разумеется, при физических или эмоциональных нагрузках происходят изменения отдельных фрагментов фазовых траекторий, в частности фрагмента реполяризации, о чем уже шла речь выше. Поэтому для решения задач идентификации и верифика-

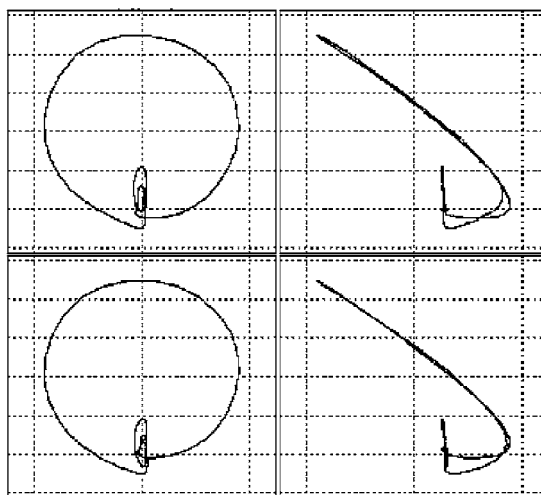


Рис. 6.52. Проекция на плоскость трехмерных фазовых портретов двух испытуемых при различных углах проектирования

ции необходимо анализировать наиболее стабильный фрагмент фазового портрета, которым, как показали эксперименты, является фрагмент, соответствующий желудочковому комплексу *QRS*.

Для построения предлагаемой ИТ идентификации личности по ЭКГ разработан комплекс вычислительных алгоритмов, направленных на решения таких задач:

- формирование эталонов фазовых портретов ЭКГ пользователей группы $V = \{V_1, \dots, V_U\}$;
- построение текущего фазового портрета испытуемого и сопоставление этого портрета со всеми имеющимися эталонами;
- выбор эталона, ближайшего к наблюдаемому фазовому портрету.

Принципиально для решения этих задач может быть использован метод усреднения и сравнения пар фазовых траекторий, основанный на вычислении хаусдорфовых расстояний (см. гл. 3). Однако эксперименты показали, что для повышения быстродействия приемлемую оценку усредненной фазовой траектории можно получить более простым путем, проведя предварительную модификацию исходных данных.

Суть модификации состоит в том, что для каждой i -й обрабатываемой траектории проводится линейная интерполяция значений между узлами $z_i[m]$ и $z_i[m-1]$, $m = 1, \dots, M_i - 1$ в нормированном трехмерном фазовом пространстве с координатами $y, \dot{y}, \ddot{y}$, а затем полученный массив векторов $\tilde{z}[k]$ прореживается таким образом, чтобы на траектории «равномерно» расположилось заданное число точек $M_0 = \text{const}$ (рис. 6.54), т. е. чтобы выполнялось условие $r_m \approx r_{m+1}$, $m = 1, \dots, M_0 - 1$, где

$$r_m = \sqrt{(y[m] - y[m+1])^2 + (\dot{y}[m] - \dot{y}[m+1])^2 + (\ddot{y}[m] - \ddot{y}[m+1])^2}.$$

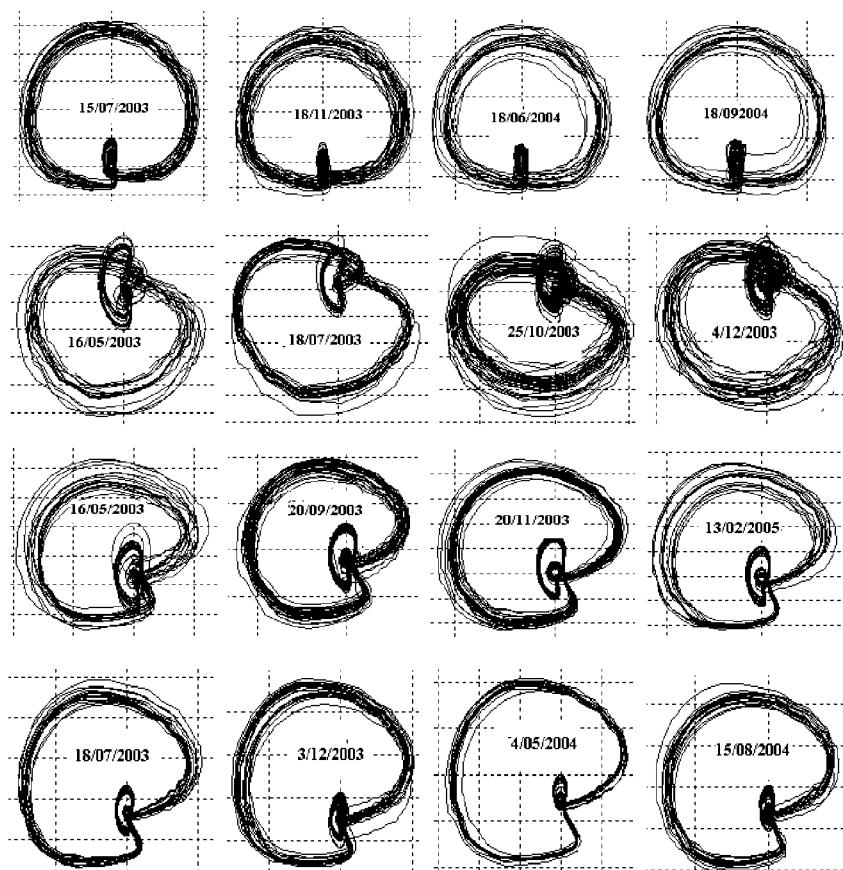


Рис. 6.53. Галерея фазовых портретов ЭКГ четырех испытуемых

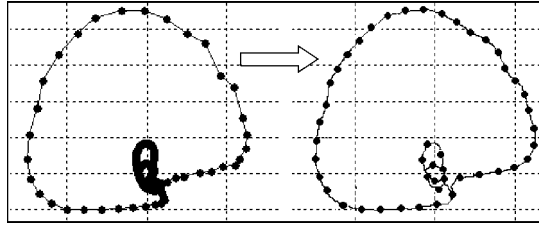


Рис. 6.54. Результат перехода от векторов $z[\cdot]$ к $\tilde{z}[\cdot]$
(показана одна из ортогональных проекций 3-мерных фазовых портретов)

Поскольку после такой модификации каждая из обрабатываемых фазовых траекторий содержит одинаковое число точек, которые к тому же синхронизированы во времени, можно получить оценку значений усредненных векторов $\hat{z}[m] = (\hat{y}[m], \hat{y}[m], \hat{y}[m])$, применив к модифицированным векторам $\tilde{z}[m]$ простую формулу

$$\hat{z}[m] = \frac{1}{N} \sum_{n=1}^N \tilde{z}[m], \quad m = 1, \dots, M_0.$$

Кроме того, модификация обрабатываемых данных позволяет достаточно просто сравнивать пары двух фазовых траекторий, используя вместо хаусдорфовых расстояний среднее или максимальное евклидово расстояние между соответствующими парами точек, принадлежащих i -й и j -й траекториям:

$$R_{\text{ср}} = \frac{1}{M_0} \sum_{m=1}^{M_0} \rho(\tilde{z}_i[m], \tilde{z}_j[m]) \quad \text{или} \quad R_{\text{max}} = \max_{1 \leq m \leq M_0} \rho(\tilde{z}_i[m], \tilde{z}_j[m]).$$

Использование такого подхода позволило существенно ускорить процедуру построения эталонов и сравнения наблюдаемых фазовых портретов с эталонами, хранимыми в базе данных.

Чтобы повысить достоверность результатов, предложено для каждого пользователя формировать несколько эталонов его фазовых портретов, используя сигналы, зарегистрированные в различных внешних условиях (до и после дозированной физической нагрузки, в разное время суток и т. п.). При этом решение о необходимости формирования нового эталона принимается самой обучающей программой на основе сравнения текущего образа с имеющимися эталонами и анализа дополнительного набора признаков.

Идентификация пользователя осуществляется на основе вычисления и сравнения по правилу минимума среднего евклидова расстояния между его фазовым портретом и одним из эталонов, хранимым в базе. Цикл идентификации проводится несколько раз по ходу регистрации ЭКГ и окончательное решение принимается методом голосования в пользу того эталона, который был выбран большее число раз.

Для выполнения экспериментальных исследований разработана программа (рис. 6.55), которая реализует предложенные алгоритмы идентификации пользователя по ЭКГ. Ввод сигнала осуществляется с пальцев пользователя с помощью микропроцессорного сенсора ИТ «ФАЗАГРАФ» (см. рис. 6.18) в течение 15–20 с. Его обработка проводится в реальном времени. Результат идентификации представляется в виде строки с фамилией идентифицированного пользователя и сопровождается голосовым сообщением. Для большей наглядности на экран выводится также фото идентифицированного пользователя. Построение базы данных с эталонами фазовых портретов пользователей происходит с помощью отдельного программного модуля обучения системы.



Рис. 6.55. Внешний вид окна программы идентификации пользователя



Рис. 6.56. Внешний вид окна программы верификации пользователя

На первых экспериментах получены обнадеживающие результаты. Создана база данных, содержащая эталоны фазовых портретов 167 различных пользователей. Для формирования эталонов использовалось от 2 до 382 ЭКГ каждого пользователя, зарегистрированных при различных внешних ситуациях. Оценки достоверности идентификации проводились по 3133 ЭКГ пользователей, эталоны которых хранились в базе данных. При экспериментах 3027 ЭКГ (96,6 %) были идентифицированы правильно.

Разработана также экспериментальная программа, обеспечивающая проверку эффективности предложенного подхода для решения второй задачи — верификации пользователя (на примере верификации владельца автомобиля).

В экспериментах участвовало 516 пользователей. Эксперимент состоял в том, что все испытуемые выполняли роль авто-владельца либо взломщика, который пытался открыть «чужой» автомобиль. Автовладельцами считались 167 пользователей, эталоны фазовых портретов которых хранились в базе данных. На экран последовательно выводились фотографии машин, «принадлежащих» разным автовладельцам (рис. 6.56).

В качестве пароля доступа использовался сигнал, снимаемый с пальцев рук испытуемого с помощью датчика ИТ «ФАЗАГРАФ». Эксперименты показали, что вероятность взлома «чужого» автомобиля зависела от количества эталонов автовладельца, хранимых в базе данных, и не превышала 1,16 %.

Разумеется, рассмотренные ИТ лишь подтверждают принципиальную возможность выдвинутой гипотезы и являются прото-

типами коммерческих продуктов, разработка которых требует проведения дополнительных исследований, направленных на совершенствование вычислительных алгоритмов и экспериментальную оценку их эффективности.

6.6.2. Анализ и интерпретация кривых биохемилюминесценции (БХЛ)

В 1961 г. Б.Н. Тарусов и его сотрудники обнаружили новое явление — биохемилюминесценцию [187], суть которого состоит в сверхслабом свечении живых организмов и биосубстратов за счет выделения тепла при химических реакциях, протекающих в тканях. Основными излучателями света являются химические молекулы, имеющие неспаренные электроны (свободные радикалы, перекиси различного происхождения и т. п.). Биохемилюминесценция наблюдается в видимой и инфракрасной частях спектра и может быть зарегистрирована как эмиссия квантов света с помощью фотоэлектронных умножителей. В результате может быть сформирован сигнал $B(t)$, несущий информацию об изменении свечения БХЛ во времени (рис. 6.57).

Явление БХЛ сразу же было воспринято исследователями как метод диагностики заболеваний человека. Преимущество этого метода заключается в том, что, используя один и тот же измерительный прибор и меняя лишь реактивы, можно обнаружить и измерить большое количество биологически активных веществ, получить информацию о функционировании и метаболизме клеток, тканей и биологических жидкостей, иметь возможность исследовать кинетику окислительных реакций, развивающихся под влиянием разнообразных физиологических и патологических агентов. По мнению специалистов, чувствительность данного метода сравнима, а в некоторых случаях даже превосходит чувствительность радиоизотопного метода.

Для того чтобы, используя метод биохемилюминесценции, можно было определить источник патологического процесса, в работах Н.И. Якубы [186 и др.] предложено его применять при постановке иммунологических реакций организма. Для этого производится сравнительный анализ кривых изменения во времени БХЛ, зарегистрированных в двух пробах — содержащей специфический антиген и контрольной пробе (без антигена). При диагностике может использоваться широкий набор специфических антигенов на определенные органы (печень, желудок,

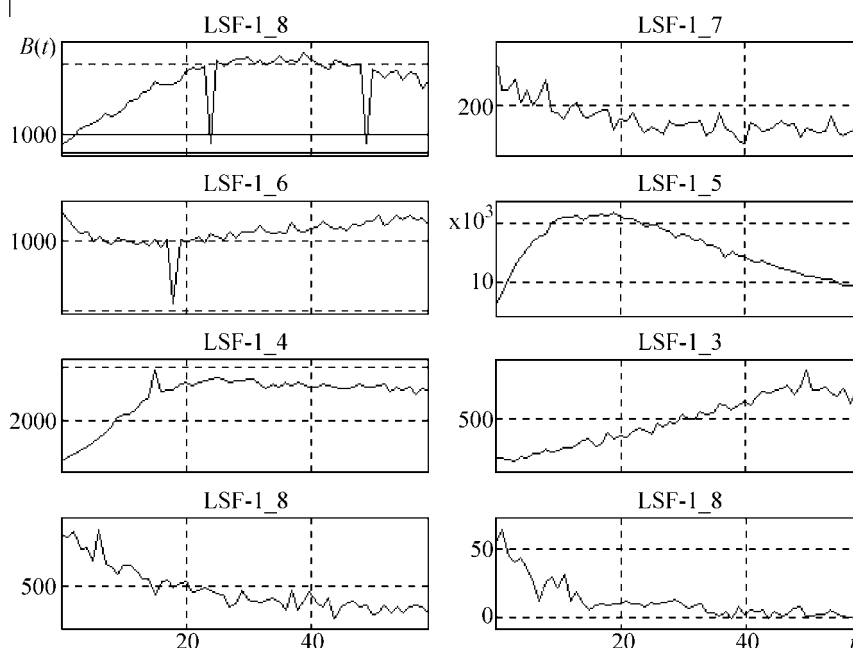


Рис. 6.57. Примеры реальных кривых БХЛ

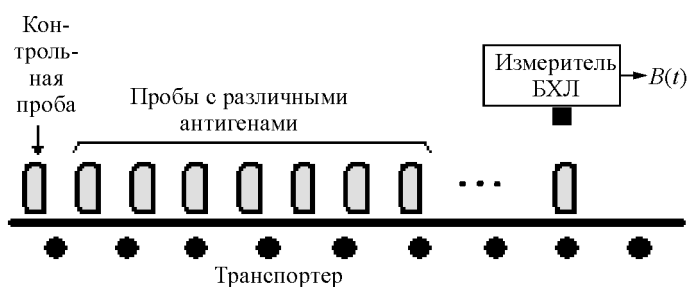


Рис. 6.58. Измерительная установка БХЛ

желчный пузырь, поджелудочная железа, селезенка, толстый и тонкий кишечник и т. д.), различные бактерии, аллергены и ряд других антигенов.

На рис. 6.58 показана упрощенная схема измерительной установки, которая формирует сигнал $B(t)$, несущий информацию об изменении БХЛ во времени. С помощью транспортера к измерителю БХЛ последовательно поступают предварительно под-

готовленные пробы с биологическим материалом, в качестве которого может использоваться кровь, моча или слюна испытуемого пациента. При этом одна из проб является контрольной, а остальные содержат специфические антигены.

Поскольку интенсивность спонтанной БХЛ очень низка, то на практике используются различные способы ее усиления, которое достигается путем воздействия электрического поля, добавления к пробе металлов с переменной валентностью, выступающих в роли катализатора, дозированного количества перекиси водорода, воздействия ультрафиолетового излучения, использования люминесцирующих соединений.

После добавления в биологическую жидкость (плазму крови, слюну, мочу и пр.) перекиси водорода свечение увеличивается примерно в 5 раз по сравнению со спонтанной хемилюминесценцией. Вспышка БХЛ длится 1–3 мин, после чего через 5–6 мин интенсивность ее понижается до стационарного уровня. На рис. 6.57 показаны реальные кривые БХЛ, полученные после индуцирования перекисью водорода проб с различными антигенами биологической жидкости (мочи) одного испытуемого.

Как видно, кривые БХЛ существенно различаются по форме, содержат характерные фрагменты, которые, по мнению специалистов, могут быть использованы как критерии при дифференциальной диагностике различных заболеваний.

В частности, при отравлениях и поражениях ионизирующей радиацией происходит существенное повышение уровня БХЛ по сравнению с нормой, а у больных со злокачественными новообразованиями, при выраженных аллергических и атеросклеротических процессах, наоборот, уровень БХЛ оказывается существенно ниже нормального уровня.

Для дальнейшего изучения диагностической ценности метода БХЛ под руководством автора разработан экспериментальный образец информационной технологии «САЛЮКС»¹⁰, которая обеспечивала ввод, предварительную обработку и визуализацию кривых БХЛ, а также хранение результатов обработки в специализированной базе данных. Используя эту технологию, Н.Я. Якуба провела серию лабораторных и клинических исследований, которые, в частности, показали, что при дискенезии желчевыводящих путей кривая БХЛ с печеночным антигеном от-

¹⁰ В разработке системы принимали участие Т.П. Потапова и В.В. Вишневский.

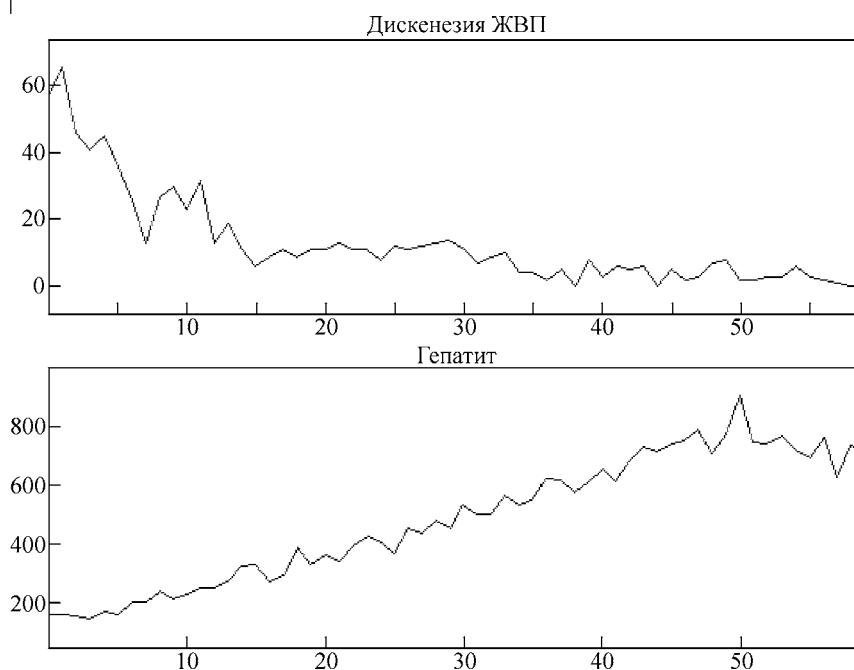


Рис. 6.59. Форма кривых БХЛ при дискенезии желчевыводящих путей (вверху) и гепатите (внизу)

личается высоким подъемом, но быстрым возвращением к данным контрольной пробы, в то время как при хроническом гепатите кривая БХЛ с печеночным антигеном характеризовалась низкими показателями уровня в начале кривой и повышенными — в конце (рис. 6.59).

Безусловно, все эти интересные факты — лишь обнадеживающие гипотезы, на основе которых могут быть созданы новые диагностические критерии. Но прежде чем метод БХЛ будет пригоден для широкого клинического использования, необходимо провести многочисленные экспериментальные исследования по сравнительной оценке и сопоставлению кривых БХЛ. В этом плане актуальна задача синтеза многофункциональной ИТ для сбора, анализа и автоматической интерпретации кривых БХЛ на основе рассмотренных в монографии методов обработки искаженных возмущениями сигналов сложной формы.

6.6.3. Интегральная оценка эластичности кровеносных сосудов

Известно, что старение организма сопровождается потерей эластичности крупных кровеносных сосудов. Усиление жесткости артерий приводит к увеличению скорости распространения пульсовой волны [261]. Этот фактор в настоящее время признается одним из главных факторов риска гипертензии (повышение артериального давления) и возникновения ишемической болезни сердца.

В основу распространенных в настоящее время методов оценки эластичности крупных сосудов положена идея измерения времени между прямой (порожденной ударом сердца) и обратной (отраженной от нижних конечностей) пульсовыми волнами. Для измерения используют различные датчики (оптические, пьезоэлектрические, емкостные), которые размещают на руке пациента. С помощью этих датчиков регистрируется электрический сигнал, несущий информацию о пульсовой волне. Задача обработки такого сигнала состоит в обнаружении на нем фрагментов, соответствующих прямой и отраженной пульсовым волнам, и измерении интервала времени между указанными фрагментами.

Проведенные нами предварительные эксперименты показали, что приемлемую оценку эластичности кровеносных сосудов можно совместить с процессом измерения артериального давления (АД) на основе так называемого осциллометрического метода. Остановимся на этом более подробно.

Хорошо известен неинвазивный метод измерения артериального давления, основанный на обнаружении восстановления кровотока через артерию после ее пережатия специальной манжетой с последующим снижением давления. Для определения момента восстановления кровотока используются два метода: аускультативный и осциллометрический.

Первый основан на определении так называемых тонов Короткова с помощью микрофона, расположенного над плечевой артерией. Однако такой метод чувствителен к внешним шумам и точности расположения микрофона над артерией. Поэтому в современных автоматических тонометрах получил распространение второй метод — осциллометрический.

Осциллометрический метод основан на том, что при прохождении крови во время систолы через сдавленный участок артерии в манжете возникают микропульсации, анализируя которые

можно получить значения систолического, диастолического и среднего АД. Анализ осцилляции проводится с помощью специальных алгоритмов. Систолическому АД обычно соответствует давление в манжете, при котором происходит наиболее резкое увеличение амплитуды осцилляции, среднему — максимальный уровень осцилляции, и диастолическому — резкое ослабление осцилляции.

Метод имеет следующие преимущества (по сравнению с аускультативным):

- из-за устойчивости к шумовым нагрузкам может использоваться при высоком уровне шума;
- показатели АД почти не зависят от разворота манжеты на руке и мало зависят от ее перемещений вдоль руки (если манжета не достигла локтевого сгиба);
- точность измерения не снижается при определении АД через тонкую ткань одежды, что важно из соображений гигиены.

Проведенные исследования показали, что применяя разработанные нами алгоритмы к обработке сигнала, несущего информацию об изменении давления в манжете, можно получить не только информацию, необходимую для измерения артериального давления, но и «увидеть» в этом сигнале фрагменты, соответствующие отраженным пульсовым волнам. Это дает ключ к построению нового подхода для определения эластичности кровеносных сосудов.

На рис. 6.60 показана схема, иллюстрирующая предлагаемый подход. На плечевую или кистевую артерию пациента накладывают манжету и за счет нагнетания воздуха постепенно повышают в ней давление до значения, превышающего на 20—30 мм рт. ст. ожидаемое значение систолического артериального давления P_c . Затем постепенно снижают давление до значения, которое на 20—30 мм рт. ст. меньше ожидаемого значения диастолического давления P_d . Этот процесс может выполняться с помощью ручного (груша) или автоматического компрессора.

Участок артерии, на которую наложена манжета, за счет дополнительного давления постепенно сжимается. Датчик формирует электрический сигнал, несущий информацию о текущем значении давления $P(t)$ в манжете. В процессе медленной (2—4 мм рт. ст./с) компрессии или декомпрессии давления в манжете проводится автоматический анализ амплитуд микропульсаций сигнала $P(t)$ (осцилляций), наблюдаемых за счет восстановления кровотока через артерию. На рис. 6.60, а показан типичный гра-

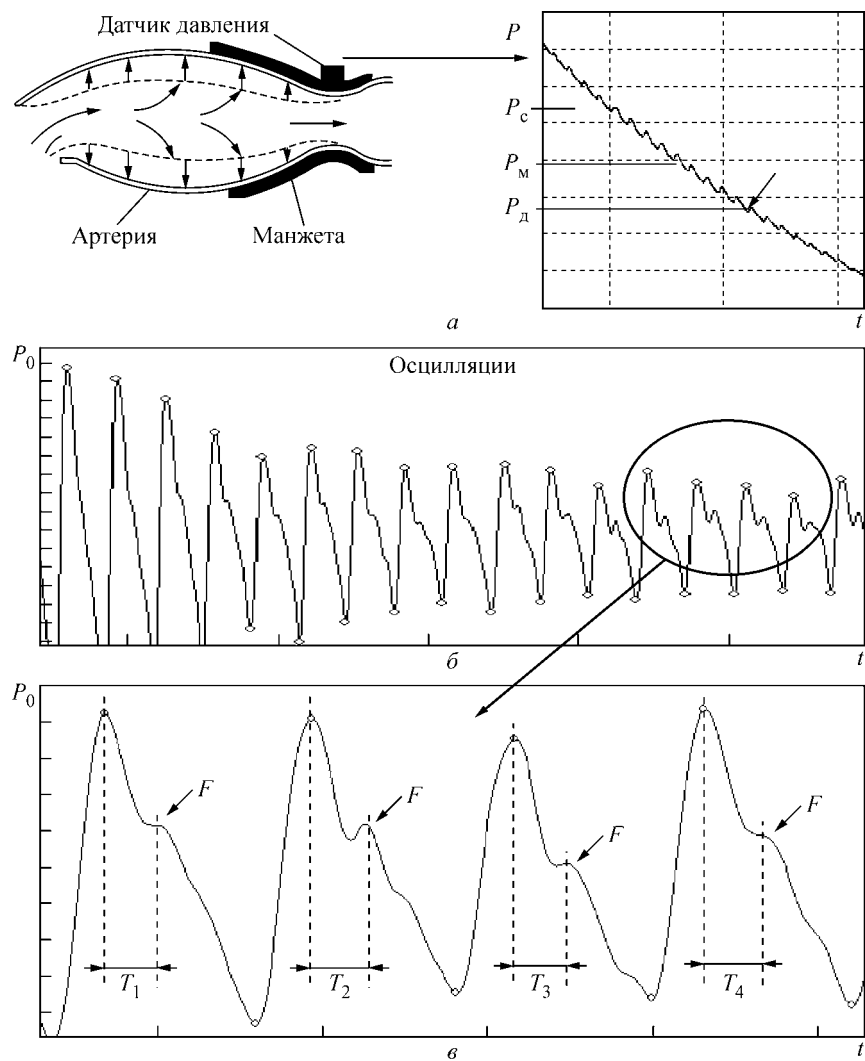


Рис. 6.60. Иллюстрация предлагаемого метода оценки эластичности кровеносных сосудов:

a — график изменения давления в манжете; b — график осцилляций; c — последовательность осцилляций в увеличенном масштабе

фик сигнала $P(t)$, зарегистрированный в процессе медленной декомпрессии.

Уровень сигнала $P(t)$, при котором в процессе медленной декомпрессии амплитуда осцилляций превышает заданное значение либо резко увеличивается, определяет значение P_c систолического артериального давления, а уровень сигнала $P(t)$, при котором в процессе медленной декомпрессии амплитуда осцилляций становится меньше заданной величины либо резко уменьшается, определяет значение P_d диастолического артериального давления.

Значения систолического P_c и диастолического P_d артериального давления могут быть также вычислены с помощью заранее построенных регрессионных зависимостей, связывающих P_c и P_d со значением среднего давления P_m , которому соответствует максимальная амплитуда осцилляций.

Эксперименты показали, что при давлениях, близких к систолическому или диастолическому, на осцилляциях (рис. 6.60, б) наблюдаются фрагменты, вызванные отраженной пульсовой волной от нижних частей тела пациента. Такие фрагменты имеют вид зубцов или площадок F (рис. 6.60, в) и могут быть автоматически обнаружены с помощью ряда оригинальных алгоритмов, реализованных в инструментальной системе «СИДИГРАФ».

Вычислим теперь временные задержки T_i , $i = 1, 2, \dots$, между моментом, когда соответствующая осцилляция принимает максимальное значение, и моментом появления указанного фрагмента F , обусловленного отраженной волной. Затем определим среднее значение временных задержек по некоторой последовательности N_F таких осцилляций:

$$T_m = \frac{1}{N} \sum_{i=1}^N T_i.$$

Полученное значение дает возможность вычислить среднюю скорость распространения кровотока и тем самым интегрально оценить эластичность кровяных сосудов пациента (так называемый индекс I_Ω жесткости сосудов) по формуле

$$I_\Omega = \frac{H}{T_m},$$

где H — рост пациента.

Для экспериментальной проверки предложенного метода разработан макетный образец информационной технологии, которая

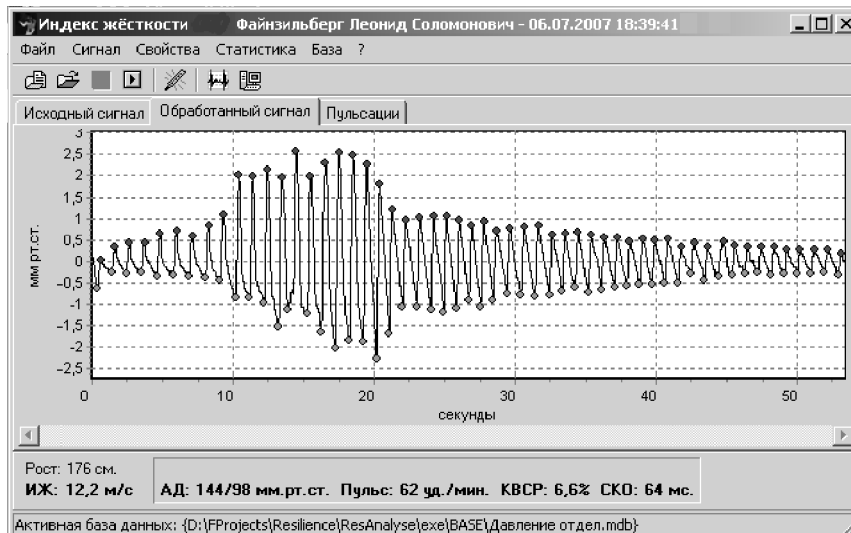


Рис. 6.61. График зарегистрированных осцилляций

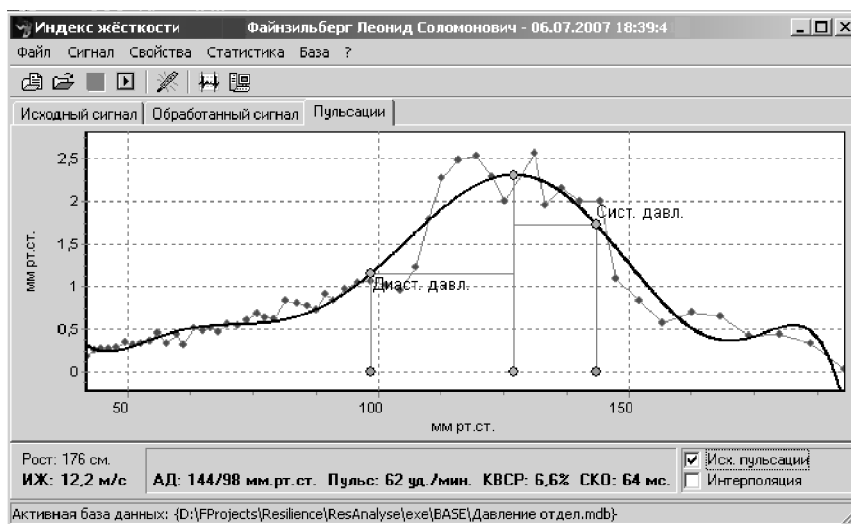


Рис. 6.62. График аппроксимирующей функции, используемой для расчета артериального давления

автоматизирует все стадии анализа и интерпретации сигнала, поступающего от датчика давления в процессе компрессии и декомпрессии манжеты. Использовался сигнал бытового тонометра ВАТ-31, реализующего осциллометрический способ измерения АД.

Оцифрованный сигнал поступал через порт RS-232 в персональный компьютер. В окнах программы последовательно отображался фильтрованный сигнал $P(t)$ и график зарегистрированных осцилляций (рис. 6.61).

Для более точного вычисления значений P_c , P_d и P_m используются алгоритмы сплайн-интерполяции и полиномиальной аппроксимации (рис. 6.62).

Индекс жесткости сосудов рассчитывается по описанной выше методике, для чего в программе реализованы вычислительные процедуры поиска и распознавания фрагментов осцилляций, порожденных отраженными волнами. Графики результатов распознавания отображаются в специальном окне программы, используя которое пользователь, в случае необходимости, может внести ручные коррективы (рис. 6.63).

Результат интерпретации сигнала отображается на специальном индикаторе, который индицирует текущие значения систолического и диастолического АД, частоту пульса (ЧСС), вариабельность ЧСС, а также индекс жесткости в графическом (в виде столбика ртути) и текстовом видах (рис. 6.64). На индикаторе отображаются границы норм указанных параметров (зеленая зона), а также незначительных (желтая зона) и существенных (красная зона) отклонений каждого из параметров от нормы. Для удобства на этих шкалах отображается стрелка, положение которой соответствует среднему значению показателя конкретного пациента, вычисленному по данным, которые хранятся в базе данных.

Как уже отмечалось, у пациентов пожилого возраста за счет склеротических изменений сосудов индекс I_Ω жесткости принимает более высокие значения. Поэтому на основе статистических исследований для каждой возрастной группы здоровых добровольцев должны быть построены шкалы пороговых значений, превышение которых будет служить маркером преждевременного старения организма.

Предварительные испытания ИТ проводились в группе из 27 добровольцев разного возраста. Результаты испытаний достаточно хорошо согласовывались с устоявшимся представлением клиницистов об увеличении значения индекса жесткости I_Ω по мере

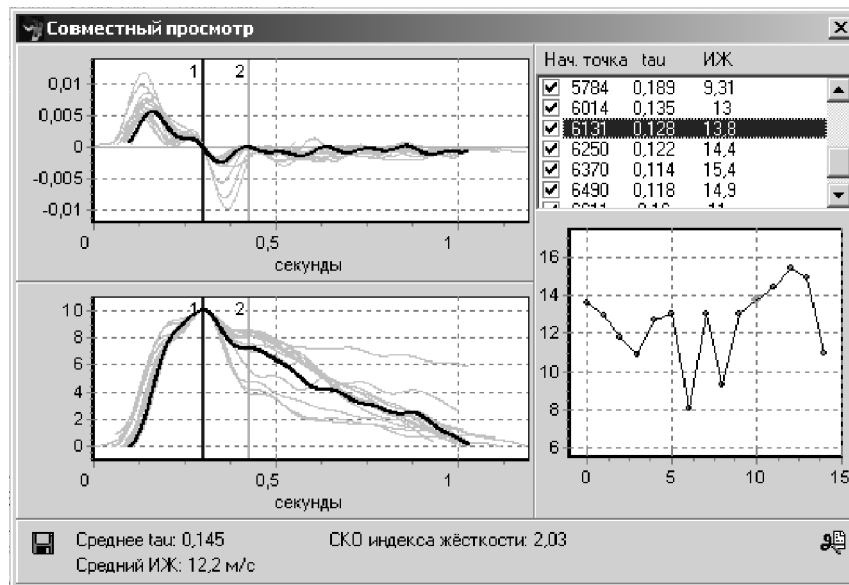


Рис. 6.63. Графики, используемые для расчета индекса жесткости



Рис. 6.64. Индикатор результатов обработки

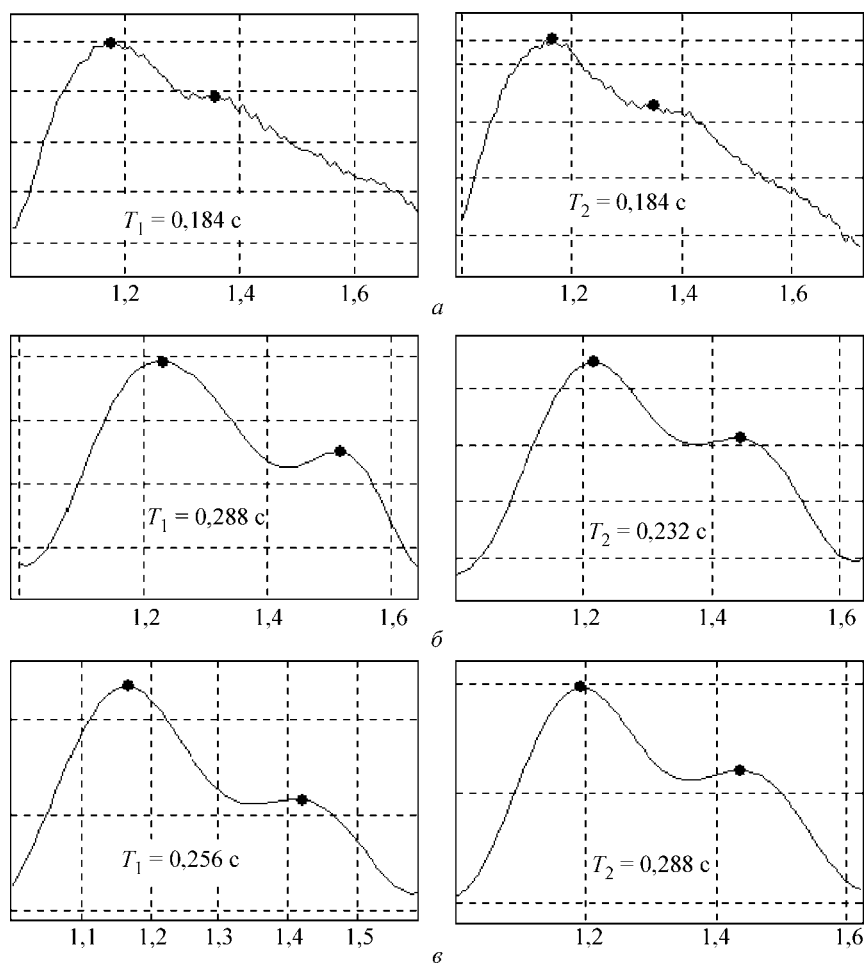


Рис. 6.65. Эксперименты по вычислению I_Ω у добровольцев разного возраста:

а — 57 лет; *б* — 35 лет; *в* — 14 лет

старения организма. На рис. 6.65 приведены результаты экспериментов по вычислению I_Ω трех здоровых добровольцев.

Возраст первого испытуемого 57 лет, рост $H = 176$ см. Временные задержки между максимальными значениями осцилля-

ций и моментами появления фрагментов, порождаемых отраженной пульсовой волной, на первой и второй осцилляциях были равными: $T_1 = T_2 = 0,184$ с (рис. 6.65, *а*). Следовательно, $T_m = 0,184$ с и $I_\Omega = 1,76/0,184 = 9,56$ м/с.

Возраст второго испытуемого 35 лет, рост $H = 178$ см. Временные задержки приняли значения $T_1 = 0,288$ с, а $T_2 = 0,232$ с (рис. 6.65, *б*). Следовательно, $T_m = 0,26$ с и $I_\Omega = 1,78/0,26 = 6,85$ м/с.

Возраст третьего испытуемого 14 лет, рост $H = 165$ см. В данном случае временные задержки приняли значения $T_1 = 0,256$ с, $T_2 = 0,288$ с (рис. 6.65, *в*). Следовательно, $T_m = 0,272$ с и $I_\Omega = 1,78/0,26 = 1,65/0,272 = 6,07$ м/с.

В заключение отметим, что рассмотренные в монографии методы и инструментальные средства построения ИТ обработки сигналов сложной формы с локально сосредоточенными диагностическими признаками успешно используются в настоящее время и, несомненно, будут в дальнейшем широко применяться при решении актуальных прикладных задач.

СПИСОК ЛИТЕРАТУРЫ

1. Автоматическое определение в цифровой форме химического состава жидкого чугуна / Л.С. Файнзильберг, Ю.В. Власенко, Э.Х. Тухин, Э.А. Шелковый // Литейное производство. — 1985. — № 8. — С. 23,24.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. — М.: Финансы и статистика, 1983. — 471 с.
3. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Метод потенциальных функций в задаче восстановления характеристик функционального преобразователя по случайно наблюдаемым точкам // Автоматика и телемеханика. — 1964. — 25, № 12. — С. 1705—1715.
4. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Метод потенциальных функций в теории обучения машин. — М.: Наука, 1970. — 384 с.
5. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. — М.: Наука, 1976. — 280 с.
6. А.с. 1520551 СССР, МКИ G 01 F 15/46. Анализатор аналоговых сигналов / Л.С. Файнзильберг. — № 4446409/24—24; Заявл. 29.04.88; Оpubл. 07.11.81, Бюл. № 41. — 7 с.
7. А.с. 1742691 СССР, МКИ G 01 N 25/06. Анализатор содержания кремния в жидком чугуне / Л.С. Файнзильберг. — № 4854658/25; Заявл. 30.07.92; Оpubл. 23.06.92, Бюл. № 24. — 8 с.
8. Анищенко В.С. Сложные колебания в простых системах: Механизмы возникновения, структура и свойства динамического хаоса в радиофизических системах. — М.: Наука, 1990. — 397 с.
9. Антонью А. Цифровые фильтры: Анализ и проектирование. — М.: Радио и связь, 1983. — 320 с.
10. Амосов Н.М., Агапов Б.Т., Паничкин Ю.В. Исследование сократительной функции миокарда методом фазовых координат // Докл. АН СССР. — 1972. — 202, № 1. — С. 245—247.

11. *Архиреева И.Н., Бородюк В.П., Полванов Ф.Ю.* Формирование оптимальной регрессионной структуры по данным пассивного эксперимента // Заводская лаборатория. — 1987. — № 10. — С. 71—74.
12. *Бадд Т.* Объектно-ориентированное программирование в действии: Пер. с англ. — СПб.: Питер, 1997. — 464 с.
13. *Баевский Р.М., Беяева Ф.Н., Сельцер В.К.* Методика массовой баллостокардиографии // Советская медицина. — 1960. — № 4. — С. 105—109.
14. *Баевский Р.М., Егоров А.Д., Казарьян Л.А.* Методика сейсмокардиографии. // Кардиология. — 1964. — № 2. — С. 87—89.
15. *Баевский Р.М., Кириллов О.И., Клецкин С.З.* Математический анализ изменений ритма сердца при стрессе. — М.: Наука, 1984. — 480 с.
16. *Бакан Г.М., Куссуль Н.Н.* Размытый эллипсоидальный алгоритм фильтрации состояний статического объекта // Проблемы управления и информатики. — 1996. — № 5. — С. 77—93.
17. *Балакришнан А.* Теория фильтрации Калмана. — М.: Мир, 1988. — 168 с.
18. *Барабаш Ю.Л.* Коллективные статистические решения при распознавании. — М.: Радио и связь, 1983. — 224 с.
19. *Барабаш Ю.Л.* Минимизация описаний в задаче автоматического распознавания образов // Изв. АН СССР. Техн. кибернетики. — 1964. — № 3. — С. 14—23.
20. *Барабаш Ю.Л.* Учет свойств признаков при распознавании // Там же. — 1965. — № 5. — С. 85—92.
21. *Барков А.В., Баркова Н.А., Азовцев А.Ю.* Мониторинг и диагностика роторных машин по вибрации. — СПб.: Государственный морской технический университет. — 2000. — 169 с.
22. *Басвилль М., Вилски А., Банвенист А.* Обнаружение изменения свойств сигналов и динамических систем. — М.: Мир, 1989. — 278 с.
23. *Белкин А.Р., Левин М.Ш.* Принятие решений: комбинаторные методы аппроксимации информации. — М.: Наука, 1990. — 160 с.
24. *Бен-Ари М.* Языки программирования. Практический сравнительный анализ: Пер. с англ. — М.: Мир, 2000. — 366 с.
25. *Берг Л.Г.* Введение в теорию термографии. — М.: Наука, 1969. — 380 с.
26. *Бидюк П.И., Баклан И.В., Гасанов А.С.* Системный подход к анализу адекватности нелинейных моделей временных рядов // Кибернетика и системный анализ. — 2003. — № 3. — С. 147—158.
27. *Биргер И.А.* Определение диагностической ценности признаков // Кибернетика. — 1968. — № 3. — С. 80—85.
28. *Биргер И.А.* Техническая диагностика. — М.: Машиностроение, 1978. — 240 с.
29. *Богге У., Боггс М.* UML и Rational Rose: Пер. с англ. — М.: ЛОРИ, 2000. — 582 с.
30. *Бодянский Е.В., Руднева И.А.* Об одном адаптивном алгоритме обнаружения разладок в случайных последовательностях // Автоматика и телемеханика. — 1995. — № 10. — С. 101—106.
31. *Боровиков В.П.* Популярное введение в программу STATISTICA. — М.: КомпьютерПресс, 1998. — 267 с.
32. *Бородюк В.П., Лецкий Э.К.* Статистическое описание промышленных объектов. — М.: Энергия, 1971. — 110 с.

33. *Браверман Э.М., Мучник И.Б.* Структурные методы обработки эмпирических данных. — М.: Наука, 1983. — 492 с.
34. *Бродский В.З.* Введение в факторное планирование эксперимента. — М.: Наука, 1976. — 223 с.
35. *Брюхович Е.И.* Плоды эволюции философии науки. Действительная роль экономики и супер-ЭВМ в жизни человечества на современной стадии развития цивилизации // Наука и науковедение. — 2001. — № 4. — С. 53—68.
36. *Буч Г.* Объектно-ориентированный анализ и проектирование с примерами приложений на С++: Пер. с англ. — М.: Бином; СПб.: Невский диалект, 1999. — 560 с.
37. *Валужис А.К., Рашимас А.П.* Статистический алгоритм структурного анализа электрокардиосигнала // Кибернетика. — 1979. — № 3. — С. 91—95.
38. *Васецкий Ю.М., Файнзильберг Л.С., Чайковский И.А.* Методы анализа структуры распределения тока в проводящей среде применительно к магнитокардиографии // Электронное моделирование, 2004. — № 3. — С. 95—115.
39. *Вапник В.Н., Червоненкис А.Я.* Теория распознавания образов. — М.: Наука, 1974. — 416 с.
40. *Васильев В.И.* Распознающие системы. — Киев: Наук. думка, 1983. — 422 с.
41. *Васильев В.И.* Индукция, дедукция и редукция в проблемах обнаружения закономерностей // Искусственный интеллект. — 1998. — № 1. — С. 14—26.
42. *Васильев В.И., Ланге Т.И.* Принцип внешнего дополнения в методе группового учета аргументов и в методе предельных упрощений // Управляющие системы и машины. — 2003. — № 3. — С. 21—30.
43. *Васильев В.И., Овсянникова Ф.П.* Оптимизация пространства в процессе обучения распознаванию образов // Автоматика. — 1985. — № 3. — С. 8—15.
44. *Ведров А.М.* CASE-технологии. Современные методы и средства проектирования информационных систем. — М.: Финансы и статистика, 1988. — 176 с.
45. *Ветвицкий Е.В., Плотников А.В.* Компьютерный сфигмограф для исследования ритма сердца. // Тез. докл. Всерос. межвуз. науч.-техн. конф. студентов и аспирантов «Микроэлектроника и информатика — 99». — М.: МИЭТ. — 1999. — С. 116.
46. *Винцюк Т.К.* Алгоритм определения эталонных элементов слова по совокупности его реализаций // Труды акустического института. — М., 1979. — Вып. 12. — С. 163—168.
47. *Винцюк Т.К.* Анализ, распознавание и интерпретация речевых сигналов. — Киев: Наук. думка, 1987. — 262 с.
48. *Винцюк Т.К.* Обучение поэлементному распознаванию речи // Распознавание образов и конструирование читающих автоматов. — Киев: Ин-т кибернетики АН УССР, 1969. — Вып. 2. — С. 23—35.
49. *Винцюк Т.К.* Распознавание слов устной речи методами динамического программирования // Кибернетика. — 1968. — № 1. — С. 81—88.
50. *Витязев В.В.* Цифровая частотная селекция сигналов. — М.: Радио и связь, 1993. — 420 с.
51. *Вишневский В.В., Рагульская М.В., Файнзильберг Л.С.* Влияние солнечной активности на морфологические параметры ЭКГ сердца здорового

- человека // Биомедицинские технологии и радиоэлектроника. — 2003. — № 3. — С. 3—12.
52. Власов В.В. Введение в доказательную медицину. — М.: Медиа Сфера, 2001. — 392 с.
53. Власов В.В. Эффективность диагностических исследований. — М.: Медицина, 1988. — 256 с.
54. Волкова Э.Г., Калаев О.Ф., Ковынев А.Р. Диагностические возможности первой производной ЭКГ в оценке состояния коронарной артерии у больных ишемической болезнью сердца // Терапевтический архив. — 1990. — № 3. — С. 35—38.
55. Воишин А.П., Дывак М.П. Планирование оптимального насыщенного эксперимента в задачах построения интервальных моделей // Заводская лаборатория. — 1993. — № 1. — С. 56—59.
56. Воишин А.П., Сотиров Г.Р. Оптимизация в условиях неопределенности. — М.: МЭИ; София: Техника, 1989. — 224 с.
57. Вопросы статистической теории распознавания / Ю.Л. Барабаш, Б.В. Варский, В.Т. Зиновьев, В.С. Кириченко, В.Ф. Сапегин. — М.: Сов. радио, 1967. — 400 с.
58. Воробьев С.А. Алгоритмы выделения и классификации фрагментов повторяющейся формы на экспериментальных данных // Автоматика и телемеханика. — 1985. — № 8. — С. 89—93.
59. Воробьев С.А. Методы структурного анализа экспериментальных кривых с участками повторяющейся формы // Автоматизация и современные технологии. — 1997. — № 7. — С. 22—25.
60. Воробьев С.А. Способы задания признака формы в задаче структурного анализа результатов медико-биологического эксперимента // Физика волновых процессов и радиотехнические системы. — 1999. — 2, № 2. — С. 60—62.
61. Воробьев С.А. Структурный анализ результатов медикобиологического эксперимента при неизвестных параметрах модели // Вестник новых медицинских технологий. — 1999. — 5, № 1. — С. 113—115.
62. Высоцкий В.Н., Ивахненко А.Г., Чеберкус В.И. Долгосрочное прогнозирование колебательных процессов с помощью выделения гармонического тренда оптимальной сложности по критерию баланса переменных // Автоматика. — 1975. — № 1. — С. 25—31.
63. Гайдышев И. Анализ и обработка данных: Спец. справочник. — СПб.: Питер, 2001. — 752 с.
64. Генкин А.А. Новая информационная технология анализа медицинских данных. — СПб.: Политехника, 1999. — 192 с.
65. Глушков В.М. Основы безбумажной информатики. — М.: Наука, 1982. — 552 с.
66. Голд Б., Рэйдер Ч. Цифровая обработка сигналов. — М.: Сов. радио, 1973. — 368 с.
67. Головкин Б.А. Машинное распознавание и линейное программирование. — М.: Сов. радио, 1973. — 100 с.
68. Гольдбергер Э.Л., Ригни Д.Р., Уэст Б.Д. Хаос и фракталы в физиологии человека // В мире науки. — 1990. — № 4. — С. 25—32.
69. Горелик А.Л. О выборе пространства признаков системы распознавания объектов и явлений // Кибернетика. — 1976. — № 3. — С. 35—39.

70. Горелик А.Л., Скрипкин В.А. Методы распознавания. — М.: Высшая школа, 1977. — 222 с.
71. Гриценко В.И., Паньшин Б.Н. Информационная технология: вопросы развития и применения. — Киев: Наук. думка, 1988. — 380 с.
72. Гришин В.Г. Образный анализ экспериментальных данных. — М.: Наука, 1982. — 274 с.
73. Гроп Д. Методы идентификации систем. — М.: Мир, 1979. — 302 с.
74. Гуляев А.П. Металловедение. — М.: Металлургия, 1966. — 478 с.
75. Гульчак М.Г. Об информативности совокупности признаков // Автоматика и телемеханика. — 1970. — № 6. — С. 161—163.
76. Демиденко Е.З. Линейная и нелинейная регрессии. — М.: Финансы и статистика, 1981. — 302 с.
77. Дивак М.П., Волощук С.В. Локалізація гарантованих оцінок параметрів технологічних процесів // Вимірювальна та обчислювальна техніка в технологічних процесах: Зб. наук. праць. — 2001. — Вип. 8. — С. 310—316.
78. Добровидов А.В. К задаче параметрического оценивания многомерного случайного параметра // Обнаружение и распознавание. Планирование экспериментов. — М.: Наука, 1970. — С. 55—61.
79. Дошницин В.Л. Клинический анализ электрокардиограммы. — М.: Медицина, 1982. — 206 с.
80. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2 кн. — М.: Финансы и статистика, 1986. — Кн. 1. — 355 с; Кн. 2. — 349 с.
81. Дьяконов В. Matlab 6: Учебный курс. — СПб.: Питер, 2001. — 592 с.
82. Дьяконов В.П. Компьютерная математика. Теория и практика. — М.: Нолидж, 2001. — 1296 с.
83. Жежель Ю.Н. Об эффективности линейной дискриминантной функции в случае произвольных распределений // Изв. АН СССР. Техническая кибернетика. — 1968. — № 6. — С. 124—128.
84. Житецкий Л.С., Файнзильберг Л.С. Об информационном подходе к оценке полезности признаков при статистическом распознавании образов // Там же. — 1983. — № 4. — С. 120—126.
85. Завалишин Н.В., Мучник И.Б. Лингвистический (структурный) подход к проблеме распознавания образов // Автоматика и телемеханика. — 1969. — № 8. — С. 33—48.
86. Загоруйко Н.В. Методы распознавания и их применение. — М.: Советское радио. — 1972. — 208 с.
87. Задирака В.К., Мельникова С.С. Цифровая обработка сигналов. — Киев: Наук. думка. — 1993. — 294 с.
88. Зайченко Ю.П., Заєць І.О. Синтез та адаптація нечітких прогнозуючих моделей на основі методу самоорганізації // Наук. Вісті НТУУ «КПІ». — 2001. — № 3. — С. 34—41.
89. Зенкин А.А. Интерактивная компьютерная графика в теоретических исследованиях // Изв. АН СССР. Техническая кибернетика. — 1987. — № 5. — С. 29—36.
90. Зенкин А.А. Когнитивная компьютерная графика. — М.: Наука, 1991. — 187 с.
91. Зюко А.Г., Коробов Ю.Ф. Теория передачи сигналов. — М.: Связь, 1972. — 282 с.

92. *Ершов А.А.* Стабильные методы оценки параметров // Автоматика и телемеханика. — 1978. — № 8. — С. 66—99.
93. *Иванцов Г.П.* Теплообмен между слитком и изложницей. — М.: Металлургиздат, 1951. — 39 с.
94. *Ивахненко А.Г., Копна Ю.В.* Регуляризация решающих функций в методе группового учета аргументов // Автоматика. — 1970. — № 2. — С. 23—36.
95. *Ивахненко А.Г., Юрачковский Ю.П.* Моделирование сложных систем по экспериментальным данным. — М.: Радио и связь, 1987. — 120 с.
96. *Ивахненко А.Г.* Индуктивный метод самоорганизации моделей сложных систем. — Киев: Наук. думка, 1982. — 360 с.
97. *Ивахненко А.Г.* Метод группового учета аргументов — конкурент метода стохастической аппроксимации // Автоматика. — 1968. — № 3. — С. 57—73.
98. *Ивахненко А.Г., Мюллер И.А.* Самоорганизация прогнозирующих моделей. — Киев: Техника, 1984. — 350 с.
99. *Ивахненко А.Г., Степашко В.С.* Помехоустойчивое моделирование. — Киев: Наук. думка, 1985. — 300 с.
100. *Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д.* Принятие решений на основе самоорганизации. — М.: Радио и связь, 1976. — 275 с.
101. *Йордон Э., Аргила К.* Структурные модели в объектно-ориентированном анализе и проектировании: Пер. с англ. — М.: ЛОРИ, 1999. — 264 с.
102. *Казначеев В.П., Баевский Р.М., Берсенева А.П.* Донозологическая диагностика в практике массовых обследований населения. — Л.: Медицина, 1989. — 208 с.
103. *Калиниченко А.Н., Гончаренко Ю.И., Родина Н.И.* Анализ спектральных параметров variability сердечного ритма в реальном масштабе времени // Биотехнические системы в медицине и биологии. — СПб.: Политехника, 2002. — С. 68—71.
104. *Карамов К.С., Базиян Ж.А., Алехин К.П.* К диагностике свежих очаговых поражений миокарда // Кардиология. — 1978. — № 10. — С. 109—112.
105. *Киселев Н.В., Сечкин В.А.* Техническая диагностика методами нелинейного преобразования. — Л.: Энергия, 1980. — 112 с.
106. *Ковалевский В.А.* Методы оптимальных решений в распознавании изображений. — М.: Наука, 1976. — 328 с.
107. *Коваль В.Н., Кук Ю.В.* Структурный метод моделирования сложных систем // Управляющие системы и машины. — 2003. — № 3. — С. 45—55.
108. *Козинец Б.Н.* Рекуррентный алгоритм разделения выпуклых оболочек двух множеств // Алгоритмы обучения распознавания образов. — М.: Сов. радио, 1973. — С. 79—92.
109. *Кокс Н.* Анализ энцефалограмм, кривых кровяного давления и электрокардиограмм на цифровой вычислительной машине // Тр. Ин-та инженеров по электронике и радиотехнике (США): Пер. с англ. — М.: ВИНТИ, 1972. — № 10. — С. 24—38.
110. *Коновалов Д.А.* Разработка математических моделей температур ликвидуса и солидуса для двухкомпонентных алюминиевых сплавов // Сварочное производство. — 1999. — № 4. — С. 29—36.

111. Коркушко О.В., Писарук А.В., Шатило В.Б. Анализ variability ритма сердца в клинической практике (Возрастные аспекты). — Киев: Ин-т геронтологии АМН Украины. — 2002. — 191 с.
112. Крайнов В.П. Соотношения неопределенности для энергии и времени // Соросовский образовательный журнал. — 1998. — № 5. — С. 77–82.
113. Кривицкий Н.М. Количественная оценка пульсового кровотока конечностей реографическими методами // Медицинская техника. — 1986. — № 1. — С. 11–15.
114. Кунцевич В.М., Кунцевич А.В. Активная идентификация и управление при ограниченных шумах // Пр. міжн. конф. з управління «АВТОМАТИКА-2000». — Львів: Держ. НДІ інформ. інфраструкт., 2000. — Т. 1. — С. 7–13.
115. Кунченко Ю.П., Лега Ю.Г. Оценка параметров случайных величин методом максимизации полинома. — Киев: Наук. думка, 1991. — 180 с.
116. Лазарев Ю.Ф. MatLAB 5.x. — Киев: Изд. группа BNV, 2000. — 384 с.
117. Лана В.Г. Математические основы кибернетики. К.: Вища школа, 1974. — 452 с.
118. Ластед Л. Введение в проблему принятия решений в медицине: Пер. с англ. — М.: Мир, 1971. — 282 с.
119. Лернер А.Я. Начала кибернетики. — М.: Наука, 1967. — 400 с.
120. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений. — М.: Физ. мат. изд., 1962. — 187 с.
121. Лисенков А.Н. О методах идентификации и управления в условиях неопределенности // Межд. конф. по интерв. и стох. методам в науке и технике «ИНТЕРВАЛ-92»: Сб. тр. — М., 1992. — Т. 1. — С. 92–96.
122. Литваков Б.М. О сходимости рекуррентных алгоритмов обучения распознаванию образов // Автоматика и телемеханика. — 1968. — № 1. — С. 92–106.
123. Литваков Б.М. Об итерационном методе аппроксимации функций по конечному числу наблюдений // Автоматика и телемеханика. — 1966. — № 4. — С. 104–113.
124. Логвинов В.С. Метод диагностики по параметрам колебательных и волновых процессов в сердечно-сосудистой системе // Пульсовая диагностика тибетской медицины. — Новосибирск: Наука, 1988. — С. 90–108.
125. Лоусон Ч., Хентон Р. Численное решение задач метода наименьших квадратов. — М.: Наука, 1986. — 246 с.
126. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979. — 254 с.
127. Лычак М.М. Идентификация и оценивание состояния объектов управления на основе множественного подхода // Проблемы управления и информатики. — 1999. — № 5. — С. 34–41.
128. Лычак М.М. О решении задачи структурной параметрической идентификации в условиях неопределенности // Автоматика. — 1990. — № 6. — С. 72–77.
129. Малиновский Л.Г. Классификация объектов средствами дискриминантного анализа. — М.: Наука, 1978. — 260 с.

130. Манин Ю.Н. Теорема Геделя // Природа. — 1975. — № 12. — С. 80—87.
131. Мизин И.А., Матвеев А.А. Цифровые фильтры. — М.: Радио и связь. — 1979. — 386 с.
132. Милстайн Л.Б. Методы подавления помех в системах радиосвязи с широкополосными сигналами // Труды инженеров по электротехнике и радиоэлектронике (ТИИЭР). — 1988. — 76, № 6. — С. 19—36.
133. Моттль В.В. Метод частичной аппроксимации в задаче выделения информативных участков экспериментальных кривых // Автоматика и телемеханика. — 1977. — № 1. — С. 97—107.
134. Моттль В.В., Мучник И.Б. Лингвистический анализ экспериментальных кривых // ТИИЭР. — 1979. — 67, № 5. — С. 12—39.
135. Моттль В.В., Мучник И.Б. Скрытые марковские модели в структурном анализе сигналов. — М.: Физматгиз, 1999. — 352 с.
136. Моттль В.В., Мучник И.Б., Яковлев В.Г. Оптимальная сегментация экспериментальных кривых // Автоматика и телемеханика. — 1983. — № 8. — С. 84—95.
137. Мурашко В.В., Струтинский А.В. Электрокардиография. — М.: Медицина, 1991. — 288 с.
138. Мучник И.Б., Мучник Р.Б. Алгоритмы формирования языка для описания экспериментальных кривых // Автоматика и телемеханика. — 1973. — № 5. — С. 86—98.
139. Неймарк Ю.И., Баталова З.С., Образцова Н.Д. К вопросу о выборе признаков при распознавании образов // Изв. АН СССР. Техническая кибернетика. — 1970. — № 1. — С. 105—112.
140. Немирко А.П. Обработка и автоматический анализ электрокардиосигналов // Известия СПб ГЭТУ ЛЭТИ. Биотехнические системы в медицине и экологии. — 2002. — Вып. 1. — С. 34—36.
141. Нильсон Н. Обучающиеся машины. — М.: Мир, 1967. — 180 с.
142. Новиков И.Я., Стечкин С.Б. Основные конструкции всплесков // Фундаментальная и прикладная математика. — 1997. — Т. 3. — № 4. — С. 999—1028.
143. Новый метод анализа ЭКГ больных ревматоидным артритом / Л.С. Файнзильберг, А.Ф. Клубова, Л.А. Стаднюк, И.А. Чайковский, Дитмар Лерхе // Український ревматологічний журнал. — 2001. — № 2. — С. 48—51.
144. Омельченко В.А. Основы спектральной теории распознавания сигналов. — Киев: Вища школа. — 1983. — 156 с.
145. Омельченко В.О. Ортогональні розклади випадкових сигналів і полів. — Киев: УМКВО. — 1991. — 142 с.
146. Орлов А.И. О развитии реалистической статистики // Статистические методы оценивания и проверки гипотез: Сб. тр. — Пермь: ПГУ, 1990. — С. 89—99.
147. Основи інформаційних систем / В.Ф. Ситник, Т.А. Писаревська, Н.А. Єрємін, О.С. Краєва. — К.: КНЕУ, 2001. — 420 с.
148. Пат. № 24516 Україна, МКИ А 61 В 5/024. Спосіб інтегральної оцінки функціонального стану серцево-судинної системи людини / Л.С. Файнзильберг, Т.П. Потапова. — № 97052324; Заявл. 21.05.97; Опубл. 21.07.98, Бюл. № 5. — 4 с.

149. Пат. № 24517 Україна, МКИ А61 В 5/024. Спосіб інтегральної оцінки поточного стану серцево-судинної системи людини / Л.С. Файнзільберг. — № 97052323; Заявл. 21.05.97; Опубл. 30.10.98, Бюл. № 5. — 4 с.
150. *Пилоян Г.О.* Введение в теорию термического анализа. — М.: Наука, 1964. — 231 с.
151. *Повышение* эффективности регистрации формы электрокардиосигнала корреляционной обработкой в цифровой осциллографии / А.М. Беркутов, С.Г. Гуржин, А.А. Дунаев, Е.М. Прошин // Биомедицинские технологии и радиоэлектроника. — 2002. — № 7. — С. 4—13.
152. *Поляк Б.Т.* Устойчивые методы оценки параметров // Структурная адаптация сложных систем. — Воронеж: Воронежский политехнический институт, 1977. — С. 66—71.
153. *Поспелов Д.А.* Когнитивная графика — окно в новый мир // Программные продукты и системы. — 1992. — № 2. — С. 4—6.
154. *Прикладна теорія випадкових процесів і полів* / Під ред. Я.П. Драгана, В.О. Омельченка. — Харків—Львів—Тернопіль: НАН України, 1993. — 378 с.
155. *Применение* структурного анализа кривых к задаче исследования конвертерного процесса / Л.С. Бительман, И.Б. Мучник, Р.Б. Мучник, Р.А. Симсарьян // Изв. вузов. Черная металлургия. — 1971. — № 12. — С. 38—47.
156. *Примин М.А., Недайвода И.В.* Алгоритмы отображения и оценивания магнитокардиосигнала по двумерному распределению вектора плотности тока // Управляющие системы и машины. — 2003. — № 5. — С. 19—25.
157. *Примин М.А., Недайвода И.В., Васильев В.Е.* Новые алгоритмы обработки магнитокардиосигнала // Управляющие системы и машины. — 1998. — № 2. — С. 48—62.
158. *Примин М.А., Недайвода И.В.* Новые алгоритмы обработки биомagnetных сигналов // Управляющие системы и машины. — 1995. — № 3. — С. 3—12.
159. *Пытьев Ю.П.* Методы анализа и интерпретации эксперимента. — М.: Изд-во МГУ, 1989. — 387 с.
160. *Пытьев Ю.П.* О точности и надежности интерпретации эксперимента. // Вестн. МГУ. Сер. физ., астрономия. — 1986. — Т. 27. — С. 14—19.
161. *Пытьев Ю.П.* Точность и надежность интерпретации косвенных измерений // ДАН СССР. — 1987. — 295, № 3. — С. 542—545.
162. *Райбман Н.С., Чадеев В.М.* Построение моделей производства. — Л.: Энергия, 1975. — 375 с.
163. *Ракчеева Т.А.* Когнитивное представление ритмической структуры ЭКГ // Программные продукты и системы. — 1992. — № 2. — С. 38—47.
164. *Распознавание* образов и медицинская диагностика / Под ред. Ю.М. Неймарка. — М.: Наука, 1972. — 328 с.
165. *Растрюгин Л.А., Эренштейн Р.Х.* Метод коллективного распознавания. — М.: Энергоиздат, 1981. — 79 с.
166. *Реброва О.Ю.* Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. — М.: МедиаСфера, 2002. — 312 с.

167. Романовский П.И. Ряды Фурье. Теория поля. Аналитические и специальные функции. Преобразование Лапласа. — М.: Гос. изд. технико-теоретической литературы, 1957. — 291 с.
168. Ронкин М.А., Иванов Л.Б. Реография в клинической практике. — М.: НМФ МБН, 1997. — 250 с.
169. Руководство по внутренним болезням. — 2-е изд., перераб. / Под ред. Н.А. Жукова. — Омск: Омская Государственная медицинская академия, 1998. — 512 с.
170. Санкин А.Н. Линейная дискриминантная функция и линейная зависимость признаков // Математическая обработка медико-биологической информации. — М.: Наука, 1976. — С. 212—217.
171. Сахаров В.Л., Андреев А.С. Методы математической обработки электроэнцефалограмм. — Таганрог: Антон, 2000. — 44 с.
172. Себер Дж. Линейный регрессионный анализ. — М.: Мир, 1980. — 456 с.
173. Сейдж А., Мелс Дж. Идентификация систем. — М.: Наука, 1976. — 495 с.
174. Сергиенко А.Б. Цифровая обработка сигналов. — СПб.: Питер, 2003. — 608 с.
175. Сергиенко И.В., Парасюк И. Н., Каспицкая М.Ф. Об одной нечеткой задаче многопараметрического выбора оптимальных решений // Кибернетика и системный анализ. — 2003. — № 2. — С. 3—15.
176. Сироджа И.Б. Математическое и программное обеспечение интеллектуальных компьютерных систем. — Харьков: ХАИ, 1992. — 100 с.
177. Система компьютерной обработки термограмм / В.И. Скурихин, Л.С. Файнзильберг, Т.П. Потапова, Э.А. Шелковый // Управляющие системы и машины. — 1990. — № 4. — С. 82—88.
178. Скворцов В.А. Примеры метрических пространств. — М.: МЦНМО, 2002. — 24 с.
179. Скурихин В.И., Файнзильберг Л.С., Потапова Т.П. Инструментальная система СИДИГРАФ для синтеза информационных технологий обработки сигналов // Управляющие системы и машины. — 1996. — № 6. — С. 36—45.
180. Скурихин В.И., Файнзильберг Л.С., Потапова Т.П. Когнитивная компьютерная графика как средство интерпретации биоциклических процессов // Управляющие системы и машины. — 1995. — № 4/5. — С. 3—10.
181. Скурихин В.И., Житецкий Л.С., Файнзильберг Л.С. Цифровой анализатор «Углерод» // Управляющие системы и машины. — 1978. — № 2. — С. 99—105.
182. А.с. 1621139 СССР, МКИ Н 03 К 4/48. Следящий аналого-цифровой преобразователь сигналов низкого уровня / Л.С.Файнзильберг. — № 4461093/24; Заявл. 13.06.88; Опубл. 15.01.91, Бюл. № 2. — 3 с.
183. Сосницкий В.Н., Будник Н.Н., Войтович И.Д. Автоматизированные системы для биомагнитных исследований // Управляющие системы и машины. — 1995. — № 3. — С. 31—46.
184. Степашко В.С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей // Автоматика. — 1981. — № 3. — С. 31—36.

185. Степашко В.С., Кочерга Ю.Л. Классификация и анализ помехоустойчивости внешних критериев селекции моделей // Там же. — 1984. — № 4. — С. 38—50.
186. А.с. 1532875 СССР, МКИ 4 G 01 N 33/53. Способ определения сенсibilизации организма бактериальными аллергенами / Н.И. Якуба. — Заявл. 04.01.88; Оpubл. 30.12.89, Бюл. № 48. — С. 227.
187. Тарусов Б.Н., Поливода А.И., Журавлев А.И. Изучение сверхслабой спонтанной хемилюминесценции животных клеток // Биофизика. — 1961. — 6, вып. 4. — С. 490—492.
188. *Технология* // Советский энциклопедический словарь. — М.: Советская энциклопедия, 1988. — С. 1330.
189. Тимченко В.П., Аль-Кхури Талал. Оптоэлектронный датчик для регистрации пульсовой волны на основе биспин-структур // Тез. докл. научн.-техн. конф. «Приборостроение-92». — Винница—Керчь, 1992. — С. 78.
190. Тимченко Л.И., Аль-Кхури Талал. Реографическая цифровая информация // Приборостроение-92: Тез. докл. научн.-техн. конф. — Винница—Керчь, 1992. — С. 68.
191. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1986. — 288 с.
192. Толуевский Ю.П., Нечаев Е.А. Информационные проблемы интенсификации сталеплавильных процессов. — М.: Металлургия, 1978. — 191 с.
193. Туник А.А., Сокол С.А. Оптимальная нестационарная фильтрация на конечном интервале в базисе функций Фурье-Уолша // Кибернетика и вычислительная техника. — 1994. — Вып. 101. — С. 70—74.
194. Ту Дж., Гонсалес Р. Принципы распознавания образов. — М.: Мир, 1978. — 411 с.
195. Турбович И.Т., Гитис В.Г. Дополнительная минимизация числа первичных признаков при использовании информации о параметрах опознающей машины // Опознающие машины. — М.: Наука, 1968. — С. 67—76.
196. Турбович И.Т., Гитис В.Г., Маслов В.К. Опознавание образов. — М.: Наука, 1971. — 246 с.
197. Турчак Л.И. Основы численных методов: Учеб. пособие. — М.: Наука, 1987. — 320 с.
198. Уиллиамс У.Т., Лане Дж.Н. Методы иерархической классификации // Статистические методы для ЭВМ / Под ред. К. Энслейна, Э. Релстона, Г.С. Уилфа. — М.: Наука, 1968. — С. 269—301.
199. А.с. 1332335 СССР, МКИ G 06 F 15/46. Устройство для цифровой обработки аналогового сигнала / Л.С. Файнзильберг. — № 4049749/24—24; Заявл. 03.04.86; Оpubл. 23.08.87, Бюл. № 31. — 11 с.
200. А.с. 1762311 СССР, МКИ G 06 F 15/46. Устройство для цифровой обработки аналогового сигнала / Л.С. Файнзильберг. — № 4896109/24; Заявл. 25.12.90; Оpubл. 15.09.92, Бюл. № 34. — 6 с.
201. А.с. 1830538 СССР, МКИ G 06 G 7/18, G 06 F 15/46. Устройство для оценки скорости изменения аналогового сигнала / Л.С. Файнзильберг. — № 4909509/24; Заявл. 11.02.91; Оpubл. 30.07.93, Бюл. № 28. — 7 с.
202. Файнзильберг Л.С. Диагностика состояния объектов по фазовым траекториям наблюдаемых сигналов с локально сосредоточенными призна-

ками // Проблемы управления и информатики. — 2004. — № 2. — С. 56—67.

203. *Файнзильберг Л.С.* Компьютерный анализ и интерпретация электрокардиограмм в фазовом пространстве // Системні дослідження та інформаційні технології. — 2004. — № 1. — С. 32—46.

204. *Файнзильберг Л.С.* Восстановление эталона циклических сигналов на основе использования хаусдорфовой метрики в фазовом пространстве // Кибернетика и системный анализ. — 2003. — № 3. — С. 20—28.

205. *Файнзильберг Л.С.* Обучаемая система поддержки принятия коллективного решения группы независимых экспертов // Управляющие системы и машины. — 2003. — № 4. — С. 62—67.

206. *Файнзильберг Л.С.* Условия полезности диагностических тестов с позиции теории статистических решений // Проблемы управления и информатики. — 2003. — № 2. — С. 100—111.

207. *Файнзильберг Л.С.* Об алгоритмах самоорганизации в информационных технологиях обработки сигналов // Управляющие системы и машины. — 2003. — № 2. — С. 112—120.

208. *Файнзильберг Л.С.* К вопросу о полезности диагностических методов в задачах скрининга // Управляющие системы и машины. — 2002. — № 6. — С. 10—17.

209. *Файнзильберг Л.С.* Байесова схема принятия коллективных решений в условиях противоречий // Проблемы управления и информатики. — 2002. — № 3. — С. 112—122.

210. *Файнзильберг Л.С.* Адаптивное сглаживание шумов в информационных технологиях обработки физиологических сигналов // Математичні машини і системи. — 2002. — № 3. — С. 96—104.

211. *Файнзильберг Л.С.* Частотно-избирательная фильтрация в информационных технологиях обработки сигналов // Управляющие системы и машины. — 2002. — № 2. — С. 54—59.

212. *Файнзильберг Л.С.* Информационная технология для диагностики функционального состояния оператора // Там же. — 1998. — № 4. — С. 40—45.

213. *Файнзильберг Л.С.* Объектно-ориентированная модель данных в информационных технологиях обработки сигналов // Там же. — 1998. — № 3. — С. 37—41.

214. *Файнзильберг Л.С.* Синтез информационных технологий обработки сигналов в задачах косвенного контроля и диагностики // Там же. — 1998. — № 2. — С. 40—47.

215. *Файнзильберг Л.С.* Оценка полезности признаков при решении задач диагностики в статистической постановке // Математические машины и системы. — 1998. — № 1. — С. 57—64.

216. *Файнзильберг Л.С., Потапова Т.П.* Комплекс аппаратных и программных средств для контроля технологических параметров железоуглеродистых расплавов // Металлургия и горнорудная промышленность. — 1992. — № 1. — С. 70—72.

217. *Файнзильберг Л.С.* Об одном подходе к преобразованию сигналов низкого уровня в числовой эквивалент // Управляющие системы и машины. — 1991. — № 6. — С. 10—15.

218. Файнзильберг Л.С., Шкляр Г.А. Об оценке полезности признаков при статистическом распознавании двух классов // Автоматика. — 1985. — № 5. — С. 78—82.
219. Файнзильберг Л.С. Метод градуировки аппаратуры для термического анализа содержания углерода в стали // Заводская лаборатория. — 1987. — № 1. — С. 33—35.
220. Файнзильберг Л.С. Об одном подходе к задаче обучения распознаванию двух классов по конечному числу наблюдений // Автоматика. — 1978. — № 5. — С. 46—49.
221. Файнзильберг Л.С. Применение методов статистического распознавания в термографическом анализе состава металла // Кибернетика. — 1978. — № 6. — С. 133—136.
222. Файнзильберг Л.С. К вопросу о безошибочном распознавании двух классов по совокупности пересекающихся признаков // Там же. — 1982. — № 4. — С. 104—109.
223. Файнзильберг Л.С. Новый подход к термическому анализу жидкой стали на углерод // Изв. вузов. Черная металлургия. — 1980. — № 6. — С. 113—119.
224. Файнзильберг Л.С. Новая информационная технология обработки ЭКГ для выявления ишемической болезни сердца при массовых обследованиях населения // Управляющие системы и машины. — 2005. — № 3. — С. 63—71.
225. Фрумин Л.Л., Штарк М.Б. О фазовом портрете электрокардиограммы // Автометрия. — 1993. — № 2. — С. 51—54.
226. Фу К. Структурные методы в распознавании образов. — М.: Мир. — 1977. — 320 с.
227. Фукунага К. Введение в статистическую теорию распознавания образов. — М.: Наука, 1979. — 368 с.
228. Хайлимен В.Х. Линейные решающие функции и их применение для распознавания образов // Труды института радиоинженеров. — М.: Мир, 1962. — Т. 50. — № 6. — С. 67—81.
229. Халфен Э.Ш., Сулковская Л.С. Клиническое значение исследования скоростных показателей зубца Т ЭКГ // Кардиология. — 1986. — № 6. — С. 60—62.
230. Хьюбер П. Робастность в статистике: Пер. с англ. — М.: Мир, 1984. — 304 с.
231. Циммерман Ф. Клиническая электрокардиография. — М.: Биком. — 1998. — 448 с.
232. А.с. 1374247 СССР, МКИ G 06 F 15/46. Цифровое устройство для анализа химического состава чугуна / Л.С. Файнзильберг, Ю.В. Власенко, Э.З. Трайнин, Э.Х. Тухин, Э.А. Шелковый. — № 3911009/24-24; Заявл. 17.06.85; Оpubл. 15.02.88, Бюл. № 6. — 14 с.
233. Цыпкин Я.З. Адаптация и обучение в автоматических системах. — М.: Наука, 1968. — 400 с.
234. Цыпкин Я.З. Основы теории обучающихся систем. — М.: Наука, 1970. — 252 с.
235. Цыпкин Я.З., Кельманс Г.К. Адаптивный байесов подход // Проблемы передачи информации. — 1970. — 6, вып. 1. — С. 52—59.

236. Чайковский И.А., Батушкин В.В., Файнзильберг Л.С. и др. Эффективность оценки течения острого коронарного синдрома по данным анализа первого отведения ЭКГ на фазовой плоскости // Журнал Академії медичних наук. — 2007. — **13**, № 1. — С. 104—113.
237. Черноусько Ф.Л. Оптимальные гарантированные оценки неопределенностей с помощью эллипсоидов // Изв. АН СССР. Техническая кибернетика. — 1980. — № 3. — С. 3—11.
238. Ширяев А.Н. Задача скорейшего обнаружения нарушений стационарного режима // ДАН СССР. — 1961. — **138**, № 5. — С. 1039—1042.
239. Шлезингер М.И. Математические средства обработки изображений. — Киев: Наук. думка, 1989. — 200 с.
240. Юрачковский Ю.П. Моделирование полигармонических многомерных процессов // Автоматика. — 1985. — № 5. — С. 39—41.
241. Юрачковский Ю.П. Восстановление полиномиальных зависимостей на основе самоорганизации // Там же. — 1981. — № 4. — С. 15—20.
242. Яковлев В.Г. Алгоритм выделения всплесков на физиологических кривых // Автоматика и телемеханика. — 1977. — № 12. — С. 94—105.
243. Abboud S., Sadeh D. The waveform's alignment procedure in the averaging process for external recording of the HIS bundle activity // Comp. Biomed. Res. — 1982. — N 15. — P. 212—219.
244. Aho A.V., Ullman I.D. The theory of languages // J. of Math. System Theory. — 1968. — N 2. — P. 97—128.
245. Akaike H. Information measures and model selection // Bull. of the Intern. Statistical Inst.: Proc. 44th Session. — 1983. — P. 277—290.
246. Akaike H. Fitting autoregressive model for prediction // Ann. Stat. — 1969. — Vol. 21. — P. 243—247.
247. Akay M. Wavelet applications in medicine // IEEE Spectrum. — 1997. — **34**, N 5. — P. 50—56.
248. Allen J.B., Rabiner L.R. A unified approach to short-time Fourier analysis and synthesis // Proc. IEEE. — 1977. — **65**, N 11. — P. 1558—1564.
249. Allen J.B. Short-term spectral analysis, synthesis, and modification by discrete Fourier transform // IEEE Trans. on Acoustics, Speech, Signal Processing. — 1997. — Vol. ASSP-25. — N 3. — P. 235—238.
250. Applying time-frequency analysis to seizure EEG activity / S. Blanco, S. Kochen, O.A. Rosso, P. Saldado // IEEE Engineering in Medicine and Biology. — 1997. — **16**, N 1. — P. 64—71.
251. Bagui S.C., Pal N.R. A multistage generalization of the rank nearest neighbour classification rule // Pattern Recognition Letters. — 1995. — **16**, N 6. — P. 801—814.
252. Bailey J.J., Berson A.S., Garson A. Recommendation for standardization and specification in automated electrocardiography: bandwidth and digital signal processing: A report for health professionals by an ad hoc writing group of the committee on electrocardiography and cardiac electrophysiology of the Council on Clinical Cardiology // American Heart Association. Circulation. — 1990. — Vol. 81. — 730 p.
253. Balas E., Niehaus W. Optimized crossover-based genetic algorithms for the maximum cardinality and maximum weight clique problems // J. Heuristics. — 1998. — **4**, N 4. — P. 107—122.

254. *Bansal R.K., Papantoni-Kazakov P.* An algorithm for detecting a change in stochastic process // *IEEE Trans. on Inform. Theory.* — 1986. — IT-32. — N 2. — P. 227–235.
255. *Baumgartner C., Deecke L., Stroink G.* Biomagnetism: Fundamental Research and Clinical Application. — Amsterdam: Elsevier, 1995. — 846 p.
256. *Ben-Bassat M.* Irrelevant features in pattern recognition // *IEEE Trans. Comput.* — 1978. — Vol. C-27. — N 8. — P. 749–766.
257. *Bertrand O., Bohorques J., Pernier J.* Time-frequency digital filtering based on an invertible wavelet transform: An application to evoked potentials // *IEEE Trans. on BME.* — 1994. — **41**, N 1. — P. 77–88.
258. *Bezdek J.C.* A note on generalized self-organizing network algorithms // *SPIE Applications of Artificial Intelligence VIII.* — 1990. — Vol. 1293. — P. 260–267.
259. *Biermann A.W.* A grammatical inference program for linear languages // *Intern. Conf. of System Science (Hawaii, 1971).* — 1971. — Vol. S1. — P. 117–141.
260. *Bigger G.T., Fleiss J.L., Steinman R.C. et al.* Frequency domain measures of heart rate period variability and mortality after myocardial infarction // *Circulation.* — 1992. — Vol. 85. — P. 164–171.
261. *Blacher J., Asmar R., Djane S. et al.* Aortic pulse wave velocity as a marker of cardiovascular risk in hypertensive patients // *Hypertension.* — 1999. — Vol. 33. — P. 1111–1117.
262. *Bracewell R.* The Fourier Transform and Its Applications, 3rd ed. — New York: McGraw-Hill, 1999. — 413 p.
263. *Brigham E.O.* The Fast Fourier Transform and Applications. Englewood Cliffs. — NJ: Prentice Hall. — 1988. — 382 p.
264. *Brown R.G., Meyer R.F.* The fundamental theorem of exponential smoothing // *Oper. Res.* — 1961. — **9**, N 5. — P. 23–34.
265. *Cho S.B., Kim J.H.* Multiple network fusion using fuzzy logic // *IEEE Trans. on Neural Networks.* — 1995. — **6**, N 2. — P. 497–501.
266. *Chu C.Y.* Genetic algorithm approach to the configuration of wavelet packet processing tree // *Proc. 12th Int. Conf. on Pattern Recognition.* — Jerusalem (Israel). — 1994. — Vol. 3. — P. 53–56.
267. *Cognitive Graphic Image of Electrocardiosignal as Means for Functional Human State Interpretation* / L.S., Fainzilberg T.I. Aksenova, T.P. Potapova, V.Ju. Shelekhova // *Intern. J. of Inform. Theories and Applications.* — 1994. — Vol. 2. — P. 20–26.
268. *Cohen A.* Biomedical Signals: Origin and Dynamic Characteristic; Frequency-Domain Analysis: The Biomedical Engineering Handbook Editor-in-Chief J.D. Bronzino. — Boca Raton (Florida): CRC and IEEE Press, 1995. — P. 805–827.
269. *Cooley J.W., Tukey O.W.* An algorithm for the machine calculation of complex Fourier series // *Math. Comput.* — 1965. — N 19. — P. 297–301.
270. *Cremer P., Leuwen P., Hailer B. et al.* Changes in magnetic field during repolarization in patients with coronary artery disease // *Med. & Biol. Engineering & Computing.* — 1999. — N 37. — Suppl. 2. — P. 1480–1481.
271. *Crochiere R.* A weighted overlap-add method of short-time Fourier analysis/synthesis // *IEEE Trans. on Acoustics, Speech, Signal Processing.* — 1980. — Vol. ASSP-28. — N 2. — P. 99–102.

272. *Design of Experiments and Data Analysis: New Trends and Results* / Letzky E.K., Voshinin A.P., Dyvak N.P., Simoff S.J., Orlov A.I., Gorsky V.G., Nikitina E.P., Nosov V.N. / Edited by E.K. Letzky. — Moscow: ANTAL. — 1993. — 192 p.
273. *Detection of coronary artery disease in patients with normal or unspecifically changed ECG on the basis of magnetocardiography* / I. Chaikovsky, J. Kohler, Th. Hecker, B. Hailer, S. Auth-Eisernitz, V. Sosnitsky, L. Feinzilberg, N. Budnik, F. Steinberg. — In: *Biomag 2000* // Proc. of the 12th Internat. Conf. on Biomagnetism / Eds. J. Nenonen, R. Ilmoniemi, T. Katila. — Helsinki Univ. of Technology. — Espoo, 2000. — P. 565–568.
274. *Dvoretzky A. On stochastic approximation* // Proc. of the Third Berkeley Symp. on Math. Statistics and Probability. — 1956. — Vol. 1. — P. 23–31.
275. *Eckmann J.P., Ruelle D. Ergodic theory of chaos and strange attractors* // Review of modern physics. — 1985. — Vol. 57. — P. 617–656.
276. *Fainzilberg L.S. Heart functional state diagnostic using pattern recognition of phase space ECG-images* // Proc. 6th European Congress on Intelligent Techniques and Soft Computing (EUFIT '98). — Aachen (Germany). — 1998. — 3, N B-27. — P. 1878–1882.
277. *Fainzilberg L.S. Interconnection between features properties and probability of error in statistical recognition of two classes* // Proc. 12th Intern. Conf. on Pattern Recognition (ICPR'94). — Vol. 2. — Jerusalem (Israel): IEEE Computer Society Press (LA, CA). — 1994. — P. 544–546.
278. *Fainzilberg L.S., Potapova T.P. Computer analysis and recognition of cognitive phase space electro-cardiographic image* // Proc. 6th Intern Conf. on Computer Analysis of Images and Patterns (CAIP'95). — Prague (Czech Republic). — 1995. — P. 668–673.
279. *Fainzilberg L.S. Why relevant features may be unusefull in statistical recognition of two classes* // Proc. 13th Intern. Conf. on Pattern Recognition (ICPR'96). — Viena (Austria). — 1996. — P. 730–734.
280. *Fainzilberg L.S. Information technologies of signal Processing and novel feature's selection in medical application* // 54th ICB Seminar *Multimedia, Data Integration, Medical Databases*. — Warsaw (Poland). — 1999. — P. 15–16.
281. *Fainzilberg L.S. ECG averaging based on hausdorff metric* // Intern. J. of Biomagnetism. — 2003. — 5, N 1. — P. 236–237.
282. *Пат. США № 4198679, МКП G 01N (USSR)*. — Оpubл. 15.04.80. L.S. Fainzilberg. Methods and device for discriminating thermal effect of phase transformation of metals of metals and alloys in the process of their cooling.
283. *Fainzilberg L.S. Nowa metoda interpretacji zapisu EKG w balaniach skriningowych oraz w opiece domowej* // *Zdrowie publiczne*. — 2005. — 115, N 4. — P. 458–464.
284. *Fainzilberg L., Chaikovsky I., Auth-Eisernitz S. et al. Sensitivity and specificity of magnetocardiography using computerized classification of current density vectors maps in ischemic patients with normal ECG and echocardiogram* // Intern. Congresses Series, 2007. — Vol. 1300. — P. 468–471.
285. *Fraser A.M., Swinney H.L. Independent coordinates for strange attractors from mutual information* // *Physical Review Letters*. — 1986. — Vol. A 33. — P. 1134–1140.

286. Furno G., Tompkins W. QRS detection using automata theory in battery-powered microprocessor system // Proc. IEEE Frontiers Eng. Health Care. — 1982. — Vol. 9. — P. 155.
287. Fu K.S., Huang T. Stochastic grammar and languages. — Intern. J. of Computer and Inf. Sci. — 1972. — N 2. — P. 135–175.
288. Fuzzy additive hybrid operators for network-based decision making / Keller J.M., Krishnapuram R., Chen Z., Nasraoui O. // Int. J. Intelligent Systems. — 1994. — N 9. — P. 1001–1023.
289. Gabor D. Perspectives of Planning. Organization of Economic Cooperation and Development. — London: Emp. College of Sci. and Technology, 1971. — 425 p.
290. Galen R.S., Gambino S.R. Beyond normality: The predictive value and efficiency of medical diagnoses. — New York: Wiley and Sons Inc., 1975. — 237 p.
291. Goldberger A.L. Fractal mechanisms in the electrophysiology of the heart // IEEE Eng. Med. Biol. — 1992. — N 11. — P. 47–52.
292. Gonzalez R.C. Syntactic pattern recognition — introduction and survey // Proc. National Electronic Conf. — 1972. — 27, N 1. — P. 27–32.
293. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors // Physics. — 1983. — Vol. D 9. — P. 189–208.
294. Gray R., Begg C.B., Greenes R.A. Construction of receiver operating characteristic curves when disease verification is subject to selection bias // Medical Decision Making. — 1984. — N 4. — P. 151–164.
295. Griebach S. Formal parsing system // Comm. ASM. — 1964. — 7, N 8. — P. 499–516.
296. Gustafsson F. Determining the initial states in forward-backward filtering // IEEE Trans. on Signal Processing. — 1996. — Vol. 44. — Issue 4. — P. 988–992.
297. Hansen L.K., Salamon P. Neural network ensembles // IEEE Trans. on Pattern Anal. and Machine Intell. — 1990. — 12, N 10. — P. 993–1001.
298. Haralick R.M. Structural pattern recognition homomorphisms and arrangements // Pattern recognition. — 1978. — 10, N 3. — P. 223–236.
299. Hashem S., Schmeiser B. Improving model accuracy using optimal linear combinations of trained neural networks // IEEE Trans. on Neural Networks. — 1995. — 6, N 3. — P. 792–794.
300. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use // European Heart J. — 1996. — N 17. — P. 354–381.
301. Hild Ch.R., Bozdogan H. The Use of information-based model selection criteria in the GMDH algorithms // SAMS. — 1995. — 20, N 1–2. — P. 29–50.
302. Ho T.K., Hull J.J., Srihari S.N. Decision combination in multiple classifier systems // IEEE Trans. on Pattern Analysis and Machine Intelligence. — 1994. — 16, N 1. — P. 66–75.
303. Holt C.C. Forecasting trend and seasonal by exponentially weighted moving averages // O.N. R. Memorandum, Carnegie Inst. of Technology. — 1957. — N 52. — P. 17–23.
304. <http://liinwww.ira.uka.de/bibliography/theory/wavelets/wlet.html>

305. *Image coding using wavelet transform* / M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies // IEEE Trans. on Image Processing. — 1992. — **1**, N 2. — P. 205—220.
306. *Jura St., Szelkowsky E.A., Maslennikov S.N.* Zabielenie zeliwa niskostopowego // Krzepniecie metali i stopow. — Gliwice, 1988. — S. 76—83.
307. *Jura St., Szelkowsky E.A., Maslennikov S.N.* Jak uzyskac wyzsza precyzje oceny zawartosci pierwiastkow w zeliwie przy pomocy metody ATD // Ibid. — S. 37—43.
308. *Kaiser W., Findeis M.* Artifact processing during exercise testing // J. of Electrocardiology. — 1999. — Vol. 32, Supplement. — P. 212—219.
309. *Kale B.K.* Trimmed means and the method of maximum likelihood when spurious observations are present // Proc. Conf. at Dalhousie Univ. — Amsterdam: North-Holland Publisher's Company, 1975. — P. 177—185.
310. *Kanjilal P.P., Bhattacharya J., Saha G.* Robust method for periodicity detection and characterization of irregular cyclical series in terms of embedded periodic components // Physical Review. — 1999. — Vol. 59. — P. 4013—4025.
311. *Kaplan J.L., Yorke J.A.* Chaotic behavior of multidimensional difference equations // Lectures Notes in Mathematics. — 1979. — Vol. 730. — P. 204—227.
312. *Khoros Programmer's Manual.* — Univ. of New Mexico, 1992. — 289 p.
313. *Kimura F., Shridhar M.* Handwritten numerical recognition based on multiple algorithms // Pattern Recognition. — 1991. — **24**, N 10. — P. 969—983.
314. *Kittler J.* Feature selection and extraction // Handbook of Pattern Recognition and Image Processing. — 1986. — 450 p.
315. *Krogh A., Vedelsby J.* Neural network ensembles, cross validation, and active learning // Advances in neural information processing systems. — Cambridge MA: MIT Press, 1995. — 278 p.
316. *Laguna P., Jane R., Caminal P.* Automatic detection of wave boundaries in multi-lead ECG signals: Validation with the CSE database // Comp. Biomedical Research. — 1994. — **27**, N 1. — P. 45—60.
317. *Lewis P.M.* The characteristic selection problem in recognition system // IRE Trans. Inform. Theory. — 1962. — **8**, N 2. — P. 171—178.
318. *Lorden S.* Procedures for reacting to change in distribution // Ann. Math. Stat. — 1971. — **4**, N 2. — P. 1897—1971.
319. *Lorenz E.N.* Deterministic nonperiodic flow // J. of Atmosphere Sci. — 1963. — N 20. — P. 130—141.
320. *Malik M., Camm A.J.* Components of heart rate variability: what they really mean and what we really measure // American J. of Cardiology. — 1993. — N 72. — P. 821—822.
321. *Mallat S.* A theory for multiresolution signal decomposition: the wavelet representation // IEEE Pattern Anal. and Machine Intell. — 1989. — **11**, N 7. — P. 674—693.
322. *Mallows C.L.* Some comments on Cp // Technometrics. — 1973. — Vol. 15. — P. 661—667.
323. *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. — **11**, N 2. — P. 342—344.
324. *Mathcad 8.* User's Guide. — Cambridge, USA: Math. Soft. Inc., 1998. — 372 p.

325. *Matteo Frigo, Steven G. Johnson*. FFTW: An adaptive software architecture for the FFT // Proc. IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing. — Seattle: WA 1998. — Vol 3. — P. 1381–1384.
326. *McNeil B.J., Hanley J.A.* Statistical approaches to the analysis of receiver operating characteristic (ROC) curves // Medical Decision Making. — 1984. — N 4. — P. 137–150.
327. *Methodology of ECG Interpretation in the Hanover Program* / Zywniez C., Borovsky D., Goettsch G., Joseph G. // Meth. Inf. Med. — 1990. — N 29. — P. 375.
328. *Metz C.E.* Fundamental ROC Analysis. Progress in Medical Physics and Psy-Chophysics // Handbook of Medical Imaging (Vol. 1). — Bellingham: WA: SPIE Press, 2000. — P. 754–769.
329. *Mostafa A.E.S., Abdel-Kader M., El-Osmay A.* Improvements of anti-jam performance of spread-spectrum system // IEEE Trans. Commun. — 1983. — Vol. COM-31. — P. 803–808.
330. *Müller J.-A., Lemke F.* Self-organizing modelling and decision support in economics // Proc. IMACS'95. — Gordon and Breach Publ. — 1995. — P. 135–138.
331. *Narendra P.M., Fukunaga K.A.* Branch and bound algorithm for feature subset selection // IEEE Trans. Comp. — 1977. — Vol. C-26. — N 9. — P. 917–922.
332. *Neter J., Wasserman W., Kutner M.H.* Applied linear regression models (2nd ed.). — Homewood, IL: Irwin, 1989. — 376 p.
333. *Nikiforov I.V.* Sequential detection of changes in stochastic processes // Preprints 9th IFAC/IFORS Symp. Identification and System Parameter Estimation. — Budapest. — 1991. — Vol. 1. — P. 11–19.
334. *Ogawa J., Cho S.J., Morinaga N., Namekawa T.* Optimum detection of M-ary RSK signal in the presence of CW interference // Trans. IECE Japan. — 1981. — Vol. E 64. — P. 800–806.
335. *On combining classifiers* /J. Kittler, M. Hatef, R.P.W. Duin, J. Matas // IEEE Trans. on Pattern Anal. and Machine Intell. — 1998. — N 20. — P. 226–239.
336. *Oppenheim A.V., R.W. Schaffer.* Discrete-Time Signal Processing. — Englewood Cliffs. — NJ: Prentice-Hall. — 1989. — P. 311, 312.
337. *Oppenheim A.V., Schaffer R.W.* Time Signal Processing. — New York: Upper Sadle River, 1999. — P. 453–574.
338. *Pan J., Tompkins W.J.* A real-time QRS detection algorithm // IEEE Trans. Biomed. Eng. — 1985. — Vol. BME-32. — N 3. — P. 230–236.
339. *Pan X., Metz C.E.* The «proper» binormal model: parametric ROC curve estimation with degenerate data // Acad. Radiology. — 1997. — N 4. — P. 380–389.
340. *Patterson J.D., Womack B.F.* An adaptive pattern classification system. — IEEE Trans. on System Science and Cybernetics. — 1966. — 2, N 1. — P. 62–67.
341. *Patterson J.D., Wagner T.J., Womack B.F.* A mean-square performance criterion for adaptive pattern classification systems. — IEEE Trans. Automatic Control. — 1967. — 12, N 2. — P. 195, 196.
342. *Partridge E.M.* A linear vector recognition function // IRE Trans. Electronic Comp. — 1962. — Vol. EC-11. — N 1. — P. 35–49.

343. *Pavaldis T.* Linguistic analysis of waveforms. — *Software Eng.* — 1971. — Vol. SE-2. — N 2. — P. 48–53.
344. *Pesquet J.C., Krim H., Carfatan H.* Time-invariant orthonormal wavelet representations // *IEEE Trans. Sign. Proc.* — 1996. — **44**, N 8. — P. 1964–1970.
345. *Pettitt A.N.* A non-parametric approach to the change-point problem // *Application on Statistics.* — 1979. — N 28. — P. 126–135.
346. *Pitas I., Venetsanosoulos A.N.* Nonlinear order statistic filters for img filtering and edge detection // *North Holland Signal Processing.* — 1986. — N 10. — P. 395–413.
347. *Poli R., Cagnoni S., Valli G.* Genetic design of optimum linear and non-linear QRS detectors // *IEEE Trans. on Biomedical Eng.* — 1995. — **42**, N 11. — P. 1137–1141.
348. *Polikar R.* Wavelets tutorial. — <http://engineering.rowan.edu/~polikar/homepage.html>.
349. *Possibilities of magnetocardiography in coronary artery disease detection in patient with normal or unspecifically changes ECG* / I. Chaikovsky, F. Steinberg, B. Heiler, S. Auth-Eisernith, Th. Hecker, V. Sosnitsky, N. Budnik, L. Feinzilberg // *Proc. 3rd Int. Congress on Coronary Artery Disease.* — Lyon (France), October 2–5, 2000. — P. 415–421.
350. *Pranke J., Mandler E. A.* Comparison of two approaches for combining the votes of cooperating classifiers // *Proc. 11th IAPR Intern. Conf. on Pattern Recognition*, 1992. — Vol. 2. — P. 611–614.
351. *Rachkovskij D.A., Kussul E.M.* DataGen: a generator of datasets for evaluation of classification algorithms // *Pattern Recognition Letters.* — 1998. — **19**, N 7. — P. 537–544.
352. *Ramirez R.W.* The FFT: Fundamentals and Concepts. Englewood Cliffs. — NJ: Prentice-Hall. — 1985. — 29 p.
353. *Richter M., Schreiber T.* Phase space embedding of electrocardiograms // *Physical Review.* — 1998. — Vol. E 58 (6392). — P.1–7.
354. *Robbins H., Monro S.* A stochastic approximation method // *Ann. of Math. stat.* — 1951. — **22**, N 1. — P. 38–51.
355. *Romanovich S.* Reconstruction of three-components dipoles within layer // *Biomedizinische Technick.* — 1997. — Vol. 42(1). — P. 227–230.
356. *Romanovich S.* Supercurrents: Forward & Inverse Problems // *Proc. 6th Int. Conf. Superconductive Electronics.* — Berlin (Germany), 1997. — P. 344–346.
357. *Rosher B.* On the detection of many outliers // *Technometrics.* — 1975. — Vol. 17. — P. 221–227.
358. *Rossler O.E.* An equation for continuous chaos // *Physical Review Letters.* — 1976. — Vol. A 57. — P. 397–398.
359. *Sauer T., Yorke J., Casdagli M.* Embedology // *J. Stat. Physics.* — 1991. — Vol. 65. — P. 579–616.
360. *Sawaragi Y., Soeda T. et al.* Statistical prediction of air pollution levels using non-physical models // *Automatica (IFAC).* — 1979. — **15**, N 4. — P. 441–452.
361. *Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis* / Ivanov P.C., Rosenblum M.G., Peng C.K., Mietus J., Havlin S., Stanley H.E., Goldberger A.L. // *Nature.* — 1996. — **383**, N 26. — P. 323–327.

362. *Shankar R.* The GMDH. Master Thesis : Univ. of Delaware, 1972. — 250 p.
363. *Shelekhova V.Yu.* Harmonic algorithm GMDH for large data volume // SAMS. — 1995. — Vol. 20. — P. 117–126.
364. *Siedlecki W., Sklansky J.* On automatic feature selection // Int. J. of Pattern Recognition and Artificial Intelligence. — 1988. — Vol. 2. — P. 197–220.
365. *Small M., Yu D., Harrison R.G.* Surrogate test for pseudoperiodic time series data // Physical Review Letters. — 2001. — **87**, N 18. — P. 34–37.
366. *Sosnitsky V.N., Voitovich I.D.* Application of superconducting electronics to registration of biomagnetic signals // J. de Physique IV Colloque 3, supplement au J. de Physique III (Belgium). — 1996. — Vol. 6. — P. 3417–3422.
367. *SQUID-magnetometer* system used to diagnostics of some cardiac diseases / A. Gapelyuk, Yu. Minin, M. Primin, V. Sosnitsky, I. Voytovych, P. Sutkovoij // Int. Conf. EUCA'93. — Gottingen (Germany), 1993. — P. 462.
368. *Srikanth T., Napper S.A., Gu H.* Bottom-up approach to uniform feature extraction in time and frequency domains for single-load ECG signal // Inter. J. of Bioelectromagnetism. — 2002. — Vol. 4. — P. 21–29.
369. *Stearns S.D.* On selecting features for pattern classifiers // Proc. Third Int. Conf. on Pattern Recognition (Coronado, CA), 1976. — P.71–75.
370. *Takens F.* Detecting Strange Attractors in Turbulence // Lecture Notes in Mathematics. — New York, 1981. — P. 366–381.
371. *Theiler J.* Estimating the fractal dimension of chaotic time series // Lincoln Lab. J. — 1990. — N 3. — P. 63–86.
372. *Tissari S., Rahola J., Nenonen J.* Source localization accuracy in MEG using the Galerkin method // Proc. 13th Int. Conf. on Biomagnetism, BIOMAG 2002. — Jena (Germany), 2002. — P. 794–796.
373. *Tou J.T., Heydorn R.P.* Some approaches to optimum feature selection // Computers and Inform. Science Academy Press. — 1967. — **11**, N 4. — P. 57–89.
374. *Trahaniyas E.* An approach to QRS-complex detection using mathematical morphology // IEEE Transactions on Biomedical Eng. — 1993. — **40**, N 2. — P. 201–205.
375. *Tree-Structured* non-linear Filter and Wavelet Transform for Microcalcification Segmentation in Digital Mammography / Clarke L.P., Kallergi M., Qian W., Li H.D., Clark R.A., Sibiger M.L. // Cancer Letters. — 1994. — N 77. — P. 173–181.
376. *Vasetsky Yu., Chaikovsky I.* Model of current distribution generated by extended electrical sources and its using for current maps classification // Proc. 13th Int. Conf. on Biomagnetism (BIOMAG 2002). — Jena (Germany), 2002. — P. 846–848.
377. *Vetterli M., Kovacevic J.* Wavelets and Subband Coding. New Jersey: Prentice Hall. — 1995. — 487 p.
378. *Wang Xiufeng L.B.* Self-organizing methods modelling with «concentrated» variables // SAMS. — 1995. — **20**, N 1–2. — P. 51–58.
379. *Whittaker E.T., Robinson G.* Graduation or the Smoothing of Data // The Calculus of Observations, Ch. 11: A Treatise on Numerical Mathematics, 4th ed. — New York: Dover, 1967. — P. 285–316.
380. *Whittaker E.T., Watson G.N.* Quasi-Periodic Functions // A Course in Modern Analysis. — Cambridge (England): Cambridge University Press, 1990. — P. 445–447.

381. *Willems J., Abreu-Lima C. et al.* Evaluation of ECG interpretation results obtained by computer and cardiologist // *Meth. Inf. Med.* — 1990. — N 29. — P. 308–316.
382. *Winters P.R.* Forecasting sales by exponentially weighted moving averages // *Management Science.* — 1960. — 6, N 3. — P. 34–46.
383. *Wolpert D.H.* Stacked generalization // *Neural Networks.* — 1992. — 5, N 2. — P. 241–260.
384. *Woods K.S., Bowyer K., Kergelmeyer W.P.* Combination of multiple classifiers using local accuracy estimates // *Proc. of CVPR'98.* — 1998. — P. 391–396.
385. *Xu L., Krzyzak A., Suen C.Y.* Methods of combining multiple classifiers and their applications to handwriting recognition // *IEEE Trans. SMC.* — 1992. — 22, N 3. — P. 418–435.
386. *Xue Q., Hu Y.H., Tompkins W.J.* Neural-network-based adaptive matched filtering for QRS detection // *IEEE Trans. on Biomedical Eng.* — 1992. — 39, N 4. — P. 317–328.
387. *Yang F., Liao W.* Modeling and decomposition of HRV signals with wavelet transforms // *IEEE Eng. in Medicine and Biology.* — 1997. — 16, N 4. — P. 17–22.
388. *Ziener R.E., Peterson R.L.* Digital Communications and Spread Spectrum. — New York: MacMillan, 1985. — 386 p.
389. *Zywienz C., Borovsky D., Goettsch G., Joseph G.* Methodology of ECG Interpretation in the Hanover Program // *Meth. Inf. Med.* — 1990. — N 29. — P. 375.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
ГЛАВА 1	
ПРОБЛЕМА ОЦЕНКИ СОСТОЯНИЯ ОБЪЕКТОВ В УСЛОВИЯХ НЕПОЛНОЙ ИНФОРМАЦИИ	11
1.1. Общая формулировка проблемы	11
1.2. Методы построения моделей сложных систем по экспериментальным данным	16
1.3. Методы построения диагностических правил	22
1.4. Анализ свойств сигналов как метод формирования диагностических признаков	30
1.4.1. Анализ гармонических компонент сигнала	31
1.4.2. Структурный анализ сигналов сложной формы	42
1.4.3. Анализ свойств сигналов в фазовом пространстве	48
1.5. Задача оценки эффективности диагностических признаков	51
1.6. Современные подходы к программной реализации информационных технологий	53
ГЛАВА 2	
МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ПОСТРОЕНИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ОБРАБОТКИ СИГНАЛОВ СЛОЖНОЙ ФОРМЫ	57
2.1. Особенности компьютерной обработки временных сигналов с признаками, сосредоточенными на локальных фрагментах	57
2.2. Примеры сигналов с локально сосредоточенными признаками	61
2.2.1. Электрокардиограмма, несущая информацию о состоянии сердечно-сосудистой системы живого организма	61

2.2.2. Электрофизиологические сигналы, также используемые в кардиологической практике	67
2.2.3. Термическая кривая, несущая информацию о свойствах многокомпонентного материала	72
2.3. Модели порождения сигналов с локально сосредоточенными признаками	74
2.3.1. Базовая модель	75
2.3.2. Модель порождения циклических сигналов	77
2.3.3. Генератор циклических последовательностей	79
2.4. Декомпозиция общей задачи компьютерной обработки сигналов с локально сосредоточенными признаками	81

ГЛАВА 3

МЕТОД ВОССТАНОВЛЕНИЯ ПОЛЕЗНОГО СИГНАЛА ПО ФАЗОВЫМ ТРАЕКТОРИЯМ ИСКАЖЕННЫХ НАБЛЮДЕНИЙ	85
3.1. Метод усреднения реализаций во временной области	85
3.2. Метод усреднения реализаций в фазовом пространстве	90
3.2.1. Основная идея метода	90
3.2.2. Базовый алгоритм усреднения траекторий	91
3.2.3. Модифицированный алгоритм усреднения	93
3.3. Алгоритмы обработки циклических сигналов	95
3.3.1. Процедура разбиения на циклы	95
3.3.2. Селекция нетипичных циклов	98
3.3.3. Модельный пример	100
3.4. Результаты численного моделирования	102

ГЛАВА 4

БАЗОВЫЕ ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ОБРАБОТКИ СИГНАЛОВ СЛОЖНОЙ ФОРМЫ	106
4.1. Анализ методов подавления аддитивных помех	106
4.1.1. Общие замечания	106
4.1.2. О критериях оптимальности процедур фильтрации	107
4.1.3. Характерные типы внешних возмущений	109
4.2. Принципы построения инструментальной системы	110
4.2.1. Концептуальная идея	110
4.2.2. Архитектура инструментальной системы	113
4.2.3. Объектно-ориентированные модели ядра инструментальной системы	119
4.2.4. Базовые компоненты инструментальной системы	125
4.3. Эффективная процедура фильтрации сосредоточенных гармонических помех	129
4.3.1. Общие замечания, касающиеся частотно-избирательной фильтрации	129
4.3.2. Предлагаемый алгоритм	135
4.3.3. Результаты модельных экспериментов	137
4.4. Эффективная процедура подавления ограниченных по уровню случайных помех	144

4.4.1. Анализ традиционных алгоритмов сглаживания	144
4.4.2. Алгоритм адаптивного сглаживания	149
4.4.3. Результаты экспериментальных исследований	152
4.5. Каскадная процедура предварительной обработки сигналов	160

ГЛАВА 5

ОЦЕНКА ЭФФЕКТИВНОСТИ ДИАГНОСТИЧЕСКИХ ТЕСТОВ С ПОЗИЦИИ ТЕОРИИ СТАТИСТИЧЕСКИХ РЕШЕНИЙ

СТАТИСТИЧЕСКИХ РЕШЕНИЙ	165
5.1. Критерии эффективности диагностических тестов	165
5.1.1. Постановка задачи	165
5.1.2. Достаточные условия полезности диагностического теста	167
5.1.3. Условия полезности совокупности диагностических признаков ..	170
5.1.4. Модельный пример	174
5.1.5. Частный случай	176
5.2. Оценка полезности диагностических тестов в задачах скрининга за- болеваний	178
5.2.1. Общая схема	178
5.2.2. Допустимые границы специфичности и чувствительности полез- ного теста	182
5.2.3. Взаимосвязь полученных результатов с традиционным <i>ROC</i> -ана- лизом	185
5.3. Модель формирования коллективного решения в условиях проти- воречий	188
5.3.1. Постановка задачи	188
5.3.2. Байесова схема разрешения противоречий	188
5.3.3. Обучаемая система принятия коллективного решения	197
5.4. Оценка полезности признаков при диагональной матрице потерь	202
5.4.1. Общее условие бесполезности релевантных признаков	202
5.4.2. Признак-симптом, принимающий два значения	204
5.4.3. Непрерывный признак, нормально распределенный в классах	207

ГЛАВА 6

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ОЦЕНКИ СОСТОЯНИЯ ОБЪЕКТОВ

6.1. Информационная технология «ТЕРМОГРАФ»	213
6.1.1. Общая характеристика информационной технологии «ТЕРМОГРАФ»	213
6.1.2. Схема формирования информационного продукта	214
6.1.3. Аппаратная реализация упрощенных процедур обработки термограмм	224
6.1.4. Модифицированная процедура параметрической идентификации моделей	228
6.1.5. Результаты промышленных испытаний	230
6.2. Информационная технология «FAZAWIN»	231
6.2.1. Общая характеристика информационной технологии «FAZAWIN»	231
6.2.2. Диагностические признаки ЭКГ в фазовом пространстве	233
6.2.3. Результаты клинических исследований	240
6.3. Информационная технология «ФАЗАГРАФ»	246

6.3.1. Задача информационной технологии	246
6.3.2. Базовые гипотезы, положенные в основу ИТ «ФАЗАГРАФ»	247
6.3.3. Общая характеристика информационной технологии	248
6.3.4. Схема формирования информационного продукта	251
6.3.5. Оценка эффективности ИТ «ФАЗАГРАФ» для решения задачи скрининга ишемической болезни сердца	253
6.3.6. Интерфейс пользователя	256
6.3.7. Практические результаты	259
6.4. Информационная технология «ИКАР»	260
6.4.1. Общая характеристика информационной технологии «ИКАР»	260
6.4.2. Интерфейс пользователя	264
6.4.3. Автономное микропроцессорное устройство «ИКАР-2006»	269
6.4.4. Результаты физиологических экспериментов и клинических испытаний	272
6.5. Информационная технология «MAGWIN»	276
6.5.1. Общая характеристика ИТ «MAGWIN»	276
6.5.2. Алгоритм интерпретации, основанный на классификации карт плотностей тока	281
6.5.3. Результаты клинических испытаний	286
6.6. Перспективные информационные технологии	287
6.6.1. Идентификация личности на основе уникальных особенностей фазового портрета ЭКГ	287
6.6.2. Анализ и интерпретация кривых биохемилюминесценции (БХЛ)	295
6.6.3. Интегральная оценка эластичности кровеносных сосудов	299
СПИСОК ЛИТЕРАТУРЫ	308

Наукове видання
НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МІЖНАРОДНИЙ НАУКОВО-НАВЧАЛЬНИЙ
ЦЕНТР ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
ТА СИСТЕМ

ФАЙНЗІЛЬБЕРГ Леонід Соломонович

**ІНФОРМАЦІЙНІ
ТЕХНОЛОГІЇ
ОБРОБКИ
СИГНАЛІВ
СКЛАДНОЇ ФОРМИ**

ТЕОРІЯ І ПРАКТИКА

(Російською мовою)

Київ, Науково-виробниче підприємство
«Видавництво “Наукова думка” НАН України», 2008

Оформлення художника *І.П. Савицької*
Художній редактор *І.Р. Сільман*
Технічний редактор *С.Г. Максимова*
Коректор *О.В. Должикова*
Оператор *В.Г. Каменькович*
Комп’ютерна верстка *Т.О. Ценцеус*

Підп. до друку 10.12.2008. Формат 60×90/16.
Папір офс. № 1. Гарн. Таймс. Друк. офс.
Ум. друк. арк. 21,0. Ум. фарбо-відб. 21,0.
Обл.-вид. арк. 18,67. Тираж 500 прим. Зам. 8—893

НВП «Видавництво “Наукова думка” НАН України»
Свідоцтво про внесення суб’єкта видавничої справи
до Державного реєстру —
серія ДК № 2440 від 15.03.2006
01601 Київ 1, вул. Терещенківська, 3

03680, Київ 680, вул. Шутова, 13^б
ПП «Видавництво “Фенікс”»

