

Л. С. ФАЙНЗИЛЬБЕРГ

МЕТОДИ
ТА СИСТЕМИ
ШТУЧНОГО
ИНТЕЛЕКТУ



2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Л. С. Файнзільберг

МЕТОДИ ТА СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ

*Затверджено Вченою радою КПІ ім. Ігоря Сікорського
як підручник для студентів, які навчаються за спеціальністю
«Комп'ютерні науки» освітньо-професійної програми
«Комп'ютерні технології в біології та медицині»*

Київ-2023

УДК 004.8

Рецензенти: *Романов В. О.*, професор, доктор технічних наук, зав. відділу Інституту кібернетики ім. В. М. Глушкова НАН України

Степашко В. С., професор, доктор технічних наук, зав. відділу Міжнародного науково-навчального центру інформаційних технологій та систем НАН України та МОН України

Відповідальний редактор *Городецька О. К.*, канд. технічних наук, доцент кафедри біомедичної кібернетики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

**Гриф надано Вченою радою КПІ ім. Ігоря Сікорського
(протокол № 7 від 3 липня 2023 р.)**

Файнзільберг Леонід Соломонович, д-р техн. наук, професор

Методи та системи штучного інтелекту: підруч. для студ. спец. «Комп'ютерні науки» освітньо-професійної програми «Комп'ютерні технології в біології та медицині» / Л. С. Файнзільберг; КПІ ім. Ігоря Сікорського. – Київ, ТОВ «7БЦ», 2023. – 316 с.

Охоплено базові поняття та сучасні методи штучного інтелекту, орієнтовані на розв'язування задач цифрової медицини. Подано стохастичні моделі породження біомедичних сигналів складної форми з локально-зосередженими діагностичними ознаками. Викладено адаптивні алгоритми приглушення адитивних завад в умовах неповної апріорної інформації та оригінальний метод відновлення корисного сигналу у фазовому просторі з використанням метрики Гаусдорфа. Розглянуто формальні умови оцінювання корисності бінарних класифікаторів у задачах скринінгу. Подано оригінальні методи оброблення циклічних сигналів на основі метрики Левенштейна та фазового портрету перестановної ентропії. Містить завдання до комп'ютерних практикумів.

Для студентів, аспірантів і викладачів закладів вищої освіти з напрямку біомедичної інженерії.

ISBN 978-617-549-255-0

© Л. С. Файнзільберг, 2023

Leonid Fainzilberg

METHODS AND SYSTEMS FOR ARTIFICIAL INTELLIGENCE

The basic concepts and modern methods of artificial intelligence, focused on solving the problems of digital medicine, are covered. Stochastic models of the generation of complex-shaped biomedical signals with locally concentrated diagnostic features are presented. Adaptive algorithms for suppressing additive interference in the conditions of incomplete a priori information and an original method of restoring the useful signal in the phase space using the Hausdorff metric are outlined. The formal conditions for evaluating the usefulness of binary classifiers in screening tasks are considered. Original methods of cyclic signal processing based on the Levenshtein metric and phase portrait of permuted entropy are presented. Contains tasks for computer workshops.

For students, postgraduates and teachers of higher educational institutions in the field of biomedical engineering.

Es werden die wichtigsten Konzepte und modernen Methoden der künstlichen Intelligenz behandelt, die auf die Lösung der Probleme der digitalen Medizin abzielen. Es werden stochastische Modelle zur Bildung komplexer biomedizinischer Signale mit lokal konzentrierten diagnostischen Merkmalen vorgestellt. Wir beschreiben adaptive Algorithmen zur Unterdrückung von additivem Rauschen bei unvollständigen A-priori-Informationen und eine originelle Methode zur Wiederherstellung des nützlichen Signals im Phasenraum mithilfe der Hausdorff-Metrik. Es werden die formalen Bedingungen zur Bewertung der Nützlichkeit binärer Klassifikatoren bei Screening-Problemen betrachtet. Es werden originelle Methoden zur Verarbeitung zyklischer Signale basierend auf der Levenshtein-Metrik und dem Phasenporträt der Permutationsentropie vorgestellt. Der Text enthält Aufgaben für Computer-Workshops.

Für Studierende, Doktoranden und Lehrende höherer Bildungseinrichtungen im Bereich Biomedizintechnik.

ЗМІСТ

ВСТУП	7
РОЗДІЛ 1	
ОСНОВНІ ОЗНАЧЕННЯ ТА ПОНЯТТЯ	
ШТУЧНОГО ІНТЕЛЕКТУ	17
1.1 Історія розвитку наукових досліджень	17
1.2. Інтелектуальні інформаційні технології та цифрова медицина.....	29
1.3. Інтелектуальні інформаційні технології оброблення сигналів складної форми	40
1.4. Інтерактивний синтез прикладних ІТ.....	44
Завдання до комп'ютерних практикумів.....	52
Запитання для самоконтролю.....	53
РОЗДІЛ 2	
МАТЕМАТИЧНІ МОДЕЛІ ПОРОДЖЕННЯ	
ШТУЧНИХ СИГНАЛІВ	54
2.1. Загальні характеристика електрокардіографії.....	54
2.2. Інтерполяційна модель породження ЕКГ	58
2.3. Генеративна модель породження ЕКГ	65
2.4. Генерація ЕКГ з альтернацією зубця T	76
2.5. Порівняльний аналіз штучних та реальних ЕКГ.....	84
Завдання до комп'ютерних практикумів.....	91
Запитання для самоконтролю.....	93
РОЗДІЛ 3	
ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ПРИГЛУШЕННЯ	
АДИТИВНИХ ЗАВАД	94
3.1. Загальні відомості з теорії сигналів	94
3.2. Адаптивна фільтрація частотних завад	98
3.3. Адаптивна фільтрація випадкових завад	119
3.4. Фільтрація завад іншого типу.....	129

Завдання до комп'ютерного практикуму.....	134
Запитання для самоконтролю.....	135
РОЗДІЛ 4	
ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ОБРОБЛЕННЯ ЦИКЛІЧНИХ СИГНАЛІВ	137
4.1. Оцінювання форми усередненого циклу	137
4.2. Оцінювання ступеня хаотичності сигналів.....	157
4.3 Лінгвістичний підхід до оброблення циклічних сигналів...	176
Завдання до комп'ютерних практикумів.....	189
Запитання для самоконтролю.....	193
РОЗДІЛ 5	
БІНАРНІ КЛАСИФІКАТОРИ В ЗАДАЧАХ МЕДИЧНОЇ ДІАГНОСТИКИ	194
5.1. Формальні стратегії прийняття рішень.....	194
5.2. Оцінювання ефективності бінарного класифікатора.....	202
5.3. Допустимі границі ймовірностей пропуску цілі та хибної тривоги.....	213
Запитання для самоконтролю.....	222
РОЗДІЛ 6	
МАШИННЕ НАВЧАННЯ	223
6.1. Змістовна та формальна постановки задачі.....	223
6.2. Лінійні класифікатори.....	225
6.3. Одношаровий персептрон.....	235
6.4. Алгоритми навчання Розенблата та Козинця.....	240
6.5. Навчання штучної нейронної мережі.....	247
6.6. Генетичний алгоритм.....	252
Запитання для самоконтролю.....	261

РОЗДІЛ 7

ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ПРЕВЕНТИВНОЇ МЕДИЦИНИ	262
7.1. Загальні характеристики превентивної медицини.....	262
7.2. Інтелектуальний електрокардіограф ФАЗАГРАФ®	270
7.3. Інтелектуальний ритмограф на смартфоні.....	277
7.4. Інтелектуальний артеріограф на смартфоні.....	280
7.5. Інтелектуальний тонометр для домашнього застосування...	283
7.6. Інтелектуальний стетоскоп на смартфоні.....	287
Завдання до комп'ютерного практикуму.....	289
Запитання для самоконтролю.....	290

РОЗДІЛ 8

ІНТЕЛЕКТУАЛЬНІ БІОМЕТРИЧНІ СИСТЕМИ	291
8.1. Загальні характеристики біометрії.....	291
8.2. Фазовий портрет ЕКГ як засіб біометрії.....	294
8.3. Ідентифікація людини за фазовим портретом одноканальної ЕКГ	297
8.4. Верифікація людини за фазовим портретом одноканальної ЕКГ	299
8.5. Перспективи подальшого розвитку.....	303
Запитання для самоконтролю.....	304
ПІСЛЯМОВА.....	305
РЕКОМЕНДОВАНА ЛІТЕРАТУРА.....	306
ПРЕДМЕТНИЙ ПОКАЖЧИК.....	307
ПРО АВТОРА.....	311

ВСТУП

Штучний інтелект – один із найбільш важливих напрямів науки та технологій, який динамічно розвивається. Сформувалися три наукові підходи, спрямовані на практичну реалізацію методів штучного інтелекту: алгоритмічний, засновником якого є британський математик Алан Тюрінг, нейрокомп'ютерний, розвиток якого розпочався з моделі нервової клітини американських фізіологів Уоррена Мак-Калок та Уолтера Пітса, та еволюційний, запропонований італійським математиком Нільсом Барічеллі. Штучний інтелект проникає у всі сфери життя та стає невід'ємною частиною сучасного суспільства.

Завдяки впровадженню інтелектуальних інформаційних технологій (ІТ) та сучасних засобів телекомунікацій сформувалась нова парадигма охорони здоров'я – цифрова медицина, основна мета якої наблизити медичні засоби безпосередньо до пацієнта. Цифрова медицина реалізує головний принцип персоніфікованої медицини – лікувати не хворобу, а хворого, ґрунтуючись на індивідуальних особливостях його організму.

За допомогою штучного інтелекту лікарі можуть проводити більш точну діагностику, заздалегідь передбачати можливі захворювання та розробляти індивідуальні лікувальні плани для пацієнтів. Штучний інтелект надає змогу прискорити процес розроблення нових ліків та методів лікування, що може значно покращити якість життя людей.

Сучасна медична діагностика ґрунтується на комп'ютерних методах оброблення сигналів, що породжує організм. Такі сигнали мають складну форму, спостерігаються в умовах внутрішніх та зовнішніх перешкод, що ускладнює видобування діагностичної інформації зі спотворених фрагментів сигналу.

Саме тому основну увагу підручника для студентів зі спеціалізації «Інформаційні технології в біології та медицині» головним чином спрямовано на наукові та практичні задачі побудови ІТ оброблення біомедичної інформації.

Досвід викладання дисциплін «Основи штучного інтелекту» та «Методи та системи штучного інтелекту» студентам з біомедичної

інженерії доводить, що наявні підручники і навчальні посібники не повною мірою враховують специфіку студентів з біомедичного напрямку. В переважній більшості в них описано лише окремі напрями з вищеназваних наукових підходів, зокрема штучні нейронні мережі. Відсутність інформації про інші методи створює у студента помилкове уявлення, що саме наведений підхід є найкращим.

Мета підручника – заповнити цю прогалину та на єдиних методологічних позиціях викласти основні засади методів та систем штучного інтелекту для сучасних ІТ оброблення біомедичної інформації.

Розглянуто специфічні питання для студентів і спеціалістів з біомедичної інженерії, зокрема, методи оцінювання бінарних класифікаторів у задачах скринінгу, адаптивних алгоритмів приглушення завад різного типу, відновлення корисного сигналу за сукупністю фазових траєкторій та інші сучасні методи. Подано багато прикладів до практичних задач, які виникають перед конструктором сучасних ІТ оброблення біомедичної інформації.

Підручник складається з восьми розділів.

У першому розділі розглянуто основні поняття штучного інтелекту, викладено інформацію про обмеження та парадокси, які траплялися науковцям під час розвитку методів штучного інтелекту, сформульовані означення інформаційної технології оброблення сигналів складної форми, подано концептуальну ідею інтерактивного синтезу прикладних ІТ.

У другому розділі розглянуто оригінальні математичні моделі породження штучних сигналів складної форми. Викладено інтерполяційну та генеративну моделі породження циклічного сигналу на прикладі ЕКГ реалістичної форми. Розглянуто модель генерації ЕКГ з електричною альтернацією серця. Надано результати якісного та кількісного порівняння штучних та реальних сигналів.

У третьому розділі розглянуто ефективні методи приглушення адитивних завад. Викладено інформацію про характерні спотворення, які трапляються під час оброблення біомедичних сигналів. Розглянуто оригінальні адаптивні алгоритми приглушення частотних та випадкових завад в умовах обмеженої апріорної інформації. Подано інформацію про методи приглушення імпульсних завад та видалення дрейфу ізоелектричної лінії.

У четвертому розділі розглянуто інтелектуальні методи оброблення циклічних сигналів. Подано оригінальні методи відновлення

корисного сигналу (на прикладі еталонного циклу ЕКГ) за сукупністю спотворених фазових траєкторій з використанням метрики Гаусдорфа. На основі доказового експерименту доведено переваги запропонованого підходу порівняно з існуючими. Розглянуто метод класифікації надійних та ненадійних циклів. Викладено нові методи оброблення циклічних сигналів, які ґрунтуються на побудові фазового портрету перестановної ентропії та лінгвістичного методу інтерпретації сигналу на основі метрики Левенштейна.

У п'ятому розділі розглянуто формальні умови синтезу бінарних класифікаторів у задачах медичної діагностики. Викладено математичні умови ефективності бінарного класифікатора, які покладено в основу вдосконаленого ROC-аналізу. Визначено граничні значення ймовірностей пропуску цілі та хибної тривоги бінарних класифікаторів, що реалізують стратегії Неймана-Пірсона.

У шостому розділі розглянуто задачу машинного навчання розпізнаванню двох класів за прецедентами. Викладено змістовну та формальну постановку задачі. Розглянуто умови лінійної роздільності двох класів у багатовимірному просторі ознак. Наведено структуру та математичну модель одношарового перцептрона. Розглянуто алгоритми навчання Розенблата та Козинця. Подано результати статистичного експерименту для визначення швидкодії алгоритмів, принципи навчання класифікації штучної нейронної мережі та генетичного алгоритму.

У сьомому розділі розглянуто загальний підхід до створення інтелектуальних систем превентивної медицини. Наведено приклади практичного застосування підходу під час створення інтелектуального електрокардіографа ФАЗАГРАФ[®], інтелектуальних приладів АІ-РИТМОГРАФ та АІ-АРТЕРІОГРАФ, інтелектуального тонометра для домашнього застосування та інтелектуального стетоскопа на смартфоні.

У восьмому розділі розглянуто загальні характеристики інтелектуальних біометричних систем. Викладено принципи систем ідентифікації та верифікації людини за фазовим портретом ЕКГ. Подано результати експериментальних досліджень за реальними даними. Розглянуто перспективи подальшого розвитку та практичного застосування систем.

INTRODUCTION

Artificial intelligence is one of the most important areas of science and technology, which is developing dynamically. Three scientific approaches aimed at the practical implementation of artificial intelligence methods have been formed: algorithmic, the founder of which is the British mathematician Alan Turing, neurocomputer, the development of which began with the model of a nerve cell by American physiologists Warren McCulloch and Walter Pitts, and evolutionary, proposed by the Italian mathematician Niels Barichelli. Artificial intelligence penetrates into all spheres of life and becomes an integral part of modern society.

Thanks to the implementation of intelligent information technologies (IT) and modern means of telecommunications, a new paradigm of health care was formed - digital medicine, the main goal of which is to bring medical devices directly to the patient. Digital medicine implements the main principle of personalized medicine - to treat not the disease, but the patient, based on the individual characteristics of his organism.

With the help of artificial intelligence, doctors can make more accurate diagnoses, predict possible diseases in advance and develop individual treatment plans for patients. Artificial intelligence makes it possible to accelerate the process of developing new drugs and treatment methods, which can significantly improve people's quality of life.

Modern medical diagnostics is based on computer methods of processing signals generated by the body. Such signals have a complex shape, are observed in conditions of internal and external interference, which makes it difficult to extract diagnostic information from distorted signal fragments.

That is why the main focus of the textbook for students specializing in "Information technologies in biology and medicine" is mainly aimed at the scientific and practical tasks of building IT processing of biomedical information.

The experience of teaching the disciplines "Fundamentals of artificial intelligence" and "Methods and systems for artificial intelligence" to students of biomedical engineering proves that the available textbooks and training aids do not fully take into account the specifics of biomedical stu-

dents. In most of them, only certain directions from the above-mentioned scientific approaches are described, in particular, artificial neural networks. The lack of information about other methods creates a false impression in the student that this particular approach is the best.

The purpose of the textbook is to fill this gap and to outline the basic principles of methods and systems of artificial intelligence for modern IT processing of biomedical information in unified methodological positions.

Specific questions for students and specialists in biomedical engineering are considered, in particular, methods of evaluating binary classifiers in screening tasks, adaptive algorithms for suppressing various types of interference, restoration of a useful signal based on a set of phase trajectories, and other modern methods. Many examples of practical problems that arise before the designer of modern IT processing of biomedical information are given.

The textbook consists of eight chapters.

In the first part, the main concepts of artificial intelligence are considered, information about the limitations and paradoxes encountered by scientists during the development of artificial intelligence methods is presented, the definitions of information technology for complex signal processing are formulated, and the conceptual idea of interactive synthesis of applied IT is presented.

In the second chapter, the original mathematical models of the generation of artificial signals of a complex shape are considered. The interpolation and generative models of generation of a cyclic signal are presented on the example of an ECG of a realistic form. The model of ECG generation with electrical alternation of the heart is considered. The results of a qualitative and quantitative comparison of artificial and real signals are provided.

In the third section, effective methods of suppressing additive interference are considered. Information about characteristic distortions encountered during biomedical signal processing is presented. The original adaptive algorithms for suppressing frequency and random disturbances in the conditions of limited a priori information are considered. Information is given on methods of suppressing impulse interference and removing the drift of the isoelectric line.

In the fourth chapter, intelligent methods of cyclic signal processing are considered. Original methods of restoring the useful signal (on the example of the reference ECG cycle) based on a set of distorted phase trajec-

tories using the Hausdorff metric are presented. On the basis of a proof-of-concept experiment, the advantages of the proposed approach in comparison with the existing ones have been proven. The method of classification of reliable and unreliable cycles is considered. New methods of cyclic signal processing are outlined, which are based on the construction of a phase portrait of permutational entropy and a linguistic method of signal interpretation based on the Levenshtein metric.

In the fifth chapter, the formal conditions for the synthesis of binary classifiers in the tasks of medical diagnostics are considered. Mathematical conditions for the efficiency of the binary classifier are outlined, which are the basis of the improved ROC analysis. Limit values of target miss and false alarm probabilities of binary classifiers implementing Neyman-Pearson strategies are determined.

In the sixth chapter, two-class recognition machine learning problem based on precedents is considered. The substantive and formal positions of the problem are outlined. The conditions of linear separation of two classes in the multidimensional feature space are considered. The structure and mathematical model of a single-layer perceptron are given. The learning algorithms of Rosenblatt and Kozinets are considered. The results of a statistical experiment to determine the performance of algorithms are presented. The principles of learning the classification of an artificial neural network and a genetic algorithm are presented.

In the seventh chapter, a general approach to the creation of intelligent systems of preventive medicine is considered. Examples of the practical application of the approach during the creation of an intelligent electrocardiograph FAZAGRAPH®, intelligent devices AI-RHYTHMOGRAPH and AI-ARTIERIOGRAPH, an intelligent tonometer for home use and an intelligent stethoscope on a smartphone are given.

The eighth chapter discusses the general characteristics of intelligent biometric systems. The principles of human identification and verification systems based on the phase portrait of the ECG are outlined. The results of experimental studies based on real data are presented. Prospects for further development and practical application of the systems are considered.

EINFÜHRUNG

Künstliche Intelligenz ist einer der wichtigsten Bereiche von Wissenschaft und Technik, der sich dynamisch entwickelt. Drei wissenschaftliche Ansätze haben sich herausgebildet, die auf die praktische Umsetzung von Methoden der künstlichen Intelligenz abzielen: Algorithmik, deren Begründer der britische Mathematiker Alan Turing ist, Neurocomputer, deren Entwicklung mit dem Modell einer Nervenzelle der amerikanischen Physiologen Warren McCulloch und Walter Pitts begann und evolutionär, vorgeschlagen von dem italienischen Mathematiker Niels Barichelli. Künstliche Intelligenz dringt in alle Lebensbereiche vor und wird zum festen Bestandteil der modernen Gesellschaft.

Dank der Implementierung intelligenter Informationstechnologien (IT) und moderner Telekommunikationsmittel wurde ein neues Paradigma der Gesundheitsversorgung geschaffen – die digitale Medizin, deren Hauptziel es ist, medizinische Geräte direkt zum Patienten zu bringen. Die digitale Medizin setzt das Hauptprinzip der personalisierten Medizin um – nicht die Krankheit, sondern den Patienten zu behandeln, basierend auf den individuellen Eigenschaften seines Organismus.

Mit Hilfe künstlicher Intelligenz können Ärzte genauere Diagnosen stellen, mögliche Krankheiten im Voraus vorhersagen und individuelle Behandlungspläne für Patienten entwickeln. Künstliche Intelligenz ermöglicht es, den Prozess der Entwicklung neuer Medikamente und Behandlungsmethoden zu beschleunigen, was die Lebensqualität der Menschen deutlich verbessern kann.

Die moderne medizinische Diagnostik basiert auf computergestützten Verfahren zur Verarbeitung körpereigener Signale. Solche Signale haben eine komplexe Form, werden unter Bedingungen interner und externer Interferenz beobachtet, was es schwierig macht, diagnostische Informationen aus verzerrten Signalfragmenten zu extrahieren.

Der Schwerpunkt des Lehrbuchs für Studierende der Vertiefungsrichtung „Informationstechnologien in Biologie und Medizin“ richtet sich daher vor allem auf die wissenschaftlichen und praktischen Aufgaben der bauinformatischen Verarbeitung biomedizinischer Informationen.

Die Erfahrung aus der Vermittlung der Fachrichtungen „Grundlagen der Künstlichen Intelligenz“ und „Methoden und Systeme der Künstlichen Intelligenz“ an Studierende der Biomedizinischen Technik beweist, dass

die verfügbaren Lehrbücher und Ausbildungshilfen die Besonderheiten der Biomedizinstudierenden nicht vollständig berücksichtigen. In den meisten Situationen werden nur bestimmte Richtungen aus den oben genannten wissenschaftlichen Ansätzen beschrieben, insbesondere künstliche neuronale Netze. Der Mangel an Informationen über andere Methoden erweckt beim Schüler den falschen Eindruck, dass dieser spezielle Ansatz der beste ist.

Ziel des Lehrbuchs ist es, diese Lücke zu schließen und die Grundlagen von Methoden und Systemen der Künstlichen Intelligenz für die moderne IT-Verarbeitung biomedizinischer Informationen in einheitlichen methodischen Positionen zu skizzieren.

Spezielle Fragestellungen für Studierende und Fachkräfte der Biomedizinischen Technik werden berücksichtigt, insbesondere Methoden zur Auswertung binärer Klassifikatoren bei Screening-Aufgaben, adaptive Algorithmen zur Unterdrückung verschiedener Störarten, Wiederherstellung eines Nutzsignals auf Basis eines Satzes von Phasentrajektorien, und andere moderne Methoden. Es werden viele Beispiele für praktische Probleme gegeben, die vor dem Designer der modernen IT-Verarbeitung biomedizinischer Informationen auftreten.

Das Lehrbuch besteht aus acht Kapiteln.

Im ersten Teil werden die Hauptkonzepte der künstlichen Intelligenz betrachtet, Informationen über die Grenzen und Paradoxien vorgestellt, auf die Wissenschaftler bei der Entwicklung von Methoden der künstlichen Intelligenz gestoßen sind, die Definitionen der Informationstechnologie für komplexe Signalverarbeitung formuliert, und die konzeptionelle Idee der interaktiven Synthese angewandter IT wird vorgestellt.

Im zweiten Kapitel werden die ursprünglichen mathematischen Modelle der Erzeugung künstlicher Signale mit komplexer Form betrachtet. Am Beispiel eines EKG in realistischer Form werden die Interpolations- und generativen Modelle der Generierung eines zyklischen Signals vorgestellt. Es wird das Modell der EKG-Erzeugung mit elektrischem Wechsel des Herzens betrachtet. Die Ergebnisse eines qualitativen und quantitativen Vergleichs von künstlichen und realen Signalen werden bereitgestellt.

Im dritten Abschnitt werden effektive Methoden zur Unterdrückung additiver Störungen betrachtet. Informationen über charakteristische Verzerrungen, die während der biomedizinischen Signalverarbeitung auftreten, werden präsentiert. Die ursprünglichen adaptiven Algorithmen zur Unterdrückung von Frequenz- und Zufallsstörungen unter Bedingungen

begrenzter a priori-Informationen werden betrachtet. Es werden Informationen zur Unterdrückung von Impulsstörungen und zur Beseitigung der Drift der isoelektrischen Linie gegeben.

Im vierten Kapitel werden intelligente Methoden der zyklischen Signalverarbeitung betrachtet. Es werden originelle Methoden zur Wiederherstellung des Nutzsignals (am Beispiel des Referenz-EKG-Zyklus) basierend auf einem Satz verzerrter Phasentrajektorien unter Verwendung der Hausdorff-Metrik vorgestellt. Anhand eines Proof-of-Concept-Experiments wurden die Vorteile des vorgeschlagenen Ansatzes gegenüber den bestehenden nachgewiesen. Die Methode der Klassifizierung von zuverlässigen und unzuverlässigen Zyklen wird betrachtet. Neue Methoden der zyklischen Signalverarbeitung werden skizziert, die auf der Konstruktion eines Phasenporträts der Permutationsentropie und einer linguistischen Methode der Signalinterpretation auf Basis der Levenshtein-Metrik basieren.

Im fünften Kapitel werden die formalen Bedingungen für die Synthese binärer Klassifikatoren in den Aufgaben der medizinischen Diagnostik betrachtet. Mathematische Bedingungen für die Effizienz des binären Klassifikators werden skizziert, die die Grundlage der verbesserten ROC-Analyse sind. Grenzwerte von Zielverfehlungs- und Fehlalarmwahrscheinlichkeiten von binären Klassifizierern, die Neyman-Pearson-Strategien implementieren, werden bestimmt.

Im sechsten Kapitel wird das Problem des maschinellen Lernens zweier Klassen auf der Grundlage von Präzedenzfällen betrachtet. Die inhaltlichen und formalen Positionen des Problems werden umrissen. Es werden die Bedingungen der linearen Trennung zweier Klassen im mehrdimensionalen Merkmalsraum betrachtet. Die Struktur und das mathematische Modell eines einschichtigen Perzeptrons werden angegeben. Dabei werden die Lernalgorithmen von Rosenblatt und Kozinets berücksichtigt. Vorgestellt werden die Ergebnisse eines statistischen Experiments zur Bestimmung der Leistungsfähigkeit von Algorithmen. Die Prinzipien des Erlernens der Klassifikation eines künstlichen neuronalen Netzes und eines genetischen Algorithmus werden vorgestellt.

Im siebten Kapitel wird ein allgemeiner Ansatz zur Schaffung intelligenter Systeme der Präventivmedizin betrachtet. Beispiele für die praktische Anwendung des Ansatzes bei der Entwicklung eines intelligenten Elektrokardiographen FASEGRAPH[®], intelligenter Geräte AIRHYTHMOGRAPH und AI-ARTIERIOGRAPH, eines intelligenten To-

nometers für den Heimgebrauch und eines intelligenten Stethoskops auf einem Smartphone werden gegeben.

Das achte Kapitel behandelt die allgemeinen Eigenschaften intelligenter biometrischer Systeme. Die Prinzipien menschlicher Identifikations- und Verifikationssysteme auf der Grundlage des Phasenporträts des EKG werden skizziert. Die Ergebnisse experimenteller Studien auf der Grundlage realer Daten werden vorgestellt. Perspektiven für die Weiterentwicklung und praktische Anwendung der Systeme werden betrachtet.

РОЗДІЛ 1

ОСНОВНІ ОЗНАЧЕННЯ ТА ПОНЯТТЯ ШТУЧНОГО ІНТЕЛЕКТУ

1.1. Історія розвитку наукових досліджень

Штучним інтелектом (Artificial Intelligence) називають науку, спрямовану на створення технічних та/або програмних систем, що мають властивості природного інтелекту. Щоб сформулювати властивості таких систем, будемо використовувати таке означення.

Означення 1.1. Інтелект це здатність пристосовуватися до нових невідомих ситуацій та навчатися, щоб розуміти й активно впливати на навколишнє середовище.

Надалі властивості, притаманні інтелектуальним системам, будуть розширюватися та уточнюватися.

Формування наукових досліджень у галузі штучного інтелекту (ШІ) розпочалося ще у першій половині ХХ століття, які стали основою трьох напрямів (рис. 1.1).



Рис. 1.1. Напрями наукових досліджень

Засновником алгоритмічного підходу в галузі ШІ вважають видатного британського математика Алана Тюрінга (Alan Mathison Turing, рис. 1.2). Метою його досліджень було вивчення зовнішнього боку інтелекту людини, зокрема аналізування того, як людина проводить обчислення. Його роздуми в 1936 році призвели до створення моделі абстрактної обчислювальної машини, яку пізніше назвали машиною Тюрінга.



Рис. 1.2. Алан Тюрінг

Машина Тюрінга формалізує поняття «обчислювальний алгоритм», який зводиться до послідовних переходів станів машини S_1, S_2, \dots, S_N за таким правилом. Залежно від поточного стану S_i , $1 \leq i \leq N$, що визначається символом з кінцевого алфавіту, записаним у i -ту комірку нескінченної стрічки, за правилом потрібно записати в цю комірку новий символ, перейти в новий стан S_{i+1} та переміститися на одну комірку вліво або вправо. Деякі стани зазначені як термінальні, і перехід у будь-який із них означає зупинку алгоритму.

Незважаючи на простоту, машина Тюрінга не тільки стандартизувала саме поняття алгоритму, а й започаткувала математичну теорію складності, алгоритмічну теорію інформації, криптографію та багато інших наукових дисциплін.

Першовідкривачами другого наукового підходу в галузі ШІ – нейрокомп'ютерного підходу, були американські фізіологи (рис. 1.3) Уоррен Мак-Калок (Warren Sturgis McCulloch) та Уолтер Пітс (Walter Pitts), які вивчали внутрішню організацію людського мозку.



Рис. 1.3. Уоррен Мак-Каллок (ліворуч) та Уолтер Пітс (праворуч)

З дослідження клітин мозку вчені вперше встановили, що процес мислення організовує складну комбінацію елементарних логічних

операцій типу I, АБО, НЕ. У результаті 1943 року було запропоновано найпростішу математична модель нейрона.



Рис. 1.4. Нільс Борічеллі

I, нарешті, формування третього наукового напрямку в галузі ШІ (еволюційного підходу) пов'язують з ім'ям італійського математика Нільса Барічеллі (Nils Bärnstrand, рис. 1.4) який вивчав процес еволюції інтелекту. В 1954 році було запропоновано модель, яка послужила основою низки алгоритмів, зокрема генетичного алгоритму. Цей евристичний алгоритм активно використовує такі механізми, як схрещування, мутації та спадковість, запозичені з теорії Дарвіна еволюції живої природи.

Протягом багатьох років зазначені три напрями робіт у галузі ШІ розвивалися паралельно і досить часто необґрунтовано протиставлялися один одному. Проте складності практичних задач, які потребують застосування засобів штучного інтелекту, призвело до певного зближення цих підходів. У наш час активно формуються наукові програми щодо створення НБІК-технологій, які ґрунтуються на конвергенції нанотехнологій, біології, інформаційних і когнітивних науках.

XX століття сміливо можна назвати «золотим» століттям розвитку наукових досліджень у галузі ШІ. Згадаємо тільки декілька прізвищ видатних учених, які запропонували та досліджували важливі наукові проблеми теорії штучного інтелекту.

З ім'ям американського математика Джона фон Неймана (John von Neumann), рис. 1.5, ліворуч) пов'язують архітектуру більшості сучасних комп'ютерів – архітектуру фон Неймана. Він запропонував теорію цифрових автоматів, зокрема, концепцію клітинних автоматів.

Саме концепція клітинних автоматів надала математичне обґрунтування можливості створення життя з мертвої матерії. На основі цієї концепції фон Нейман запропонував структуру абстрактної машини, яка може відтворювати сама себе. Ідею клітинних автоматів почали застосовувати майже у всіх галузях сучасної науки.

Френк Розенблат (Frank Rosenblatt, рис. 1.5, по центру) вперше створив на технічному пристрої відповідної структури (персептроні)

обчислювальну систему, здатну навчатися і розв'язувати найпростіші задачі. Сама ця структура була основою сучасних нейронних мереж, які широко застосовують у різних галузях.



Рис. 1.5. Дж. фон Нейман (ліворуч), Френк Розенблат (по центру) та Клод Шенон (праворуч)

Американський математик Клод Шенон (Claude Elwood Shannon, рис. 1.5, праворуч) вважається основоположником інформаційного віку. Він запропонував теорію інформації та кодування, яка сформувала основу сучасних комунікаційних технологій.

Американський вчений-лінгвіст Ноам Хомський (Avram Noam Chomsky, рис. 1.6, ліворуч) є автором формальних мов, що називають ієрархією Хомського. Ідеї цього видатного вченого покладено не тільки в основу математичної лінгвістики, а й застосовуються в різних галузях комп'ютерних наук. Зокрема, ідею породжувальної граматики Хомського використовують в інтелектуальних методах структурного розпізнавання об'єктів різної фізичної природи.



Рис. 1.6. Ноам Хомський (ліворуч) та Лафті Заде (праворуч)

Професор Каліфорнійського університету Лафті Заде (Lotfi Askar Zadeh, рис. 1.6, праворуч) є засновником теорії нечітких множин та нечіткої логіки. На відміну від класичної логіки, яка використовує

поняття «ІСТИНА» та «ХИБНІСТЬ», що кодують одиницею та нулем, у нечіткій логіці (fuzzy logic) застосовують лінгвістичну змінну, що характеризує функції належності до нечіткої множини та набуває значення між нулем та одиницею.

Незважаючи на те, що спочатку таке узагальнення класичної логіки не сприймалось традиційними математиками, теорія Заде виявилась дуже плідною та конструктивною для розв'язування важливих задач у різних галузях.

Значний вклад в розвиток математичних методів штучного інтелекту зробили виданні математики Андрій Миколайович Колмогоров (рис. 1.7, ліворуч) та Олександр Миколайович Тихонов (рис. 1.7, праворуч). Їм належать фундаментальні дослідження методів розв'язування слабоформалізованих та некоректно поставлених математичних задач.



Рис. 1.7. А. М. Колмогоров (ліворуч) та О. М. Тихонов (праворуч)

Відомий український учений, академік Віктор Михайлович Глушков (рис. 1.8, ліворуч) – засновник кібернетичного центру національної академії наук України, автор фундаментальних праць у галузі кібернетики, математики і обчислювальної техніки. Він суттєво розвинув теорію цифрових автоматів, брав безпосередню участь у створенні першого в континентальній Європі цифрового комп'ютера. Під керівництвом В. М. Глушкова створено оригінальні обчислювальні машини, зокрема, машини сімейства МІР для інженерних розрахунків.



Рис. 1.8. В. М. Глушков (ліворуч), М. М. Амосов (по центру) та О. Г. Івахненко (праворуч)

Український кардіохірург, академік Микола Михайлович Амосов (рис. 1.8, по центру) також працював у кібернетичному центрі. Він відомий не тільки як автор новаторських методик у кардіології, а й як засновник нового напрямку – біологічної та медичної кібернетики. Опублікував фундаментальні наукові праці з моделювання мислення та психіки, разом зі своїми учнями створив ефективні кібернетичні моделі регуляції життєвих функцій та нейрокомп'ютерних технологій.

Академік Олексій Григорович Івахненко (рис. 1.8, праворуч) зробив значний внесок у теорію автоматичного керування технічними об'єктами, розробив оригінальні методи індуктивного моделювання та теорію самоорганізації моделей за експериментальними даними, поклавши в основу метод групового урахування аргументів (МГУА), який широко використовують для розв'язування практичних задач, зокрема, задач розпізнавання образів і прогнозування складних систем.

Здатність до абстрактного мислення, узагальнення, навчання розпізнаванню образів тривалий час вважалось нероздільною властивістю людського мозку. Цікаво, що людина може навіть за малою кількістю прикладів навчитися безпомилково знайти узагальнюючі ознаки, які надають змогу відрізнити зображення одного класу від другого. Це наочно ілюструють картинки М. М. Бонгарда, наведені на рис. 1.9.

Водночас важко відповісти на запитання, за якими ознаками людина може відрізнити досить схожі зображення, наприклад кішок та собак (рис. 1.10). Хоча навіть трирічна дитина робить це дуже вдало.

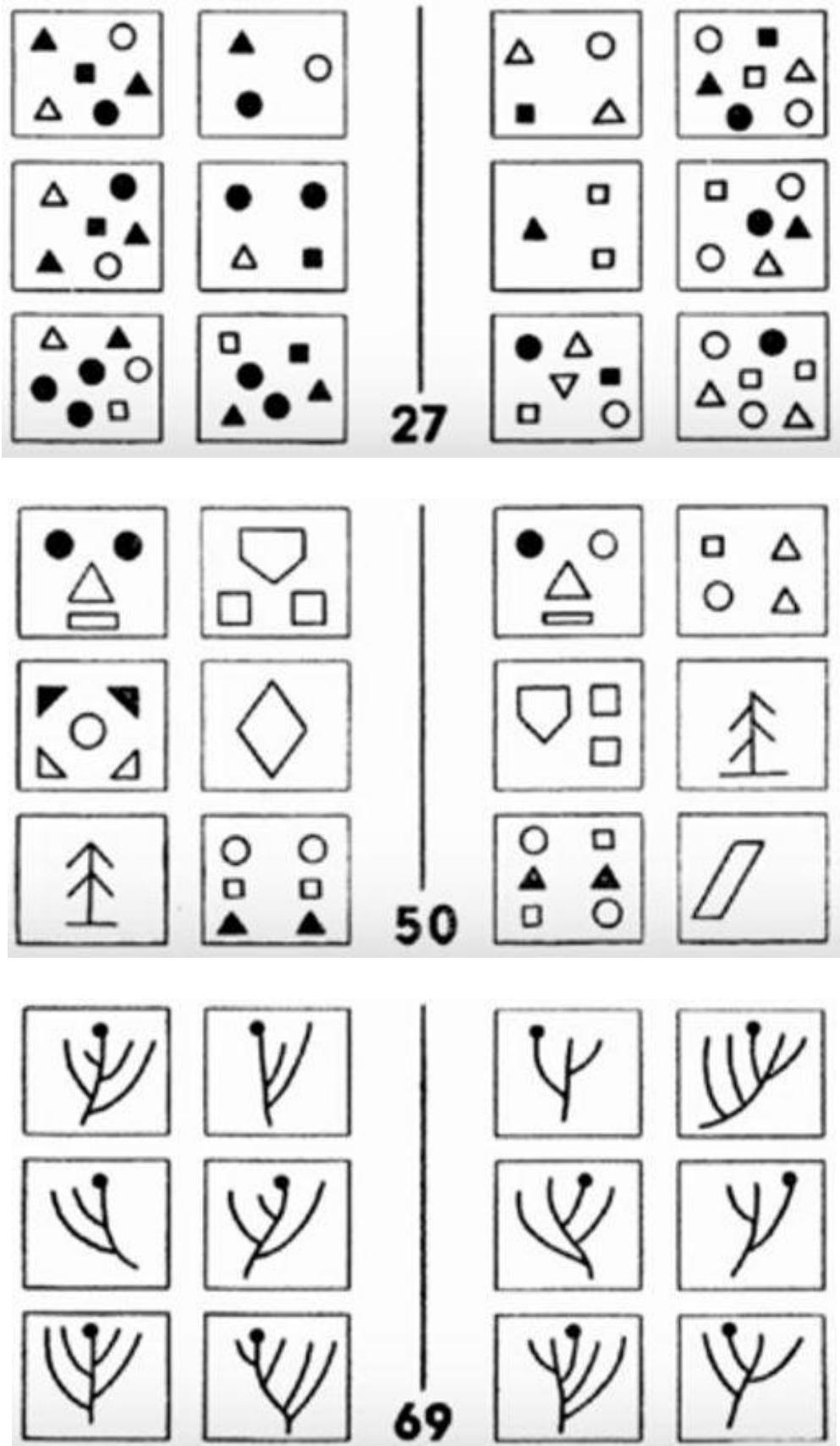


Рис. 1.9. Приклади картинок М. М. Бонгарда



Рис. 1.10. Зображення кішок (ліворуч) та собак (праворуч)

Інтелект людини – це не тільки розумний спосіб поведінки. В повній темряві і тиші, розмірковуючи і усвідомлюючи себе самого і навколишній світ людина залишається розумною.

Підтвердженням такого факту є сліпоглухоніма з дитинства О. І. Скороходова, яка за допомогою українського вченого лікаря-дефектолога І. О. Соколянського повернулася до майже повноцінного життя, захистила дисертацію та написала книгу «Як я сприймаю, уявляю і розумію навколишній світ».

Слід зазначити, що на етапі інтенсивного розвитку теорії штучного інтелекту вчені виявили низку обмежень та сформулювали на перший погляд парадоксальні принципи. Розглянемо деяких з них.

Так, наприклад, Ганс Моравек (Hans Moravec) сформулював принцип, за яким всупереч поширеній думці високорівневі міркуван-

ня вимагають невеликих обчислень, тоді як низькорівневі сенсомоторні навички вимагають величезних обчислювальних витрат. Парадокс Моравека (Moravec's paradox) полягає в тому, що в галузі штучного інтелекту відносно легко досягти рівня дорослої людини в таких задачах, як гра в шахи, проте дуже складно досягти навичок дитини надійно відрізнити будь-яку кішку від будь-якого собаки.

Проблема машинного навчання класифікації образів (Machine Learning) вже давно привертає увагу вчених. Поява перцептрона Розенблата відкрила епоху нейронних мереж, які були дуже популярними до появи в 1969 монографії Мартіна Мінського (Martin Minsky, рис. 1.11, ліворуч) і Саймона Пейперта (Seymour Papert, рис. 1.11, праворуч). Автори цієї монографії продемонстрували обмежені можливості перцептрону. Їм належить крилата фраза: «Нейронні мережі вирішують практичні задачі, але ми не розуміємо як вони це роблять».

Прихильники нейрокомп'ютерного підходу пропонували різні структури та методи навчання нейронних мереж. У 1980 роки з'явилися теоретичні передумови нового підходу до навчання багатосарових нейронних мереж, яке пізніше назвали глибоким навчанням (Deep Learning).

Незважаючи на те, що глибоке навчання досі інтенсивно розвивається, монографія Мінського та Пейперта на десятки років змінила інтерес дослідників та розробників штучного інтелекту від нейрокомп'ютерного спрямування до алгоритмічного, в якому були отримані серйозні теоретичні та прикладні результати, зокрема у задачі навчання розпізнавання образів.

Українські вчені, зокрема, співробітники кібернетичного центру НАН України, проводячи фундаментальні наукові дослідження, сформували нові наукові напрями у розпізнаванні зображень, мовленнєвих сигналів та інтелектуального оброблення текстової інформації.

Перед фахівцями в галузі штучного інтелекту виникає важливе запитання – яку технічну систему справді можна вважати інтелектуальною. Першим, хто спробував відповісти на це запитання, був Алан Тюрінг, який ще в 1950 році у філософському журналі *Mind* запропонував тест, названий згодом його ім'ям.



Рис. 1.11. Марвін Мінський (ліворуч) та Сеймур Пайперт (праворуч)

На думку Тюрінга, тест дає змогу оцінити рівень штучного інтелекту. Ідея перевірки передбачає діалог протягом 5 хвилин експертів за допомогою текстових повідомлень з живою людиною та роботом (комп'ютерною програмою), які перебувають в окремій кімнаті. Якщо щонайменше 30 % експертів визнають комп'ютер людиною, то тест вважається пройденим.

Однією з перших комп'ютерних програм, здатних вести такий діалог – програма Еліза (ELIZA), створена американським фахівцем Джозефом Вейценбаумом у 1966 році. Програма пародіює діалог психотерапевта, реалізуючи техніку активного слухання на основі нескладного синтаксичного аналізатора, який вміло перефразовує висловлювання користувача та створюється хибне уявлення про інтелектуальні властивості програми.

Наприклад, фразу «У мене болить голова» перефразовує у запитання «Чому Ви кажете, що у вас болить голова?». Не знаходячи варіантів відповіді, Еліза формує невизначене повідомлення такого типу: «У кожного свої проблеми!» та переключає розмову на іншу тему.

Програми, здатні проводити подібний діалог (комп'ютерні роботи-боти), постійно вдосконалюються та змагаються у проходженні тесту Тюрінга. З 1990 року навіть відбувається щорічний конкурс AI Loebner, переможцю якого вручається премія Лебнера в 2000 доларів. Першій програмі, яка повністю пройде текстовий тест Тюрінга, буде вручено срібну медаль і винагороду в 25000 \$, а програмі, яка пройде цей тест з використанням текстового, візуального та звукового підтвердження, буде вручено золоту медаль і приз у 100000 \$.

Інформацію про комп'ютерних роботів, здатних проводити «інтелектуальний» діалог з користувачами, буде не повною, якщо не згадати про останні досягнення в цій галузі. Одне з таких досягнень – робот Софія (Sophia) у вигляді жінки, розроблений компанією Hanson Robotics (рис. 1.12). Цікаво, що у 2017 році робот Софія вперше в історії людства отримав громадянство Саудівської Аравії!



Рис. 1.12. Людиноподібний робот Софія (Sophia)

Софія оснащена функціями оброблення візуальної інформації, технологією розпізнавання облич та мови, може імітувати близько 60 виразів обличчя та людських емоцій, а також проводити прості бесіди з певних тем (наприклад, про погоду). Програмне забезпечення, що розробляє компанія Singularity NET, постійно вдосконалюється.

Безумовно, такі розробки є гарною рекламою та демонструють високий рівень сучасних науково-технічних рішень у галузі «слабкого» штучного інтелекту, який, на відміну від «сильного», не претендує на такі фундаментальні поняття, як самосвідомість та цілепокладання.

Доречно буде згадати цитату американського вченого Джона Серля (John Searle): «Імітувати інтелект і бути інтелектом не те саме. Мишоловка ловить мишей, але не є кішкою». Він також сформулював оригінальний експеримент, названий «Китайська кімната» (Chinese Room Argument), який певною мірою спростував тест Тюрінга.

Ідея експерименту полягає в тому, що в ізольованій кімнаті знаходиться людина, якій передають осмислені повідомлення китайською мовою і вона відповідає на них також китайською мовою (рис. 1.13).

Сенс експерименту полягає в тому, що спостерігач, який знає китайські ієрогліфи, через щілину передає до кімнати ієрогліфи з питаннями, а на виході має отримати усвідомлену відповідь.



Рис. 1.13. Ідея експерименту «Китайська кімната»

Людина не знає жодного ієрогліфа, але у нього є записані в книзі точні інструкції типу «Візьміть такий ієрогліф із кошика номер один і помістіть його поруч із таким ієрогліфом із кошика номер два». Тобто інструкція складена таким чином, що ієрогліфи запитання мають перетворитися на ієрогліфи відповіді. В інструкціях відсутня інформація про значення цих ієрогліфів, людина лише дотримується вказаних інструкцій подібно до комп'ютера.

У такій ситуації спостерігач може відправити до кімнати будь-яке осмислене запитання й отримати на нього осмислену відповідь, як під час розмови з людиною, яка вільно володіє китайською писемністю.

Експеримент «Китайська кімната» значною мірою для багатьох дослідників виявився складнішим, ніж вони припускали, оскільки їхня мета була створити штучну особистість та реалізувати ідеї «сильного штучного інтелекту».

Аналіз доступних публікацій показує, що в наш час відбувається поступовий відхід від самого поняття «штучний інтелект» та перехід до прагматичної мети створення інтелектуальних інформаційних технологій та систем.

1.2. Інтелектуальні інформаційні технології та цифрова медицина

За останні роки і в науковій, і в науково-популярній літературі дедалі частіше згадуються терміни «ІТ – інформаційні технології» та «ІІТ – інтелектуальні інформаційні технології». І це не випадково, оскільки на сучасному етапі розвитку нашого суспільства такі технології відіграють важливу роль у вирішенні актуальних прикладних задач.

Водночас необґрунтоване застосування цих термінів до прикладних систем, які не є ні ІТ, ні ІІТ, може призвести до дискредитації цих важливих наукових напрямів, як це вже неодноразово було в минулому, наприклад, до дискредитації терміна «АСУ – автоматизована система управління».

Оскільки наука починається з означень, спробуємо сформулювати такі поняття. Зокрема надамо означення поняттю «ІТ оброблення сигналів складної форми», які становитимуть основний зміст наступного матеріалу.

Розглянемо базове визначення слова «технологія» (рис. 1.14), яке поєднує два терміни: «техно» (грец. «τέχνη» – мистецтво, майстерність, вміння) та «логія» (грец. «λογος» – наука).

У загальному випадку можна сформулювати таке означення:

Означення 1.2. Технологія – це мистецтво найбільш ефективним способом перетворити деяку сировину в продукт.

Наприклад, технологія виплавки чавуну передбачає сукупність операцій, які реалізують вміння найбільш ефективним способом перетворити залізну руду (сировину) в метал із заданими властивостями (продукт).

Сформулюємо тепер такі означення:

Означення 1.3. Інформаційна технологія – це сукупність методів та засобів, які найбільш ефективним способом перетворюють початкові дані (сировина технології) в інформацію, орієнтовану на конкретного користувача (продукт технології).



Рис. 1.14. Пояснення до терміну «ТЕХНОЛОГІЯ»

Означення 1.4. Інтелектуальна інформаційна технологія – це інформаційна технологія, яка для формування продукту використовує комп’ютерні процедури, що мають властивості природного інтелекту, а саме принаймні такі властивості:

- узагальнення – здатність класифікувати ситуації довкілля;
- навчання – здатність покращувати свої споживчі властивості під час експлуатації;
- адаптація – здатність пристосовуватися до змін довкілля;
- інваріантність – нечутливість до збурень довкілля;
- прогнозування – можливість відновлювати загальне бачення та передбачати майбутні ситуації зі спостереження лише частин (фрагментів) довкілля;
- розуміння – здатність осмислювати дійсність з урахуванням порівняння поточних характеристик довкілля зі своїми минулими значеннями;
- гнучкість – стійкість до можливих невдач і здатність корекції прийнятих рішень;
- взаємозамінність – використання альтернативних методів аналізу довкілля;
- комунікабельність – здатність надавати інформацію у формі, зрозумілій для сприйняття конкретним користувачем з урахуванням його кваліфікації.

Побудова інтелектуальних ІТ є одним із пріоритетних напрямів науково-технічного прогресу та інноваційного розвитку. На відміну від традиційних ІТ, оснований на процедурах оброблення числових даних, інтелектуальні ІТ використовують узагальнені поняття (образи), що дають більш повну інформацію про навколишнє середовище, а аналіз таких образів породжує цілісне бачення явищ, що вивчаються.

Інтелектуальні ІТ є головним інструментом створення засобів біомедичної кібернетики, спрямованих на підвищення ефективності діагностики, лікування та реабілітації.

Цей напрям сформувався ще на початку 60-х років минулого століття і ґрунтувався на специфічних математичних, інструментальних та інформаційних методах отримання знань про живі складноорганізовані системи.

На основі проведених досліджень було створено математичні моделі як окремих органів у нормі та патології, так і математичні моделі, орієнтовані на діагностику, прогнозування та лікування різних захворювань. Велику увагу було приділено також автоматизації організаційно-управлінських завдань у клініках та органах охорони здоров'я.

Зазначаючи безперечну важливість цих результатів, слід, однак, визнати, що на початковому етапі становлення біологічна та медична кібернетика не забезпечила ефективне вирішення низки проблем. Зокрема, не вдалося створити умови для масових профілактичних обстежень, які б забезпечили виявлення небезпечних захворювань на ранніх стадіях, знизити вартість медичних послуг, забезпечити їх доступність широким верствам населення, наблизивши засоби діагностики та лікування безпосередньо до пацієнтів.

Зрозуміло, що ці проблеми не могли бути подолані лише незначними доопрацюваннями та спрощенням приладів клінічного використання. Виявилося, що розробка простих та надійних персоналізованих медичних засобів потребує застосування нових підходів до оброблення та специфічних методів подання інформації користувачеві, який не має спеціальних медичних знань.

Використання інтелектуальних ІТ призвело до нової парадигми охорони здоров'я – цифрової медицини. Поняття «цифрова медицина» не слід ототожнювати з будь-яким медичним виробом, який побудований на елементах цифрової обчислювальної техніки. Напри-

клад, у медичній практиці широко використовують цифрові електрокардіографи третього покоління, які реалізують традиційні алгоритми діагностики, що ґрунтуються на багаторічному досвіді кардіологів візуально оцінювати електрокардіограму (ЕКГ).

Безумовно, застосування таких електрокардіографів полегшує працю медичного персоналу та скорочує час отримання результатів діагностики. Однак, цифрові електрокардіографи, що реалізують лише традиційні підходи до обробки ЕКГ, не належать до виробів цифрової медицини, оскільки вони не сприяли досягненню головної мети – підвищенню достовірності результатів діагностики.

Згідно з визначенням International Society of Digital Medicine, цифровою медициною називають галузь науки, в якій вчені прагнуть пояснити раніше незрозумілі патофізіологічні явища в організмі людини та дослідити нові медичні процедури з використанням сучасних цифрових технологій для покращення якості життя людини.

Впровадження цифрових технологій у медицину відкриває величезні перспективи, дозволяючи значно покращити якість медичного обслуговування та принципово змінити підхід до ранньої діагностики та лікування небезпечних захворювань.

Важливе місце посідають також технології реабілітації, що надає змогу відновити функції, втрачені внаслідок перенесених захворювань, зокрема уражень центральної нервової системи. В останні десятиліття спостерігається швидкий розвиток нових цифрових технологій для реабілітації пацієнтів, що страждають на неврологічні розлади. На відміну від традиційних, такі технології надають змогу коригувати інтенсивність і дозу реабілітаційної процедури залежно від індивідуальних особливостей організму конкретного пацієнта.

Оригінальні засоби цього напрямку – прилади серії ТРЕНАР розроблені в Міжнародному науково-навчальному центрі інформаційних технологій та систем НАН та МОН України (далі – МННЦ ІТiС). Прилади реалізують технологію побудови електронного образу руху, засновану на реєстрації сигналів зі здорового органу (наприклад, здорової руки) з подальшим формуванням електростимулюючих імпульсів на пошкоджену кінцівку (рис. 1.15).



Рис. 1.15. Прилад ТРЕНАР-01:
зовнішній вигляд (ліворуч) та технологія реабілітації (праворуч)

Важливе місце у програмах цифрової медицини приділено розробці спеціальних засобів для пацієнтів, яким з певних причин неможливо повністю відновити природні фізіологічні функції – зір, слух, рух тощо. Використання «інтелектуальних» протезів (рис. 1.16) надає змогу таким пацієнтам повернутися до нормального способу життя та трудової діяльності.



Рис. 1.16. Інтелектуальний протез руки

Особливе місце в цьому напрямі займають екзоскелетони, що є зовнішнім скелетом, який надає змогу людям з паралічем нижніх кінцівок ходити (рис. 1.17). Останні дослідження підтверджують позитивний ефект від використання екзоскелетонів пацієнтами з травмами спинного мозку та перенесли інсульт головного мозку.



Рис. 1.17. Екзоскелетони

Стрімкий розвиток робототехніки та інтелектуальних маніпуляторів відкриває нові можливості у багатьох сферах, зокрема і в медицині. Лікарня – ідеальне середовище для використання роботів. Оригінальні наукові дослідження, спрямовані на створення автономного робота, здатного орієнтуватися у навколишньому середовищі, проводяться в МННЦ ІТІС.

Коротко розглянемо інформацію про медичних роботів, які вже використовують на практиці і здатні взяти на себе розв'язування низки повсякденних задач практичної медицини, наприклад, роздачу ліків пацієнтам, а також більш складніші завдання, такі як допомога хірургу у виконання складних операцій або навчання лікаря-стоматолога (рис. 1.18).

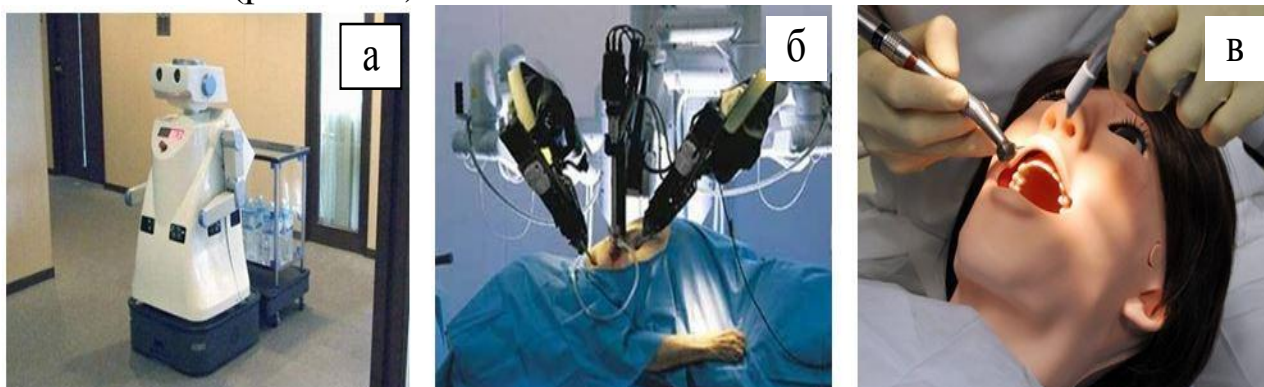


Рис. 1.18. Роботи та маніпулятори у медичній практиці:
а – розвізник ліків, б – помічник хірурга, в – тренажер стоматолога

Робот, якого успішно використовують у медичному центрі «Veterans Affairs Medical Center» для ветеранів, здатний пересуватися у межах клініки, виконуючи роботу середнього медичного працівника, вирішуючи цим проблему нестачі медичних сестер.

Хірургічні маніпулятори, зокрема, відома система “da Vinci” надає змогу значно зменшити розріз і крововтрати під час операції, підвищити точність рухів, скоротити час самої операції і період післяопераційної реабілітації.

Спеціальна відеокамера надає змогу проектувати на екран тривимірне зображення операційного поля. Саме створення технології тривимірного зору забезпечило реалізацію всієї технології, оскільки звичайна двовимірна картинка на екрані монітора не дозволяє оператору точно позиціонувати хірургічний інструмент у просторі, особливо по «глибині» зображення.

Основна перевага «робота-хірурга» – дає змогу лікарю проводити дистанційно мікроманіпуляції з будь-яким інструментом без небезпеки здійснити випадковий незручний рух, який автоматично блокує система.

Техніка проведення сучасних хірургічних операцій постійно ускладнюється. Тому підвищується ризик негативних наслідків для хворого як під час самої операції, так і у післяопераційному періоді. Один із можливих способів зниження таких ризиків – комп’ютерне планування оперативних втручань, яке дозволяє підготувати лікаря-хірурга до можливих помилок та передбачити способи їх запобігання. Такі дослідження проводять у Харківському національному університеті радіоелектроніки.

Став популярним також новозеландський «робот-терапевт», який за командами оператора вміє вимірювати тиск і температуру пацієнта, визначати наявність хрипів у легенях та виконувати низку інших операцій.

Завдяки появі роботів-маніпуляторів імпульс почала розвиватися променева терапія онкологічних хворих. Висока точність позиціонування пучка іонізуючого випромінювання та можливість стежити за рухомою мішенню критично важливі під час опромінення пухлини на таких органах, як мозок, серце чи легені.

Радіохірургічний робот-ніж Cyberknife (рис. 1.19) надає змогу проводити подібні процедури в автоматичному режимі. Вбудована система синхронізації стежить за рухами пацієнта та його органів за допомогою рентгенівської камери та оптичних маркерів на шкірі пацієнта, а спрямований пучок частинок з невеликого лінійного прискорювача руйнує ракові клітини, не зачіпаючи здорову тканину навіть у разі значних зміщень пацієнта.



Рис. 1.19. Радіохірургічний робот-ніж Cyberknife

Інтелектуальні медичні тренажери – ще один важливий напрям у галузі цифрової медицини. За допомогою таких систем лікарі-початківці можуть отримати необхідні знання і підвищити свою кваліфікацію, не завдаючи проблем реальним пацієнтам.

Так, наприклад, японська компанія Кокото створила робот-симулятор “Simroid”, який може реагувати на біль та імітувати відповідні реакції. Якщо студент-стоматолог припускається помилки, то віртуальний пацієнт, у роті якого розміщені спеціальні датчики, корчить гримасу, демонструючи дискомфорт і у разі необхідності подає потрібні голосові повідомлення (рис. 1.18, в).

Норвезька фірма Laerdal розробила інтелектуальний тренажер “SimMan 3G”, який може імітувати різні реакції людини за нещасного випадку: характерні пульсації серця, патологічні рухи легень, скорочення кишечника тощо. Така система може надати неоціненну користь під час навчання фахівців-початківців, які працюють у службах медицини катастроф.

Сучасні можливості телекомунікаційних технологій, що забезпечують фактично миттєву передачу даних на великі відстані, хмарні послуги (Cloud Service), що дають змогу віддалено зберігати величезні масиви біомедичних даних, і системи Грід (Grid computing), орієнтовані на розподілені обчислення, започаткували телемедицину, яка забезпечує інтерактивні консультативні та діагностичні послуги за допомогою дистанційних засобів телекомунікацій.

Консультації у лікаря через Інтернет, отримання лікарняного або рецепту в електронній формі, зберігання всіх медичних документів пацієнта в його особистому кабінеті вже незабаром можуть стати звичайною практикою більшості пацієнтів.

Телемедичний напрям інтенсивно розвивається і в Україні. Згадаємо такі сервіси телемедичних послуг, як LEKARIS.com, що є єдиною комунікаційною платформою для лікарів та пацієнтів. Його завдання – перевести спілкування лікаря та пацієнта в онлайн, прибравши різні бар'єри, заощадивши тим самим час на пошуку лікаря, консультації та особисті прийоми. За допомогою програми користувачі можуть знайти лікаря потрібної спеціалізації, екстрено отримати пораду (повідомлення, аудіодзвінки), запланувати консультацію на зручний день або записатися на особистий прийом.

Інтернет-сервіс ЛІКІ 24 забезпечує цілодобову доставку ліків територією України. Система щохвилини аналізує ціни на продукцію у більш ніж 1300 аптеках, знаходить мінімальну, приймає замовлення та доставляє ліки пацієнтові. Завдяки математичній моделі, що розраховує оптимальний маршрут, один кур'єр за день виконує до 60 доставок.

Мета продукту MYLAB компанії Terra Lab – інтегрувати лабораторні дослідження, клініки та пацієнтів у єдину екосистему. Використовуючи цей сервіс, лікар може отримувати необхідну інформацію про лабораторний ринок, створювати електронні довідки, вказуючи передбачуваний діагноз, отримувати «підказки» про дослідження та відслідковувати статус дослідження в особистому кабінеті.

Наведений короткий та далеко не повний огляд показує, що на сучасному розвитку суспільства наявні методи та засоби цифрової медицини вже активно використовують для розв'язування фактично важливих задач.

Виникає природне запитання: які ж головні перспективні напрями цифрової медицини очікуються в майбутньому.

Аналізуючи доступні публікації, можна виділити два такі напрями (рис. 1.20), які мають і надалі підвищувати ефективність діагностики та лікування на основі принципів персоніфікованої медицини. Обидва напрями стрімко розвиваються. Враховуючи важливість цих напрямів, розглянемо їх більш детально.



Рис. 1.20. Перспективні напрями цифрової медицини

Фірма ІВМ випустила програмні продукти для побудови моделей машинного навчання, які на запит лікаря здатні всього за три секунди прочитати та структурувати близько 200 мільйонів сторінок тексту з Інтернету. Таку інформацію лікар може використати для підтримки прийняття діагностичних рішень та вибору оптимальної тактики лікування конкретного пацієнта на основі зіставлення фізіологічних параметрів, симптомів історії хвороби, ДНК та інших індивідуальних характеристик його організму зі схожими випадками у світовій лікувальній практиці.

Подальший розвиток цього напрямку фірма орієнтує на вирішення таких проблем:

- забезпечення індивідуального підходу на лікування онкологічних хворих;
- вивчення мутації ракових клітин, що призводять до появи стійкості до дії лікарських засобів;
- пошук нових, більш ефективних, лікарських препаратів на основі аналізу генетичних даних сотень тисяч пацієнтів та здорових людей.

Важливі результати отримано також у галузі другого напрямку досліджень. Останнім часом ринок медичних виробів суттєво змінив свій напрям від складних систем клінічного використання, на якому

спостерігається відносна стагнація, до портативних цифрових приладів, якими пацієнти можуть користуватися самостійно.

Актуальність цього напряму очевидна. Тяжкі хвороби досить часто потребують стаціонарного лікування, але мало хто почувається комфортно у лікарні. Госпіталізація завдає значної психологічної травми літнім людям та дітям. Крім того, стаціонарне лікування зазнає значних економічних втрат як для пацієнта, так і для системи державної охорони здоров'я.

Так з'явився новий термін – домашній стаціонар. Під цим розуміють постійний нагляд та реабілітацію в домашніх умовах людей похилого віку, які перенесли тяжкі хвороби – інсульт, гострий інфаркт міокарда, перелом шийки стегна тощо. Значні успіхи науки в галузі сучасних інформаційних технологій, зокрема інтелектуальних ІТ з елементами мислення людини, дозволяють уже сьогодні суттєво підвищити ефективність домашнього стаціонару.

Домашній стаціонар майбутнього дозволить наблизити сучасні засоби цифрової медицини безпосередньо до пацієнтів, зокрема, до хворих на серцево-судинні захворювання, які досі залишаються головною причиною смертності та інвалідності працездатного населення у розвинених країнах.

За статистичними даними в 53 країнах Європи щорічно від серцево-судинних захворювань помирають понад 4 млн осіб, що перевищує смертність від усіх злоякісних новоутворень разом узятих. Тільки за один рік у США економічні втрати від інвалідності та передчасної смерті кардіологічних хворих становлять астрономічну суму – понад 160 мільярдів доларів.

В Україні ситуація ще драматичніша: у структурі смертності серцево-судинні захворювання становлять 65–68 % і цей показник один з найвищих у світі. Близько чверті померлих від серцево-судинних захворювань становлять люди до 65 років. Відбувається значне омоложення серцево-судинних захворювань – раптово помирають школярі на уроках фізкультури та спортсмени на тренуваннях та змаганнях. Приблизно в половині випадків госпіталізація з приводу інфаркту міокарда є першим у житті контактом хворого з кардіологом.

Один із способів профілактики ішемічної хвороби серця (ІХС) – пошук методів ранньої діагностики ішемічних змін у міокарді, оснований на оптимізації наявних та розробці нових методів оцінювання стану серця.

1.3. Інтелектуальні інформаційні технології оброблення сигналів складної форми

Одним із напрямів застосування сучасних інформаційних технологій є діагностика стану технічних та біологічних об'єктів. Зазвичай такі технології ґрунтуються на комп'ютерному обробленні сигналів, що надають інформацію про кількісні або якісні показники поточного стану об'єкта.

Загальним закономірностям отримання інформації зі спостережень присвячені фундаментальні наукові праці Р. Л. Хартлі, К. Шеннона, Н. Вінера, О. М. Колмогорова, В. О. Котельнікова. Значний внесок у розвиток теорії цифрового оброблення сигналів зробили Т. К. Вінцюк, Я. П. Драган, В. К. Задирака, Ю. П. Кунченко, В. О. Омельченко та інші українські фахівці.

Діагностика стану об'єктів здебільшого пов'язана з необхідністю вилучення інформації з сигналів складної форми, коли діагностичні ознаки зосереджені на невеликих інтервалах $\Delta t_i \ll T_0$, $\forall i = 1, 2, \dots$ області визначення $T_0 = [t_1, t_2]$ часового сигналу $z(t)$. Типові приклади – електрокардіограми (ЕКГ), реограми, сфігмограми та інші фізіологічні сигнали, локальні фрагменти яких показують інформацію про стан серцево-судинної системи.

Класичні методи ортогональних розкладів

$$z(t) = \sum_{n=1}^N C_n \varphi_n(t) \quad (1.1)$$

за системою базисних функцій $\varphi_n(t)$ практично не придатні для оброблення таких сигналів, оскільки коефіцієнти C_n функціонального ряду (1.1), побудованого по всій області $T_0 = [t_1, t_2]$, майже нечутливі до змін діагностичних ознак, зосереджених на локальних інтервалах.

Проблема видобування діагностичної інформації істотно ускладнюється, коли спостережуваний сигнал $z(t)$ спотворений обуреннями, які не завжди зводяться лише до адитивних завад.

Розглянемо загальну ідею побудови інформаційних технологій для обробки ЕКГ та інших сигналів з локально зосередженими ознаками.

Введемо узагальнений показник s стану об'єкта, під яким розумітимемо вектор $s = (s_1, \dots, s_N)$ з числовими компонентами s_1, \dots, s_N , якщо потрібний кількісний стан об'єкта або значення індикаторної змінної $s = 1, \dots, M$, то якісний стан об'єкта з кінцевої множини $V = \{V_1, \dots, V_M\}$.

Вважатимемо, що безпосереднє визначення s неможливе. Наприклад неможливо в умовах масових профілактичних обстежень визначати стан серцево-судинної системи на основі інвазивного методу коронарографії. Цей метод достатньо дорогий і, найголовніше, небезпечний для пацієнта, що обстежується. Список таких прикладів можна було б продовжити.

У таких ситуаціях не залишається нічого іншого, як опосередковано оцінювати s , використовуючи доступну інформацію, зокрема, електрокардіограму. Вирішення цієї проблеми є головною метою інтелектуальних ІТ оброблення сигналів.

В основу синтезу описаних технологій покладемо таку гіпотезу. Вважатимемо, що досліджуваний об'єкт за фіксованого стану $s \in S$ породжує фізичне поле, яке змінюється у часі:

$$z_0 = \Psi_s(u, t), \quad (1.2)$$

де t – час, u – точка простору навколо об'єкта, а Ψ_s – у загальному випадку невідома функція, яка залежить від стану s .

Наприклад, досліджуваний об'єкт – серце породжує електромагнітне поле у вигляді (1.2), яке має діагностичну інформацію про поточний стан серцево-судинної системи.

Якщо вимірювати величину поля (1.2) у фіксованій точці простору $u = u_0$, то функція $\Psi_s(u, t)$ вироджується у скалярну функцію часу

$$z_0 = \Psi_s(u_0, t) \equiv z_0(t), \quad (1.3)$$

а якщо поле (1.2) вимірюється в кількох фіксованих точках u_1, u_2, \dots, u_p простору, то маємо вектор-функцію часу

$$z_0(t) = (z_1(t), \dots, z_p(t)). \quad (1.4)$$

Стосовно електромагнітного поля серця функції (1.3) та (1.4) можна інтерпретувати як ЕКГ-сигнали відповідно в одному та в декількох відведеннях.

Таким чином, з формального погляду ІТ оброблення сигналу в загальному випадку має забезпечувати розв'язок зворотної задачі: потрібно оцінити невідоме поточне значення s стану об'єкта за результатами вимірювання сигналу $z_0(t)$ на відрізьку спостереження $T_0 = [t_1, t_2]$ (рис. 1.21).



Рис. 1.21. Абстрактна модель ІТ оброблення сигналу

У переважній більшості задач функція $\Psi_s(\cdot)$ є невідомою. І хоча можна вважати, що є закономірність (1.2), побудувати адекватну модель об'єкта з урахуванням фізичних уявлень важко, а іноді майже неможливо. Тому розглядатимемо (1.1) лише як робочу гіпотезу, що дозволяє обґрунтувати принципову можливість побудови ІТ для оцінювання стану об'єкта s за сигналом $z_0(t)$, що спостерігається.

Оскільки за відсутності аналітичного опису функції $\Psi_s(\cdot)$ проблема побудови ІТ не може бути зведена лише до математичної задачі, яка має аналітичний розв'язок, тому будемо розглядати цю проблему більш загально з боку методології розв'язування практичних задач на основі формальних та неформальних прийомів вибору та налаштування базового набору обчислювальних компонентів.

Проблема побудови ефективної ІТ оброблення сигналу істотно ускладнюється, коли сигнал $z(t)$, що обробляється, спостерігається в умовах внутрішніх неадитивних $\xi(t)$ та зовнішніх адитивних $h(t)$ завад. Тобто загалом будемо вважати, що

$$z(t) = \Phi[z_0(t), \xi(t)] + h(t), \quad (1.5)$$

де $z_0(t)$ – корисний сигнал, а $\Phi[\cdot]$ – невідома функція.

У задачі комп'ютерного оброблення ЕКГ корисний сигнал $z_0(t)$ надає інформацію про патологічні зміни серцевого м'яза, внутрішні завади $\xi(t)$ породжують система органів дихання та вегетативна нервова система, а зовнішні збурення $h(t)$ створюють мережеві електричні перешкоди, м'язовий тремор тіла пацієнта у місцях накладання електродів та низка інших збурень.

На рис. 1.22 зображено узагальнену структуру прикладної ІТ оброблення сигналів з локально зосередженими ознаками.



Рис. 1.22. Узагальнена структура ІТ

Як видно, прикладна ІТ має забезпечити:

- відновлення корисного сигналу $z_0(t)$ за спостереженням $z(t)$ в умовах дії внутрішніх $\xi(t)$ та зовнішніх $h(t)$ збурень;
- розпізнавання інформативних фрагментів $z_0^{(1)}(t), \dots, z_0^{(F)}(t)$ відновленого сигналу $z_0(t)$, на яких зосереджено діагностично цінна інформація;
- аналіз значень діагностичних ознак x_1, \dots, x_N , зосереджених на розпізнаних фрагментах $z_0^{(1)}(t), \dots, z_0^{(F)}(t)$;
- реалізацію діагностичного правила $s = s(x_1, \dots, x_N)$, що забезпечує оцінювання стану об'єкта s за ознаками x_1, \dots, x_N .

Можливі формальний та неформальний (евристичний) підходи до побудови обчислювальних процедур, що реалізують окремі етапи ІТ.

Перший підхід ґрунтується на вирішенні оптимізаційної задачі: обчислювальна процедура будується формально з умови мінімуму (максимуму) деякого критерію (суперкритерію) \mathfrak{Z} , що характеризує ефективність конкретного етапу оброблення.

Якщо ж неможливо побудувати обчислювальну процедуру на основі формального підходу, то не залишається нічого іншого, як будувати обчислювальну процедуру неформально на основі інтуїції, а критерій \mathfrak{Z} використовувати лише для оцінювання ефективності процедури. В таких ситуаціях вибір та налаштування комп'ютерних процедур можна проводити методом проб та помилок.

Розглянемо підхід до інтерактивного синтезу прикладних інформаційних технологій

1.4. Інтерактивний синтез прикладних ІТ

Кожен об'єкт конкретної предметної області породжує унікальний сигнал, що накладає специфічні умови на комп'ютерну технологію вилучення прихованої інформації. Однак, якщо дотримуватися лише такої логіки, то розробку нової ІТ потрібно розпочинати майже з нуля.

Завдання технології, як науки, полягає у виявленні загальних закономірностей з метою визначення та використання фактично найбільш ефективних і економних виробничих процесів. У нашому ви-

падку такими процесами є інтелектуальні обчислювальні процедури, які забезпечують ефективний перехід від «сировини» технології $z(t)$ до інформаційного продукту, орієнтованого конкретного користувача.

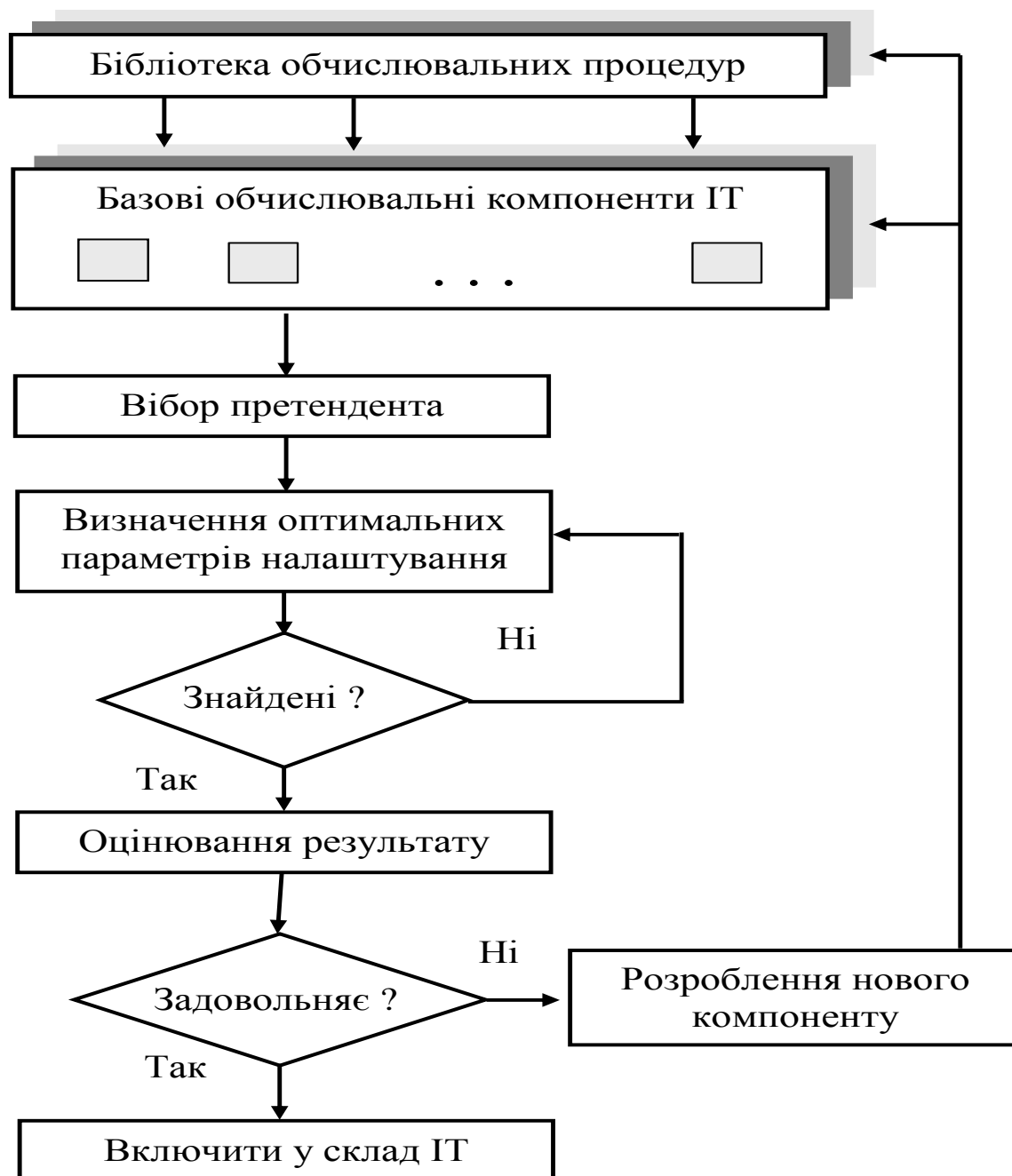


Рис. 1.23. Концептуальна ідея синтезу прикладної ІТ

Тому для побудови ефективної ІТ необхідно:

- розробити узагальнені математичні моделі породження сигналів певного класу та вивчити загальні закономірності таких сигналів;

- запропонувати сукупність обчислювальних процедур, орієнтованих на розв'язування окремих етапів комп'ютерного оброблення;
- створити програмні компоненти, які реалізують необхідні обчислювальні процедури;
- сформулювати критерії «оптимального» налаштування та вибору найбільш ефективних обчислювальних компонентів для побудови конкретної прикладної ІТ.

Для практичної реалізації ідеї інтерактивного синтезу прикладних ІТ обробки сигналів складної форми в МННЦ ІТІС було розроблено спеціальну інструментальну систему СІДІГРАФ (сигнал-діалог-графік). Під час побудови системи було враховано величезний науковий потенціал у сфері цифрового оброблення сигналів, методів розпізнавання образів і моделювання складних систем за експериментальними даними.

Водночас виявилось, що для реалізації прикладних систем оброблення сигналів з локалізованою інформацією необхідно підвищити «інтелект» цілої низки відомих алгоритмів. Тому до бібліотеки обчислювальних процедур включено широкий набір як традиційних, так і оригінальних алгоритмів, зокрема обчислювальних процедур, що забезпечують:

- генерацію детермінованих та випадкових послідовностей;
- модифікацію даних (інтерполяцію, передискретизацію, формування сумішей, ланцюжків тощо);
- фільтрацію, зокрема адаптивну частотно-виборчу та рангову;
- адаптивне згладжування даних;
- статистичний та спектральний аналіз;
- аналіз хаотичності параметрів сигналів, що спостерігаються на основі традиційних та оригінальних ентропійних оцінок;
- перехід від сигналу $z(t)$ до його відображення у нормованих фазових координатах $z(t)$, $\dot{z}(t)$ або $z(t)$, $z(t - \tau)$, де τ – затримка в часі;
- виявлення характерних точок, розпізнавання фрагментів та сегментацію одновимірних та двовимірних послідовностей.

Корисний інструмент образного розв'язання завдання – когнітивна комп'ютерна графіка, яка дозволяє або відразу розв'язати задачу, або отримати підказку для його знаходження. Такі можливості когнітивної графіки обумовлені тим, що людський мозок набагато легше сприймає та інтерпретує графічний образ, ніж числові дані, що його

породили. На основі аналізу когнітивного графічного образу розв'язок задачі може бути знайдено без складних обчислень.

Зважаючи на це, графічний інтерфейс інструментальної системи реалізує широкий набір процедур відображення та маніпуляції з графічними образами.

Для надання інструментальній системі певної гнучкості та можливості подальшого розширення її ядро побудовано на двох програмних класах – узагальненої моделі носія даних (МНД) та узагальненої моделі обробки даних (МОД).

Узагальнена модель носія даних МНД (рис. 1.24) є абстрактним класом, на основі якого створюються об'єкти (спадкоємці), що забезпечують:

- зберігання цифрових даних, які необхідні для організації елементарних обчислювальних процедур на різних стадіях оброблення сигналу;
- відображення результатів обробки окремих стадій у вигляді екранних графічних образів та текстової інформації;
- доступ до даних через їхні графічні образи;
- модифікацію графічних образів.

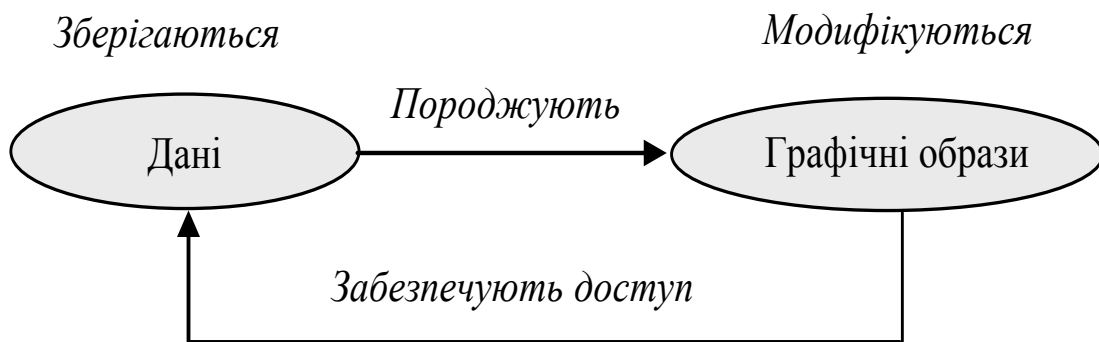


Рис. 1.24. Концептуальна ідея побудови класу МНД

Клас МНД виконано у вигляді трійки

$$\text{МНД} = \langle D_C, D_G, G \rangle, \quad (1.8)$$

де D_C – поля змістовних даних; D_G – поля графічних даних; G – методи відображення та модифікації графічних образів.

Об'єднання в одному об'єкті даних та методів їх візуалізації відкриває можливість отримувати доступ до даних через їх екранні графічні образи та забезпечити необхідне прив'язування графічного об-

разу, що візуалізує результат, до графічного образу вихідних даних. Такі можливості є дуже корисними, а часом і необхідними під час синтезу прикладних ІТ оброблення сигналів.

Узагальнена модель оброблення даних (МОД) також є абстрактним класом з універсальними полями, властивостями і методами, інваріантними щодо конкретної процедури оброблення, причому МОД містить віртуальний метод «порожньої» обробки, який визначається в спадкоємцях цього класу.

Тим самим абстрактний клас МОД дозволяє уніфікувати поняття елементарного оброблення даних, упорядкувати процес створення програмних компонентів, що реалізують окремі стадії обробки, та організувати зручний інтерфейс взаємодії між активними екземплярами (об'єктами) класу МНД та конкретною процедурою оброблення (рис. 1.25).

Клас МОД підтримує уніфікований інтерфейс виклику будь-якої бібліотечної обчислювальної процедури:

$$\langle \text{Имя} \rangle (I_1, \dots, I_{N_1}; L_1^{(I)}, \dots, L_{N_1}^{(I)}; O_1, \dots, O_{N_1}; L_1^{(O)}, \dots, L_{N_2}^{(O)}; \alpha_1, \dots, \alpha_Q; \text{Mess}), (1.9)$$

де $I_1, \dots, I_{N_1}; O_1, \dots, O_{N_1}$ – покажчики на масиви вхідних (Input) та вихідних (Output) даних; $L_1^{(I)}, \dots, L_{N_1}^{(I)}; L_1^{(O)}, \dots, L_{N_2}^{(O)}$ – кількість елементів у відповідних масивах; $\alpha_1, \dots, \alpha_Q$ – параметри налаштування процедури обробки; Mess – рядок повідомлення про виняткову (аварійну) ситуацію.

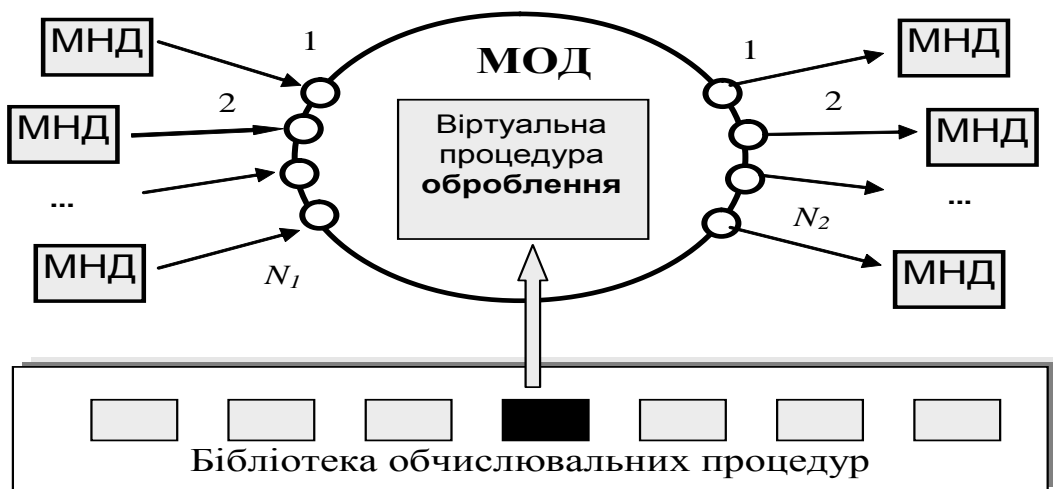


Рис. 1.25. Схема взаємодії класів МНД та МОД в інструментальній системі

Такий інтерфейс надав змогу організувати паралельну роботу програмістів, що беруть участь у створенні інструментальної системи: прикладні програмісти мають можливість зосередитися на їхньому головному завданні – реалізації ефективних обчислювальних процедур обробки даних, не дбаючи про організацію діалогу з користувачем та методи графічного подання результатів, а системні програмісти – розглядати процедуру оброблення як «чорну скриньку».

В основу ІТ обробки сигналів покладено технологічний ланцюжок, що складається з кінцевого числа N_Q обчислювальних процедур (обробок), за допомогою яких вектор дискретних значень сигналу, що спостерігається, послідовно перетворюється на вихідний продукт – оцінку \hat{s} поточного стану об'єкта:

$$D_R^{(1)} = A_1(\alpha_1, Y) \rightarrow \dots \rightarrow D_R^{(i)} = A_i(\alpha_i, D_S^{(i)}) \rightarrow \dots \rightarrow D_R^{(N_Q)} = \hat{s}, \quad (1.10)$$

де $A_i(\cdot)$ – алгоритм перетворення вихідних даних $D_S^{(i)}$ у результат $D_R^{(i)}$ обробки на i -й стадії; α_i – параметри налаштування алгоритму, причому під вихідними даними i -ї стадії в загальному випадку розуміється не тільки результат $D_R^{(i-1)}$, а й будь-які попередні результати.

Якщо конструктор ІТ має повну інформацію про кожну i -ту стадію оброблення $D_R^{(i)} = A_i(D_S^{(i)})$, то синтез ІТ зводиться до традиційних етапів життєвого циклу розробки програмного забезпечення.

Однак досвід показує, що під час синтезу нових ІТ оброблення сигналів, що породжуються складними об'єктами, основний час витрачається не стільки на програмну реалізацію обчислювальних алгоритмів, як на експериментальні дослідження, спрямовані на пошук відповідних алгоритмів та їх оптимальне налаштування.

Інтерактивний синтез прикладної ІТ зводиться до вибору та оптимального налаштування базових обчислювальних компонент самої інструментальної системи (рис. 1.26).

Для зручності налаштування параметрів $\alpha_1, \dots, \alpha_Q$ обчислювальних процедур в інструментальній системі організовано «віртуальний» зв'язок органу управління та графічного образу, що відображає результат оброблення. Одночасно з переміщенням повзунка налаштування відбувається зміна графічного образу і конструктор може візу-

ально визначити оптимальні значення параметрів налаштування $\alpha_1, \dots, \alpha_Q$ у випадках, коли відсутні формальні методи оптимізації вибору параметрів (рис. 1.27).

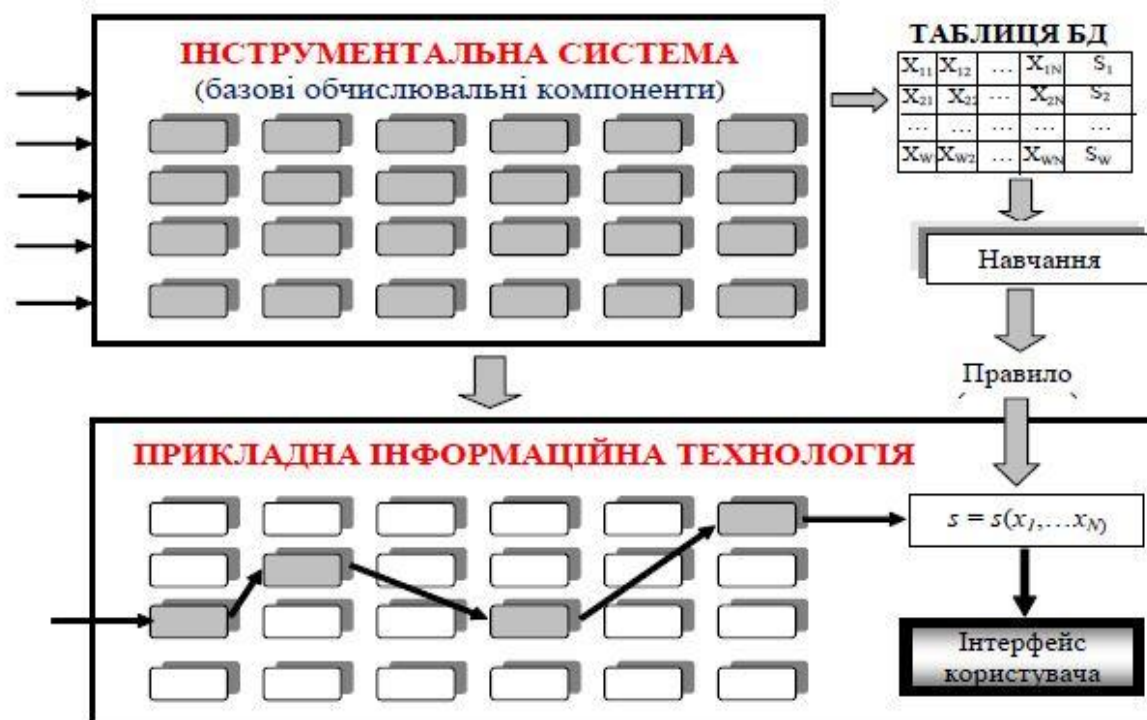


Рис. 1.26. Організація ланцюжка обчислювальних процедур прикладної ІТ

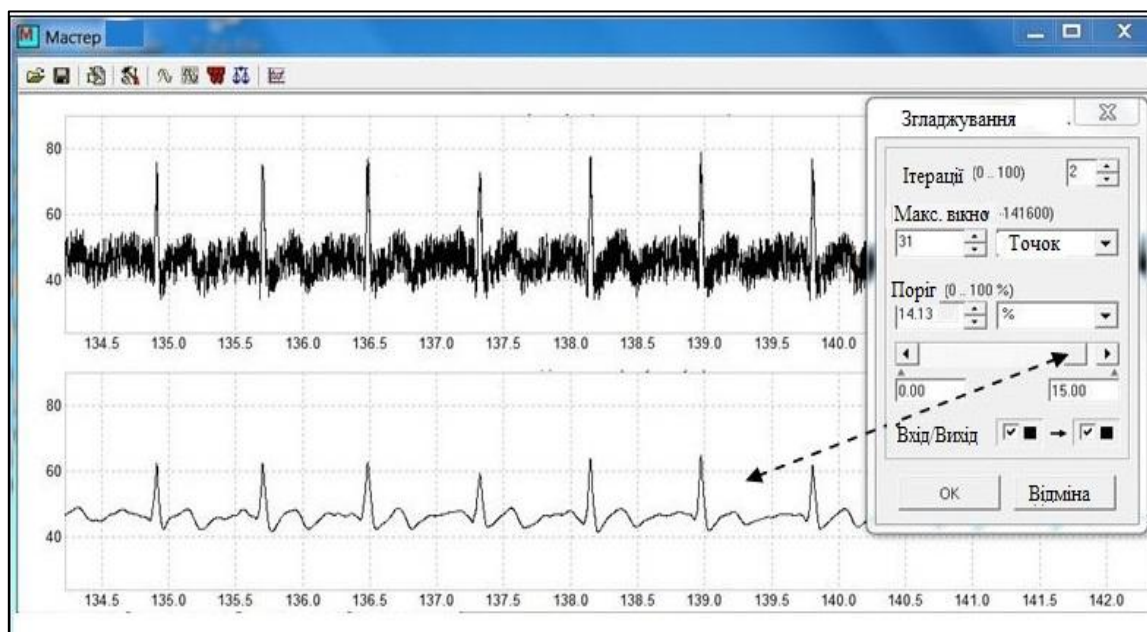


Рис. 1.27. Робоче вікно інструментальної системи в режимі інтерактивного налаштування параметрів

Таким чином, за допомогою інструментальної системи СІДІГРАФ можна:

- проводити необхідні експерименти для вибору та оптимального налаштування обчислювальних процедур, що реалізують окремі елементарні стадії обробки, спираючись на результати чисельного моделювання та засобів когнітивної комп'ютерної графіки;
- реалізувати технологічний ланцюжок алгоритмів обробки сигналу, за можливості використовуючи готові обчислювальні компоненти інструментального середовища;
- розширювати за необхідності склад компонентів інструментальної системи (базових обчислювальних процедур та засобів відображення результатів), якщо наявний набір цих засобів є недостатнім для створення конкретної прикладної системи.

Аналіз проблеми інтерактивного синтезу ІТ обробки сигналів складної форми буде не повним, якщо не розглянути сучасні тенденції побудови програмних продуктів, що ґрунтуються на методології об'єктно-орієнтованого програмування (ООП). Фундаментальне поняття ООП – програмний клас, який розглядають як абстрактну сукупність об'єктів (примірників класу), що мають спільні властивості.

Важливою особливістю класів є можливість їх організації у вигляді деякої ієрархічної структури завдяки механізму спадкування, причому об'єкти спадкоємців можуть модифікувати свою зовнішню поведінку відносно однойменних методів батьківського класу, використовуючи механізм поліморфізму.

Методологія ООП підтримується найсучаснішими мовами програмування високого рівня, при цьому взаємодія окремих компонентів програми один з одним і зовнішнім світом (користувачем) здійснюється з використанням механізму передачі повідомлень (керуючих подій).

Звідси випливає, що головною проблемою, що виникає під час програмної реалізації прикладної ІТ на основі методології ООП – конструювання програмних класів, достатніх для досягнення поставленої мети, зокрема визначення властивостей та методів, необхідних для надання класам необхідної поведінки та можливості їхньої взаємодії.

На жаль, цю проблему не можна вирішити формальними методами, а значить, успіх «оптимальної» декомпозиції цілком залежить від досвіду системного аналітика – архітектора проектованої системи.

Все більша складність програмних систем призвела до створення інженерних засобів розробки програмного забезпечення – CASE-технологій (Computer Aided Software Engineering). Слід, однак, визнати, що такі засоби переважно орієнтовані на створення організаційних ІТ, у яких основними завданнями є проектування складних баз даних, організація специфічних запитів, підготовка звітів тощо.

Водночас основні проблеми побудови ІТ оброблення сигналів перебувають в «алгоритмічній» площині. Тому під час побудови таких ІТ потрібно проводити додаткові експериментальні дослідження щодо вивчення сигналів, що створюються об'єктом досліджень. Традиційною інструментальною підтримкою таких досліджень є сучасні засоби комп'ютерної математики.

Однак такі засоби зручні і корисні лише під час розробки прототипу прикладної ІТ. Досвід показує, що якщо на етапі експериментальних досліджень вдається вибрати відповідний ланцюжок обчислювальних процедур, доступних у пакеті комп'ютерної математики, то на кінцевому етапі побудови прикладної ІТ однаково доводиться писати власні програми мовою програмування високого рівня.

Саме тому базові обчислювальні компоненти інструментальної системи СІДІГРАФ реалізовані мовами програмування високого рівня у вигляді готових до використання програмних компонентів (плагінів), що суттєво прискорює та підвищує якість функціонування прикладних ІТ.

Завдання до комп'ютерних практикумів

Завдання 1.1. Необхідно розробити програму, яка імітує інтелектуальний діалог з користувачем. У користувача має виникнути ілюзія, що розмова відбувається з реальною людиною, подібно до спілкування у чаті.

Однак співрозмовник користувача – це чітко продумані відповіді розробника програми. Для перевірки якості створеної програми необхідно реалізувати лічильник, що розраховує загальну кількість питань-відповідей у проведеному діалозі, які виявили те, що співрозмовник є програма, а не людина.

Програма може реалізувати, наприклад, такий діалог користувача з віртуальним лікарем:

ЛІКАР (програма):	<i>Як Вас звати?</i>
ЛЮДИНА (користувач):	<i>Микола.</i>
ЛІКАР (програма):	<i>Я Вас слухаю, Миколо.</i>
ЛЮДИНА (користувач):	<i>Мене бентежить погода.</i>
ЛІКАР (програма):	<i>Відтоді, як померла моя мати.</i>
ЛЮДИНА (користувач):	<i>Розкажіть про Вашу сім'ю.</i>
ЛІКАР (програма):	<i>Вони полюбляють</i>
	<i>чери.</i>
ЛЮДИНА (користувач):	<i>Комп'ютери Вас лякають.</i>
ЛІКАР (програма):	<i>Ні.</i>
ЛЮДИНА (користувач):	<i>Ви такий небагатослівний...</i>
ЛІКАР (програма):	<i>...</i>

Завдання 1.2. Необхідно розробити програму, яка імітує інтелектуальний прогноз за знаком зодіаку. За введеною користувачем датою народження програма має визначити його знак зодіаку, надати відомі факти, характерні для цього знаку, та надати прогноз на поточний день – заздалегідь підготовлене випадкове повідомлення зі списку, що імітує «інтелектуальні здібності» програми.

Запитання для самоконтролю

1. Назвіть три напрями наукових досліджень у галузі штучного інтелекту.
2. Хто є засновником алгоритмічного підходу до створення штучного інтелекту?
3. У чому полягає тест «Китайська кімната»?
4. Дайте визначення поняттям «Технологія», «Інформаційна технологія» та «Інтелектуальна інформаційна технологія».
5. Опишіть основні властивості інтелектуальних систем.
6. Що таке сигнал з локально зосередженими діагностичними ознаками. Наведіть приклад.
7. Назвіть основні задачі засобів цифрової медицини.
8. У чому полягає задача інтелектуальної ІТ оброблення біомедичних сигналів.
9. На яких принципах може бути реалізовано інтерактивний синтез прикладних ІТ.

РОЗДІЛ 2

МАТЕМАТИЧНІ МОДЕЛІ ПОРОДЖЕННЯ ШТУЧНИХ СИГНАЛІВ

2.1. Загальна характеристика електрокардіографії

У розділі 1 було показано, що основне завдання створення нових інформаційних технологій оброблення сигналів з локально зосередженими ознаками є створення інтелектуальних обчислювальних процедур, які забезпечують ефективний перехід від сигналу, що спостерігається, до інформації, орієнтованої на конкретного споживача. Можливі формальний та неформальний підходи до створення обчислювальних процедур, що реалізують окремі етапи ІТ.

Формальний підхід ґрунтується на вирішенні оптимізаційної задачі: обчислювальна процедура будується за умови мінімуму (максимуму) деякого критерію \mathfrak{Z} , що характеризує ефективність конкретного етапу оброблення.

Наприклад, байєсова стратегія класифікації об'єктів, яку буде розглянуто далі, ґрунтується на мінімумі критерію \mathfrak{Z} , що є середнім ризиком прийнятих рішень.

Однак не завжди формальний підхід є можливий. У таких випадках обчислювальну процедуру будують неформально, спираючись на інтуїцію та попередній досвід конструктора ІТ, а критерій \mathfrak{Z} використовувати лише для оцінювання ефективності побудованої процедури.

Означення 2.1. Процедура \mathfrak{R} вважається ефективною, якщо виконується строга нерівність:

$$\mathfrak{Z}_1 < \mathfrak{Z}_0,$$

де \mathfrak{Z}_0 – апріорне значення критерію \mathfrak{Z} певного етапу ІТ, який оцінено до використання обчислювальної процедури \mathfrak{R} , а \mathfrak{Z}_1 – апостеріорне значення \mathfrak{Z} , що оцінено після використання процедури \mathfrak{R} .

Означення 2.2. Процедура $\mathcal{R}^{(i)}$ вважається ефективнішою порівняно з процедурою $\mathcal{R}^{(j)}$, якщо для апостеріорних значень відповідних критеріїв справедлива строга нерівність:

$$\mathfrak{Z}_1^{(i)} < \mathfrak{Z}_1^{(j)}.$$

Отже, неформальна побудова обчислювальної процедури не обов'язково має вирішувати оптимізаційне завдання, спрямоване на пошук глобального мінімуму. Більше того, відповідно до наведених означень, процедура \mathcal{R} залишається ефективною навіть якщо вона не задовольняє локальному мінімуму критерію \mathfrak{Z} , а лише надає змогу зменшити значення критерію порівняно з його апіорним значенням \mathfrak{Z}_0 .

Зрозуміло, що сформульовані означення природним чином можуть бути переформульовані, якщо процедура \mathcal{R} спрямована на збільшення значення.

На підставі введених означень вибір обчислювальної процедури можна забезпечити методом спроб і помилок за схемою, наведеною на рис. 1.23.

Електрокардіограма (ЕКГ), яка показує інформацію про електричну активність серця, є наочним прикладом природних сигналів з локально зосередженими ознаками. Розглянемо на прикладі ЕКГ дві математичні моделі генерації штучних сигналів реалістичної форми. Ці моделі будуть використані далі для обґрунтування ефективності запропонованих оригінальних обчислювальних процедур, які мають властивості природного інтелекту.

Анатомічно серце складається з чотирьох камер – двох передсердь і двох шлуночків (рис. 2.1). Артеріальна кров, насичена киснем O_2 , від лівого шлуночка надходить до органів, а венозна кров, насичена вуглекислим газом CO_2 , повертається до серця. Зауважимо, що на рис. 2.1 ліве передсердя та лівий шлуночок показано справа, а праве передсердя та правий шлуночок навпаки показано зліва. Це називають великим колом кровообігу.

Венозна кров від правого шлуночка надходить до легенів, завдяки диханню перетворюється в артеріальну кров насичену киснем O_2 та повертається до серця. Це є малим колом кровообігу.

Треба звернути увагу на те, що може викликати непорозуміння. Медики назвали кровоносні судини, за якими кров рухається від серця, артеріальними, а судини, що входять до серця – венозними. Тому на малому колі кровообігу венозна кров рухається за легеневою артерією, а насичена киснем артеріальна кров рухається за легеневою веною.



Рис. 2.1. Спрощена структура серцево-судинної системи

Серце має низку функцій, що визначають особливості його роботи. До них належать функції

- **автоматизму**, що полягає у здатності серця самостійно виробляти електричні імпульси (за відсутності зовнішніх подразників);
- **збудливості**, що полягає у здатності ділянок серця збуджуватись під впливом електричних імпульсів;
- **провідності**, що полягає у здатності збуджуватись від однієї ділянки серця до інших відділів серцевого м'яза (міокарда);
- **скоротливості**, що полягає у здатності міокарда скорочуватись у відповідь на збудження.

Внаслідок послідовного скорочення різних відділів серця здійснюється його основна насосна функція.

В основі виникнення електричних явищ у серці є механізм проникнення іонів через м'язову мембрану, головним чином іонів калію K^+ та натрію Na^+ . В області серця одночасно функціонує велика кількість елементарних джерел струму – серцевих диполів, кожен з яких обумовлений збудженням окремих м'язових волокон і відрізняється від інших таких диполів як за величиною, так і за напрямом.

Скорочення м'яза серця поділяється на дві фази – **деполяризацію**, яка полягає в зменшенні різниці потенціалів, що перебуває між внутрішньою та зовнішньою стороною мембран клітин серця, та **реполяризацію**, яка полягає у відновленні цієї різниці потенціалів.

Завдяки механізму електромеханічного сполучення процес електричного збудження міокарда викликає його механічне скорочення – систолу серця, а після чого настає розслаблення – діастола серця. Систола і діастола разом утворюють серцевий цикл ЕКГ (рис. 2.2).

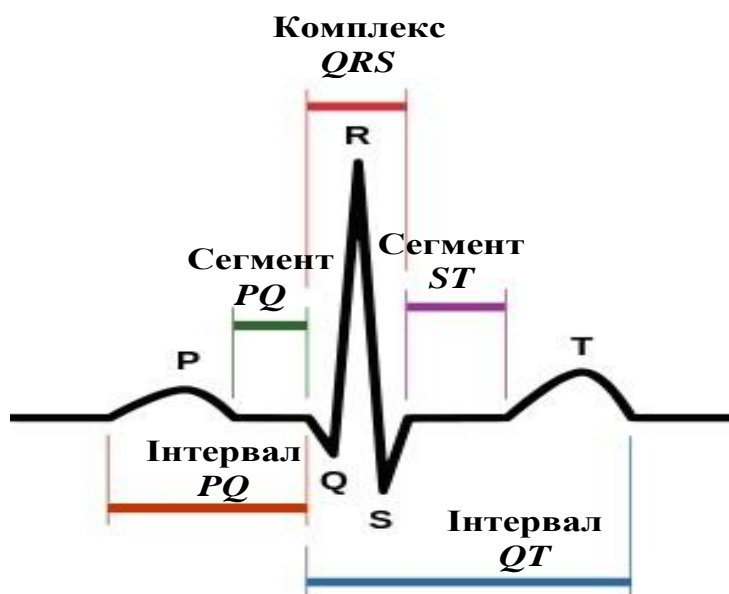


Рис. 2.2. Серцевий цикл здорової людини

Зубець P утворюється внаслідок збудження передсердь. Сегмент $P-Q$ зазвичай розміщується на нульовій лінії (ізолінії) і відповідає часу проходження збудження по атріовентрикулярному (передсердно-шлуночковому) з'єднанню. Комплекс QRS відбиває складний процес збудження (деполяризації) шлуночків. Початковий зубець Q ре-

еструється під час збудження лівої частини міжшлуночкової перегородки.

Зубець R (найчастіше найбільш виражений зубець ЕКГ), обумовлений збудженням основної маси міокарда лівого та правого шлуночків. Зубець S переважно обумовлений збудженням основи лівого шлуночка. Інтервал ST називають кінцевою частиною шлуночкового комплексу, який відображає реполяризацію шлуночків. Він поділяється на сегмент ST , який відповідає періоду згасання збудження шлуночків, і зубець T , форма якого відповідає процесу швидкої реполяризації шлуночків. Інтервал $T - P$ від закінчення зубця T і до початку зубця P наступного циклу ЕКГ відповідає розслабленню серця – діастолі.

У кардіологічній практиці найчастіше ЕКГ вимірюється у кількох відведеннях. Кожне відведення містить інформацію про різниці потенціалів між двома певними точками електричного поля серця, в яких встановлені електроди. Зазвичай використовують 12 відведень ЕКГ – три стандартні відведення, які позначають римськими цифрами I, II, III, три посилені відведення aVR , aVL , aVF та шість грудних відведень V_1, \dots, V_6 (рис. 2.3). Кожен із відведень переважно відображає стан ділянок серцевого м'яза певної локалізації.



Рис. 2.3. Традиційні 12 відведень ЕКГ

2.2. Інтерполяційна модель породження ЕКГ

Згідно з сучасними уявленнями фізіологів нерегулярність та непередбачуваність частоти серцевих скорочень (ЧСС) є важливими характеристиками здорового організму, причому тривалість комплексу QRS щонайменше пов'язана зі зміною ЧСС, ніж тривалість зубців P і T .

Розглянемо один із можливих підходів до побудови моделі породження циклу реальної ЕКГ із нерівномірними за часом змінами тривалостей окремих фрагментів.

Будемо вважати, що сигнал $z(t)$, що спостерігається в одному з відведень, наприклад, у першому стандартному відведенні (ліва та права руки), є результатом спотворення корисного сигналу $z_0(t)$ електричної активності серця внутрішніми $\xi(t)$ та зовнішніми $h(t)$ збуреннями, де $\Phi(\cdot)$ – невідома функція (рис. 2.4).

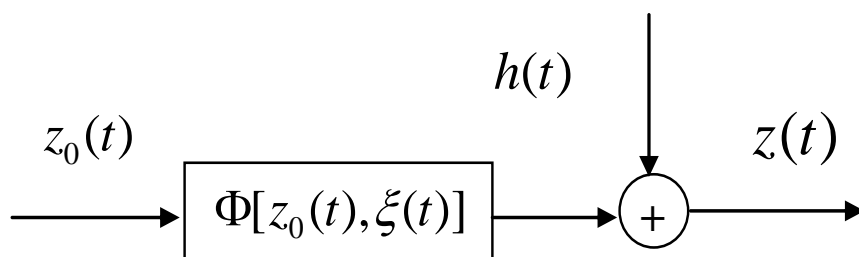


Рис. 2.4. Механізм породження реальної ЕКГ

Вважаємо, що внутрішні збурення $\xi(t)$ викликає сам біологічний об'єкт, зокрема, система органів дихання та вегетативна нервова система, а адитивні збурення $h(t)$ обумовлені докільям – мережевими перешкодами, м'язовим тремором у місцях накладання електродів та іншими завадами.

Розглянемо корисний сигнал $z_0(t)$ у вигляді послідовності функцій, що описують форму зубців P, Q, R, S, T , сегментів PQ , ST та елементів базової лінії:

$$z_0(t) = \begin{cases} z_0^{(1)}(t), & 0 \leq t < t_P^{(1)}, \\ \dots & \\ z_0^{(K)}(t), & t_T^{(2)} \leq t < T_0 \end{cases}, \quad (2.1)$$

де $t_P^{(1)}, t_P^{(2)}, \dots, t_T^{(1)}, t_T^{(2)}$ – моменти початку та закінчення відповідних зубців. За допомогою кусково-заданої функції (2.1) можна описати фактично будь-яку типову форму серцевого циклу нормальної або патологічної ЕКГ.

Припустимо, що у процесі породження m -го циклу реальної ЕКГ, $m = 1, 2, \dots$ внутрішнє збурення $\zeta(t)$ призводить до незалежних змін

тривалостей (областей визначення) окремих фрагментів корисного сигналу (еталону) $z_0^{(i)}(t)$, $i = 1, \dots, K$, а сама функція $z_0(t)$ лінійно розтягується (стискається) за амплітудою. Інакше кажучи, вважатимемо, що процес спотворення i -го фрагменту m -го циклу ЕКГ, що моделюється, здійснюється на основі операторного перетворення

$$z_m^{(i)}(t) = \lambda_m z_0^{(i)}\left(\frac{t - \tau_m^{(i)}}{\vartheta_m^{(i)}}\right), \quad m = 1, 2, \dots, i = 1, \dots, K, \quad (2.2)$$

де $\lambda_m, \vartheta_m^{(i)}$ – відповідно параметри лінійного розтягування (стиснення) за амплітудою та часом, а $\tau_m^{(i)}$ – зсув за часом.

Припустимо далі, що для кожного m -го циклу параметр λ_m набуває фіксованого значення

$$\lambda_m = 1 + \xi_m, \quad m = 1, 2, \dots, \quad (2.3)$$

де ξ_m – випадкова величина, яка з нульовим математичним сподіванням $M\{\xi\} = 0$ розподілена на інтервалі $[-\varepsilon_0, \varepsilon_0]$, обмеженому фіксованим числом $\varepsilon_0 \in [0, 1)$, а параметр $\vartheta_m^{(i)}$ набуває фіксованого значення у процесі породження кожного i -го фрагменту m -го циклу

$$\vartheta_m^{(i)} = 1 + \delta_m^{(i)}, \quad m = 1, 2, \dots, i = 1, \dots, K, \quad (2.4)$$

де $\delta_m^{(i)}$ – послідовність незалежних випадкових величин, що з нульовими математичними сподіваннями $M\{\delta\} = 0$ розподілені на інтервалах $[-\Delta_0^{(i)}, \Delta_0^{(i)}]$, обмеженими фіксованими числами $\Delta_0^{(i)} \in [0, 1)$.

Таким чином, вважатимемо, що сигнал, що спостерігається,

$$z(t) = \Phi[z_0(t), \zeta(t)] + h(t) \quad (2.5)$$

є послідовністю функцій $z_1(t), z_2(t), \dots, z_m(t), \dots$, пов'язаних з еталоном $z_0(t)$ операторним співвідношенням (2.2), у якому параметри λ_m і $\vartheta_m^{(i)}$ визначаються співвідношеннями (2.3) та (2.4).

Для забезпечення безперервності сигналу, що породжується, потрібно додатково, щоб $z_0(0) = 0$. Останню вимогу завжди можна забезпечити, виконавши центрування еталону $\check{z}_0(t) = z_0(t) - z_0(0)$.

За таких припущень тривалість i -го фрагменту m -го циклу пов'язана із тривалістю $\Delta t_0^{(i)} = t_i^{(2)} - t_i^{(1)}$ відповідного фрагмента еталону (2.1) співвідношенням

$$\Delta t_m^{(i)} = (t_i^{(2)} - t_i^{(1)})(1 + \delta_m^{(i)}),$$

а початку i -го фрагменту m -го циклу ЕКГ буде відповідати момент часу

$$t_m^{(i-1)} = \sum_{l=1}^{i-1} (t_l^{(2)} - t_l^{(1)})(1 + \delta_m^{(l)}). \quad (2.6)$$

Отже, загальна тривалість m -го циклу сигналу $z(t)$, що породжується, становить

$$T_m = T_0 + \sum_{i=1}^K (t_i^{(2)} - t_i^{(1)}) \delta_m^{(i)},$$

початку m -го циклу відповідає момент часу

$$t_m = (m-1)T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^K (t_i^{(2)} - t_i^{(1)}) \delta_j^{(i)},$$

а початку i -го фрагменту m -го циклу – момент часу

$$t_m^{(i-1)} = (m-1)T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^K (t_i^{(2)} - t_i^{(1)}) \delta_j^{(i)} + \sum_{l=1}^{i-1} (t_l^{(2)} - t_l^{(1)})(1 + \delta_m^{(l)}). \quad (2.7)$$

Якщо тепер до i -го фрагменту еталону $z_0(t)$ застосувати операторне перетворення (2.2), поклавши параметр зсуву

$$\tau_m^{(i)} = t_m^{(i-1)} - (1 + \delta_m^{(i)})t_i^{(1)}, \quad (2.8)$$

то за виразом (2.1) з урахуванням (2.2)–(2.7) процес породження i -го фрагменту m -го циклу ЕКГ можна подати функцією

$$z_m^{(i)}(t) = (1 + \xi_m)z_0^{(i)}(\theta) + h(t), m = 1, 2, \dots, i = 1, \dots, K, \quad (2.9)$$

аргумент якої θ нелінійно залежить від часу та визначається виразом

$$\theta = [t - (m-1)T_0 + (1 + \delta_m^{(i)})t_i^{(1)} - \sum_{j=1}^{m-1} \sum_{i=1}^K (t_i^{(2)} - t_i^{(1)})\delta_j^{(i)} -$$

$$- \sum_{l=1}^{i-1} (t_l^{(2)} - t_l^{(1)})(1 + \delta_m^{(l)})] (1 + \delta_m^{(i)})^{-1}. \quad (2.10)$$

Таким чином забезпечується моделювання нерівномірних за часом змін тривалостей окремих фрагментів еталону $z_0(t)$ на циклах ЕКГ, що породжується, що відповідає сучасним уявленням фізіологів (рис. 2.5).

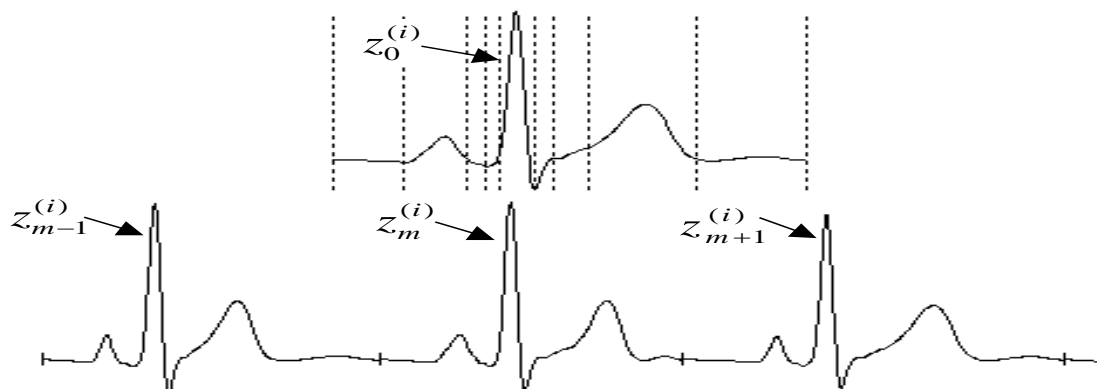


Рис. 2.5. Послідовність спотворених циклів, породжених одним еталоном $z_0(\cdot)$

Розглянута модель більш придатна для опису реальних ЕКГ, ніж її спрощений варіант

$$z_m^{(i)} = (1 + \xi_m) z_0^{(i)} [(t - (m-1)T_0 - T_0 \sum_{j=1}^{m-1} \delta_j)(1 + \delta_m)^{-1}] \quad m = 1, 2, \dots, \quad (2.11)$$

отриманий з (2.9) та (2.10) у припущенні, що випадковий параметр δ , що наявний у (2.10) залежить тільки від номера m циклу, але не залежить від номера i фрагмент.

Неважко показати, що стохастична модель (2.9), (2.10) є прямим узагальненням відомих математичних моделей строго періодичного та майже періодичного процесів. Дійсно, якщо покласти в (2.10)

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

то модель (2.10) можна подати у вигляді співвідношення

$$|z_m[t] - z_0[t - (m-1)T_0]| \leq \varepsilon_0, \quad m = 1, 2, \dots,$$

яке описує майже періодичний процес, а за додаткової умови $\varepsilon_0 = 0$ модель зводиться до строго періодичної функції

$$z[t - (m - 1)T_0] = z_0[t].$$

Запропоновану модель можна узагальнити для опису процесу породження ЕКГ із змінною морфологією окремих циклів (екстрасистолами).

Для цього достатньо передбачити, що є $G > 1$ еталонів $z_{01}(t), \dots, z_{0G}(t)$ і припустити, що кожен m -й цикл породжується за допомогою аналогічних спотворень одного з цих еталонів, що вибираються випадково відповідно до ймовірностей P_g , $\sum_{g=1}^G P_g = 1$ (рис. 2.6).

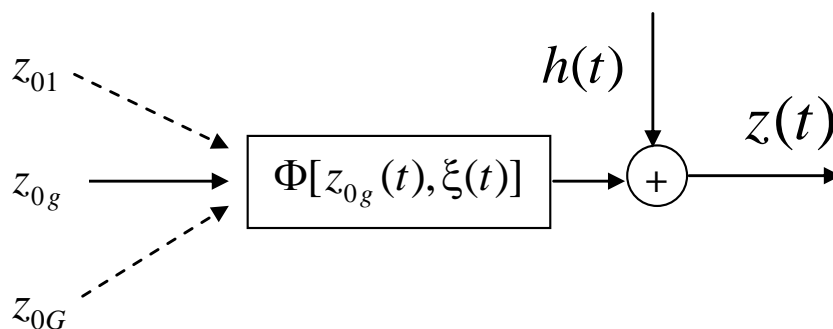


Рис. 2.6. Механізм породження ЕКГ за кількома еталонами

Розглянемо програмний генератор штучної ЕКГ за заданими еталонами у припущенні, що зовнішнє збурення відсутнє, тобто $h(t) \equiv 0$.

Припустимо, що кожен з $G > 1$ еталонів $z_{0g}(t)$, $g = 1, \dots, G$ є кінцевим числом N_g дискретних значень $z_{0g}[1], \dots, z_{0g}[N_g]$ з постійним кроком квантування за часом. Задамо також загальну кількість $K^{(g)}$ фрагментів кожного зразка та номери точок $N_g^{(i)} \in [1, N_g]$, $i = 1, \dots, K^{(g)} - 1$, які визначають межі i -го та $i + 1$ -го фрагментів g -го зразка.

Тоді процедура генерації штучної ЕКГ зводиться до таких кроків:
Крок 1. Задаємо число M циклів генерованої послідовності.

Крок 2. Визначаємо число M_g циклів, що породжуються g -м еталоном, за формулою $i < K^{(g)}$, де $R[\eta]$ – операція округлення до цілого числа η .

Крок 3. Обираємо номер g_m зразка, що породжує m -й цикл ($m = 1, \dots, M$), за значенням реалізації ρ_m цілої випадкової величини ρ_m , що розподілена на інтервалі $[1, G]$, тобто $g_m = \rho_m$.

Крок 4. Якщо $M_g = 0$, то повторюємо крок 3.

Крок 5. Визначаємо кількість точок i -го фрагмента m -го циклу за формулою

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

де $\delta_m^{(i)}$ – реалізація випадкової величини δ , яку з нульовим математичним сподіванням розподілено на інтервалі $[-\Delta_0, \Delta_0]$, $\Delta_0 \in [0, 1)$.

Крок 6. За дискретними значеннями i -го фрагмента g -го еталону в $N_g^{(i)} - N_g^{(i-1)}$ вузлах будь-яким з методів інтерполяції обчислюємо значення $z[\cdot]$ генерованої послідовності в $N_m^{(i)}$ точках.

Крок 7. Модифікуємо кожне обчислене значення $z[\cdot]$ на основі мультиплікативної процедури $\tilde{z}[\cdot] = (1 + \xi_m)z[\cdot]$, де ξ_m – реалізація випадкової величини ξ , яку з нульовим математичним сподіванням розподілено на інтервалі $[-\varepsilon_0, \varepsilon_0]$, $\varepsilon_0 \in [0, 1)$.

Крок 8. Якщо $i < K^{(g)}$, то повертаємось до кроку 5.

Крок 9. Привласнюємо $M_g := M_g - 1$.

Крок 10. Якщо $m < M$, то повертаємось до кроку 3.

На рис 2.7 наведено приклади штучних ЕКГ, генерованих за описаним вище алгоритмом.



Рис. 2.7. Штучні ЕКГ, породжені за одним (а) та двома (б) еталонами

Досвідчені кардіологи, яким показували штучні ЕКГ, що згенерували таким чином, визнавали їх реальними, що підтверджує адекватність моделей.

У разі необхідності імітувати зовнішні збурення $h(t) \neq 0$ достатньо до кожного дискретного значення $\tilde{z}[\cdot]$ генерованої послідовності додати дискретне значення $h[\cdot]$, обчислене відповідно до моделі адитивної перешкоди.

Хоча розглянутий підхід дозволяє синтезувати послідовність спотворених циклів ЕКГ реалістичної форми, запропоновану модель не можна повною мірою зарахувати до генеративних моделей.

До того ж, ця модель не надає можливість генерувати сигнал під час незалежних спотворень амплітуд окремих фрагментів еталона. Як буде показано далі, такі особливості інтерполяційної моделі обмежують її властивості для генерації штучних ЕКГ, що відповідають серцевим порушенням, зокрема, ефекту електричної альтернації серця.

Тому розглянемо інший підхід генерації штучних ЕКГ реалістичної форми.

2.3. Генеративна модель породження ЕКГ

Модель ґрунтується на механізмі, що описує динаміку серцевого ритму, який оснований на русі зображувальної точки (рис. 2.8) у тривимірному просторі координат (x, y, z) .

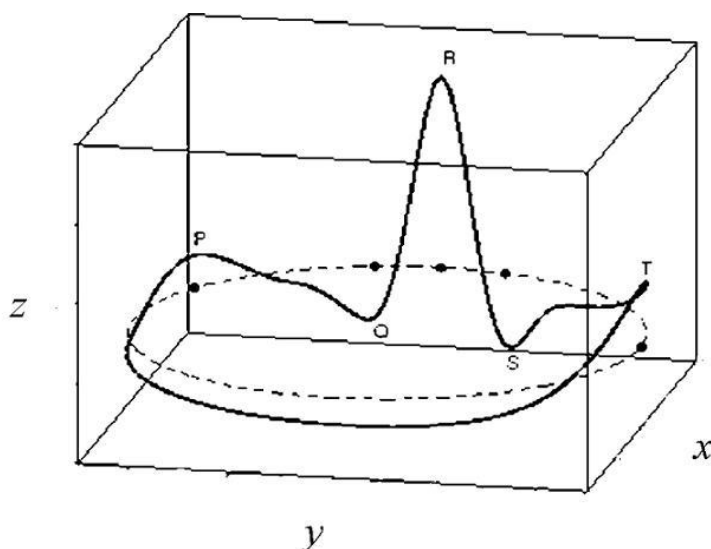


Рис. 2.8. Механізм генерації ЕКГ

Циклічність ЕКГ моделюється рухом зображувальної точки у площині (x, y) за траєкторією змінної довжини, яка «розмита» щодо кола одиничного радіусу. Кожен оберт вздовж цього кола відповідає одному серцевому циклу (RR -інтервалу).

Інформативні фрагменти ЕКГ моделюються рухом зображувальної точки в напрямку z . Моменти часу t , за яких з'являються зубці P, Q, R, S, T , визначають точки, що відповідають кутам $\theta_P, \theta_Q, \theta_R, \theta_S, \theta_T$. Коли зображувальна точка наближається до цих кутів, вона рухається вгору або вниз у напрямку z , а через певний час повертається до площини (x, y) .

Формально рух зображувальної точки описується системою з трьох звичайних диференціальних рівнянь

$$\dot{x} = \alpha x - \omega y, \quad (2.12)$$

$$\dot{y} = \alpha y + \omega x, \quad (2.13)$$

$$\dot{z} = - \sum_{i \in \{P, Q, R, S, T\}} a_i \Delta \theta_i \exp\left(-\frac{\Delta \theta_i^2}{2b_i^2}\right) - (z - z_B). \quad (2.14)$$

У системі (2.12)–(2.14) взято такі позначення: $\alpha = 1 - \sqrt{x^2 + y^2}$, ω – кутова швидкість руху точки в площині (x, y) , $\Delta \theta_i = (\theta - \theta_i) \bmod 2\pi$, $i \in \{P, Q, R, S, T\}$, де $\theta \in [-\pi, \pi]$ – поточний кут, який визначається виразом

$$\theta = \begin{cases} \operatorname{arctg}\left(\frac{y}{x}\right), & \text{якщо } x > 0; \\ \pi + \operatorname{arctg}\left(\frac{y}{x}\right), & \text{якщо } y \geq 0, x < 0; \\ -\pi + \operatorname{arctg}\left(\frac{y}{x}\right), & \text{якщо } y < 0, x < 0; \\ \frac{\pi}{2}, & \text{якщо } y > 0, x = 0; \\ -\frac{\pi}{2}, & \text{якщо } y < 0, x = 0. \end{cases}$$

У рівнянні (2.14) змінна $z_B = B \cdot \sin(2\pi f_B t)$ моделює дрейф базової лінії з амплітудою B та частотою дихання f_B . Параметри a_i , b_i , θ_i визначають амплітудно-часові характеристики зубців P , Q , R , S , T генерованої ЕКГ.

На основі чисельного розв'язання системи рівнянь (2.12)–(2.14) методом Рунге-Кутта четвертого порядку можна побудувати низку дискретних значень $z(k) \equiv z(t_k)$ у рівновіддалені моменти часу $t_k \equiv k\Delta$, $k = 1, 2, \dots$, що породжують штучну ЕКГ реалістичної форми.

Враховуючи вищесказане, виявилось, що можна знайти аналітичний вираз для обчислення функції $z = z(t)$. Для цього із розв'язання рівнянь (2.12), (2.13) знайдемо поточне значення кута θ , що використано в рівнянні (2.14), та запишемо (2.12), (2.13) у полярних координатах

$$\dot{r} = (1 - r)r, \quad (2.15)$$

$$\dot{\theta} = \omega. \quad (2.16)$$

З рівняння (2.16) випливає, що якщо кутова швидкість ω є кусково-постійною величиною, тобто, якщо на кожному окремому циклі $\omega = \text{const}$, то для моменту часу t , що належить цьому циклу, поточний кут θ визначається виразом

$$\theta = \omega t + \theta_0. \quad (2.17)$$

Запишемо функцію z , що перебуває у правій частині рівняння (2.14), у вигляді суми функцій

$$z = \sum_{i \in \{P, Q, R, S, T\}} z_i + z_1, \quad (2.18)$$

де z_i – функції, що описують форму зубців P, Q, R, S, T , а z_1 – коригувальна домішка.

Підставляючи z з (2.18) до (2.14), отримаємо

$$\dot{z} = - \sum_{i \in \{P, Q, R, S, T\}} a_i \Delta \theta_i \exp\left(-\frac{\Delta \theta_i^2}{2b_i^2}\right) - \sum_{i \in \{P, Q, R, S, T\}} z_i - z_1 + z_B.$$

Звідси випливає, що розв’язок рівняння (2.14) еквівалентне розв’язку такої системи:

$$\dot{z}_i = -a_i \Delta \theta_i \exp\left(-\frac{\Delta \theta_i^2}{2b_i^2}\right) - z_i, \quad i \in \{P, Q, R, S, T\}, \quad (2.19)$$

$$\dot{z}_1 = -z_1 + z_B. \quad (2.20)$$

З урахуванням (2.17) розв’язок системи лінійних неоднорідних рівнянь (2.19), (2.20) має вигляд:

$$z_i = \frac{a_i b_i^2}{\omega} \exp\left(-\frac{1}{2} \frac{\Delta \theta_i^2}{b_i^2}\right) - \sqrt{2\pi} \frac{a_i b_i^3}{\omega^2} \exp\left(-\frac{\Delta \theta_i}{\omega} + \frac{b_i^2}{2\omega^2}\right) \Phi\left(\frac{\Delta \theta_i}{b_i} - \frac{b_i}{\omega}\right) + C_0 \exp\left(-\frac{\Delta \theta_i}{\omega}\right), \quad (2.21)$$

$$z_1 = C_0 e^{-t} + \frac{B}{(2\pi f_B)^2 + 1} [\sin(2\pi f_B t) - 2\pi f_B \cos(2\pi f_B t)], \quad (2.22)$$

де $\Phi(\cdot)$ – функція нормального розподілу $N(0,1)$, а $C_0 = \text{const}$.

Таким чином, остаточний розв’язок рівняння (2.14) можна записати так:

$$z = \sum_{i \in \{P, Q, R, S, T\}} \left[\frac{a_i b_i^2}{\omega} \exp\left(-\frac{\Delta \theta_i^2}{2b_i^2}\right) - \sqrt{2\pi} \frac{a_i b_i^3}{\omega^2} \times \right.$$

$$\begin{aligned} & \times \exp\left(-\frac{\Delta\theta_i}{\omega} + \frac{b_i^2}{2\omega^2}\right) \Phi\left(\frac{\Delta\theta_i}{b_i} - \frac{b_i}{\omega}\right) \Big] + C_0 \exp\left(-\frac{\Delta\theta_i}{\omega}\right) + \\ & + C_0 e^{-t} + \frac{B}{(2\pi f_B)^2 + 1} [\sin(2\pi f_B t) - 2\pi f_B \cos(2\pi f_B t)]. \end{aligned} \quad (2.23)$$

Аналітичний вираз (2.23) повною мірою можна назвати генеративною моделлю породження штучної ЕКГ реалістичної форми. Однак, на відміну від інтерполяційної моделі, така модель надає змогу лише генерувати однакові розтягування (стиснення) у часі всіх інформативних фрагментів, що обмежує сферу її застосування. Така модель прийнятна лише для дослідження закономірностей варіабельності серцевого ритму.

Тому розглянемо інший підхід до моделювання штучної ЕКГ. Суть цього підходу полягає в тому, щоб на першому етапі згенерувати еталонний кардіоцикл $z_0(t)$ із заданими амплітудно-часовими параметрами інформативних фрагментів, а потім на основі $z_0(t)$ породжувати послідовність серцевих циклів, незалежно спотворюючи ці параметри випадковим чином.

З цією метою спростимо диференціальне рівняння (2.14), виключивши з нього коригуючу домішку $z_1 = z - z_B$, яка моделює дрейф базової лінії під впливом дихальних хвиль, і запишемо це рівняння у вигляді:

$$\dot{z} = - \sum_{i \in \{P, Q, R, S, T\}} a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right). \quad (2.24)$$

Неважко переконатися, що один з розв'язків рівняння (2.24) має вигляд

$$z = \sum_{i \in \{P, Q, R, S, T\}} \frac{a_i b_i^2}{\omega} \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right). \quad (2.25)$$

Запишемо співвідношення (2.25) у дещо іншій формі

$$z_0(t) = \sum_{i \in \{P, Q, R, S, T\}} A_i \exp\left(-\frac{(t - \mu_i)^2}{2b_i^2}\right), \quad (2.26)$$

Де A_i , μ_i , b_i – параметри, що визначають відповідно амплітуду, момент досягнення максимуму та тривалість i -го фрагменту.

Щоб забезпечити можливість генерації еталонного циклу з відхиленнями сегмента ST щодо ізоелектричної лінії та несиметричними зубцями, дещо розширимо модель $z_0(t)$:

$$z_0(t) = \sum_{i \in \{P, Q, R, S, ST, T\}} A_i \exp\left(-\frac{(t - \mu_i)^2}{2[b_i(t)]^2}\right), \quad (2.27)$$

де

$$b_i(t) = \begin{cases} b_i^{(1)}, & \text{якщо } t \leq \mu_i, \\ b_i^{(2)}, & \text{якщо } t > \mu_i, \end{cases} \quad (2.28)$$

а $b_i^{(1)}$, $b_i^{(2)}$ – параметри, що визначають симетрію i -го фрагмента.

Використовуючи в (2.27) функцій (2.28), можна згенерувати несиметричні фрагменти еталонного циклу, зокрема, породжувати зубець T з різним значенням параметра симетрії, який, як буде показано далі, має додаткову діагностичну цінність.

Твердження 2.1. Незважаючи на те, що функції $b_i(t)$, які використано у (2.27), відповідно до визначення (2.28) не є неперервними, коли $b_i^{(1)} \neq b_i^{(2)}$, функція $z_0(t)$ та її перша похідна $\dot{z}_0(t)$ неперервні у всіх точках області визначення, зокрема, в точках $t = \mu_i$ розриву функцій $b_i(t)$.

Доведення. Розглянемо одну із складових виразу (2.27), а саме функцію

$$f_i(t) = \exp\left(-\frac{(t - \mu_i)^2}{2[b_i(t)]^2}\right),$$

в якій $b_i(t)$ визначається співвідношенням (2.28). Легко побачити, що $f_i(\mu_i) = e^0 = 1$.

Знайдемо границі функції $f_i(t)$ зліва та справа

$$f_i(\mu_i - 0) = \lim_{t \rightarrow \mu_i, t < \mu_i} f_i(t) = \lim_{t \rightarrow \mu_i} \exp\left(-\frac{(t - \mu_i)^2}{2[b_i^{(1)}]^2}\right) = e^0 = 1,$$

$$f_i(\mu_i + 0) = \lim_{t \rightarrow \mu_i, t > \mu_i} f_i(t) = \lim_{t \rightarrow \mu_i} \exp\left(-\frac{(t - \mu_i)^2}{2[b_i^{(2)}]^2}\right) = e^0 = 1.$$

Отже, ліва та права границі функції $f_i(t)$ рівні між собою та дорівнюють значенню $f_i(t)$ у точці $t = \mu_i$, що підтверджує неперервність $f_i(t)$ у цій точці.

Знайдемо першу похідну функції $f_i(t)$:

$$\dot{f}_i(t) = -\exp\left(-\frac{(t-\mu_i)^2}{2[b_i(t)]^2}\right) \cdot \frac{(t-\mu_i)[b_i(t)]^2 - (t-\mu_i)^2 b_i(t)\dot{b}_i(t)}{[b_i(t)]^4}.$$

Оскільки функція $b_i(t)$ не диференційована в точці $t = \mu_i$, то для знаходження похідної у цій точці скористаємося означенням

$$\dot{f}_i(t) = \lim_{\Delta t \rightarrow 0} \frac{f_i(t + \Delta t) - f_i(t)}{\Delta t}, \text{ якщо границя існує.}$$

Щоб переконатися в наявності границі $\lim_{\Delta t \rightarrow 0} \frac{f_i(t + \Delta t) - f_i(t)}{\Delta t}$ у точці $t = \mu_i$, знайдемо односторонні границі:

$$\lim_{\Delta t \rightarrow 0, \Delta t < 0} \frac{f_i(\mu_i + \Delta t) - f_i(\mu_i)}{\Delta t} = \lim_{\Delta t \rightarrow 0, \Delta t < 0} \left[-\frac{\Delta t}{[b_i^{(1)}]^2} \cdot \exp\left(-\frac{(\Delta t)^2}{2[b_i^{(1)}]^2}\right) \right] = 0.$$

Аналогічно

$$\lim_{\Delta t \rightarrow 0, \Delta t > 0} \frac{f_i(\mu_i + \Delta t) - f_i(\mu_i)}{\Delta t} = \lim_{\Delta t \rightarrow 0, \Delta t > 0} \left[-\frac{\Delta t}{[b_i^{(2)}]^2} \cdot \exp\left(-\frac{(\Delta t)^2}{2[b_i^{(2)}]^2}\right) \right] = 0.$$

Оскільки границя не залежить від знака Δt , то вона існує, а значить, похідна функції $f_i(t)$ у точці $t = \mu_i$ існує і дорівнює нулю.

Таким чином, ми показали, що

$$\lim_{\Delta t \rightarrow 0, \Delta t < 0} \frac{f_i(\mu_i + \Delta t) - f_i(\mu_i)}{\Delta t} = \lim_{\Delta t \rightarrow 0, \Delta t > 0} \frac{f_i(\mu_i + \Delta t) - f_i(\mu_i)}{\Delta t} = 0,$$

тобто ліва та права границі похідної функції $f_i(t)$ у точці $t = \mu_i$ збігаються зі значенням похідної у цій точці.

Отже, для всіх $i \in \{P, Q, R, S, ST, T\}$ функції $f_i(t)$ і похідні $\dot{f}_i(t)$ неперервні в точках $t = \mu_i$, що підтверджує неперервність $z_0(t)$ та $\dot{z}_0(t)$ у всіх точках області визначення. Твердження 2.1 доведено.

Оскільки гаусова функція головним чином зосереджена на «три-сигмовому» інтервалі, можна вважати, що моменти початку $t_i^{(1)}$ та закінчення $t_i^{(2)}$ кожного i -го фрагмента, $i \in \{P, Q, R, S, ST, T\}$ пов'язані з параметрами $b_i^{(1)}, b_i^{(2)}$ і μ_i таким чином:

$$t_i^{(1)} = \mu_i - 3b_i^{(1)},$$

$$t_i^{(2)} = \mu_i + 3b_i^{(2)}.$$

Зауважимо, що допустима структура еталонного циклу накладає такі обмеження на моменти часу $t_i^{(1)}, t_i^{(2)}$:

$$0 \leq t_P^{(1)} < t_P^{(2)} \leq t_Q^{(1)} < t_Q^{(2)} = t_R^{(1)} < t_R^{(2)} = t_S^{(1)} < t_S^{(2)} = t_{ST}^{(1)} \leq t_{ST}^{(2)} \leq t_T^{(1)} < t_T^{(2)} \leq T_0,$$

де T_0 (мс) – загальна тривалість еталонного циклу, пов'язана із частотою F_H серцевих скорочень (уд/хв) співвідношенням

$$T_0 = \frac{60 \cdot 1000}{F_H}.$$

Змінюючи значення параметрів $A_i, \mu_i, b_i^{(1)}, b_i^{(2)}$ моделі (2.27), можна згенерувати еталонні цикли $z_0(t)$ найрізноманітнішої форми, зокрема, зі зсувом сегмента ST нижче (депресія) або вище (елевація) ізоелектричної лінії, інверсією зубця T та іншими електрокардіографічними ознаками серцевих патологій (рис. 2.9).

За сформованим еталоном $z_0(t)$ необхідної форми можна генерувати штучну ЕКГ на основі розглянутої вище інтерполяційної моделі (2.9), (2.10). Однак можна використати інший спосіб. Якщо в нашому розпорядженні тепер вже є аналітичний вираз (2.27), за яким генерується еталон $z_0(t)$, можна породжувати спотворені цикли штучної ЕКГ через незалежні випадкові спотворення самих параметрів $A_i, \mu_i, b_i^{(1)}, b_i^{(2)}$, що наявні у (2.27).

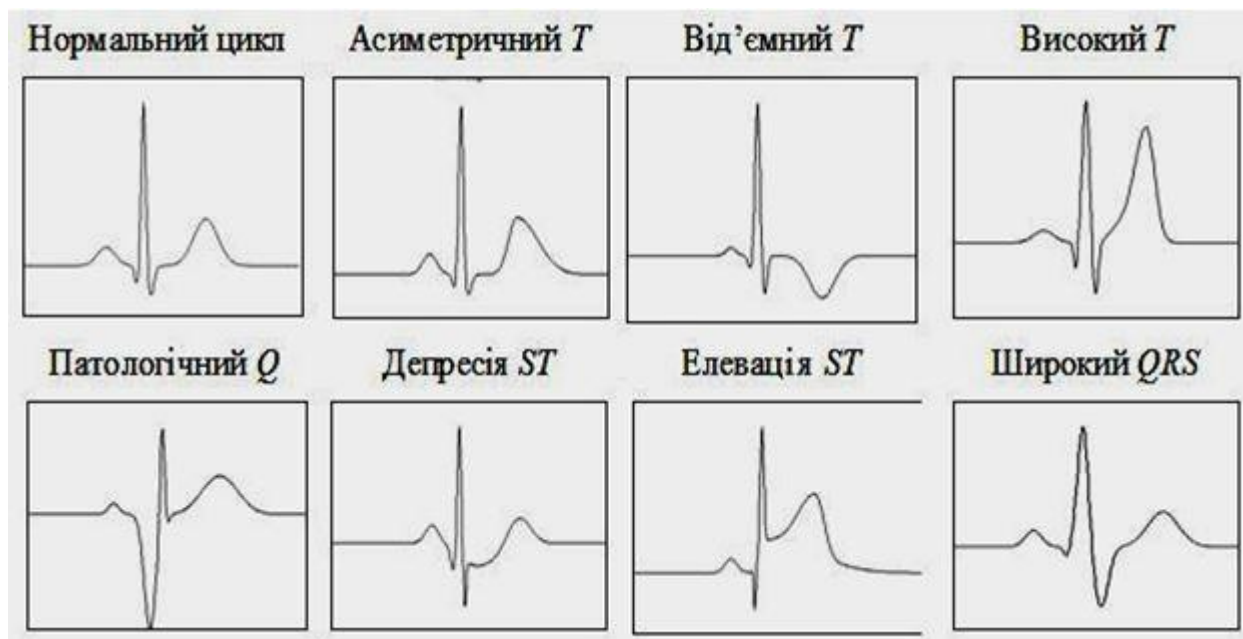


Рис. 2.9. Приклади штучних циклів ЕКГ різної форми

Таким способом вдається генерувати послідовність спотворених циклів, у яких нерівномірно змінюються тривалості, амплітуди і навіть симетричність окремих фрагментів.

Формально така модель генерації штучної ЕКГ має вигляд:

$$z_m(t) = \sum_{i \in \{P, Q, R, S, ST, T\}} \tilde{A}_{im} \exp\left[-\frac{(t - \tilde{\mu}_{im})^2}{2\tilde{b}_{im}^2}\right] + h(t), \quad m = 1, 2, \dots, M, \quad (2.29)$$

Де M – кількість циклів штучної ЕКГ, а

$$\tilde{A}_{im} = A_i(1 + \alpha_{im}), \quad (2.30)$$

$$\tilde{\mu}_{im} = \mu_i(1 + \delta_{im}), \quad (2.31)$$

$$\tilde{b}_{im} = \begin{cases} b_i^{(1)}(1 + \varepsilon_{im}^{(1)}) & \forall t \leq \mu_{im}, \\ b_i^{(2)}(1 + \varepsilon_{im}^{(2)}) & \forall t > \mu_{im}. \end{cases} \quad (2.32)$$

Відповідно до (2.29)–(2.32) моделюються спотворення параметрів A_i , μ_i , $b_i^{(1)}$, $b_i^{(2)}$ еталону на кожному m циклі штучної ЕКГ за допомогою реалізацій незалежних випадкових величин α_{im} , δ_{im} , $\varepsilon_{im}^{(1)}$, $\varepsilon_{im}^{(2)}$, які з нульовими математичними сподіваннями розподілені на обмежених інтервалах:

$$\alpha_{im} \in [-\alpha_i^0, \alpha_i^0], \delta_{im} \in [\delta_i^0, -\delta_i^0], \varepsilon_{im}^{(1)} \in [-\varepsilon_i^0, \varepsilon_i^0], \varepsilon_{im}^{(2)} \in [-\varepsilon_i^0, \varepsilon_i^0], \quad (2.33)$$

де $\alpha_i^0, \delta_i^0, \varepsilon_i^0$ – фіксовані числа.

Функція

$$h(t) = h_G(t) + h_R(t) + h_D(t), \quad (2.34)$$

що перебуває в правій частині (2.29), моделює зовнішні адитивні збурення різного типу. Мережеві завади моделюються полігармонічною функцією $h_G(t)$ із заданими амплітудами та частотами, м'язовий тремор – обмеженою за рівнем випадковою перешкодою $h_R(t)$ із заданим законом розподілу, а дрейф ізоелектричної лінії – низькочастотною функцією $h_D(t)$ бажаного рівня.

Для ілюстрації на рис. 2.10 наведено штучну ЕКГ, генеровану на основі (2.29)–(2.34) з такими параметрами:

$$A_P = 0,08; A_Q = -0,227; A_R = 0,72; A_S = -0,18; A_{ST} = 0; A_T = -0,2;$$

$$b_P^{(1)} = 0,016; b_Q^{(1)} = 0,01; b_R^{(1)} = 0,01; b_S^{(1)} = 0,015; b_{ST}^{(1)} = 0,04; b_T^{(1)} = 0,05;$$

$$b_P^{(2)} = 0,016; b_Q^{(2)} = 0,01; b_R^{(2)} = 0,01; b_S^{(2)} = 0,015; b_{ST}^{(2)} = 0,04; b_T^{(2)} = 0,023;$$

$$\mu_P = 0,18; \mu_Q = 0,25; \mu_R = 0,28; \mu_S = 0,315; \mu_{ST} = 0,38; \mu_T = 0,531;$$

з урахуванням обмежень на внутрішні спотворення

$$\alpha_P^0 = 5\%; \alpha_Q^0 = 15\%; \alpha_R^0 = 20\%; \alpha_S^0 = 15\%; \alpha_{ST}^0 = 5\%; \alpha_T^0 = 7\%;$$

$$\varepsilon_P^0 = 2\%; \varepsilon_T^0 = 2\%; \delta_T^0 = 20\%; \gamma_0 = 7\%;$$

випадкових спотворень $h_R(t)$ з амплітудою 2 % та дрейфом ізоелектричної лінії $h_D(t)$ з амплітудою 82 % щодо діапазону зміни корисного сигналу.

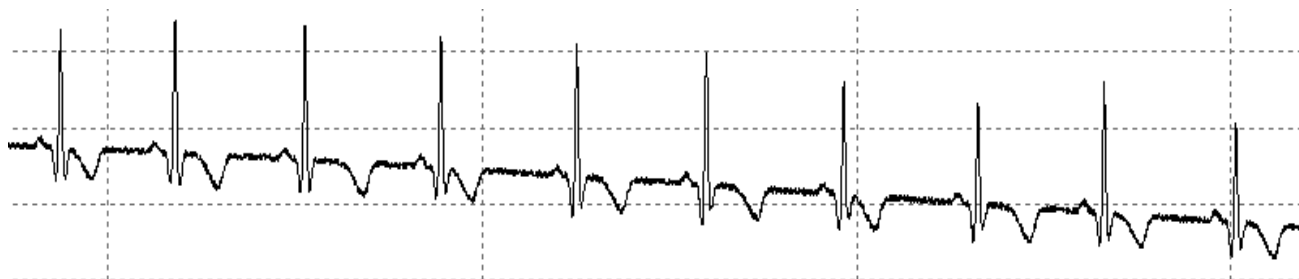


Рис. 2.10. Штучна ЕКГ з інверсією зубця T на тлі тренда ізоелектричної лінії

Для генерації штучних ЕКГ з нетиповими циклами (екстрасистолами та артефактами) пропонується випадковим чином сформувати дві підмножини упорядкованих індексів

$$I_E \subset \{1, \dots, M\} \text{ і } I_F \subset \{1, \dots, M\}, \quad I_E \cap I_F = \emptyset$$

та перейти від (2.29)–(2.34) до вдосконаленої моделі

$$z_m(t) = \begin{cases} \sum_{i \in \{P, Q, R, S, ST, T\}} \tilde{A}_{im} \exp\left(-\frac{(t - \tilde{\mu}_{im})^2}{2\tilde{b}_{im}^2}\right) + h(t), & \forall m \in \{1, \dots, M\} / (I_E \cup I_F), \\ z_E(t) + h(t), & \forall m \in I_E, \\ z_F(t) + h(t), & \forall m \in I_F. \end{cases} \quad (2.35)$$

$$\tilde{A}_{im} = A_i(1 + \alpha_{im}), \quad (2.36)$$

$$\tilde{\mu}_{im} = \mu_i(1 + \delta_{im}), \quad (2.37)$$

$$\tilde{b}_{im} = \begin{cases} b_i^{(1)}(1 + \varepsilon_{im}^{(1)}) & \forall t \leq \mu_{im}, \\ b_i^{(2)}(1 + \varepsilon_{im}^{(2)}) & \forall t > \mu_{im}. \end{cases} \quad (2.38)$$

Функція $z_E(t) \in Z_E$ імітує екстрасистоли, яка випадково вибирається з кінцевої множини заздалегідь генерованих функцій, а $z_F(t) \in Z_F$ – функція, що імітує артефакт, яка випадково вибирається з іншої кінцевої множини $Z_F = \{Z_F^{(1)}(t), \dots, Z_F^{(K_F)}(t)\}$ заданих реальних чи модельних функцій (рис. 2.11).

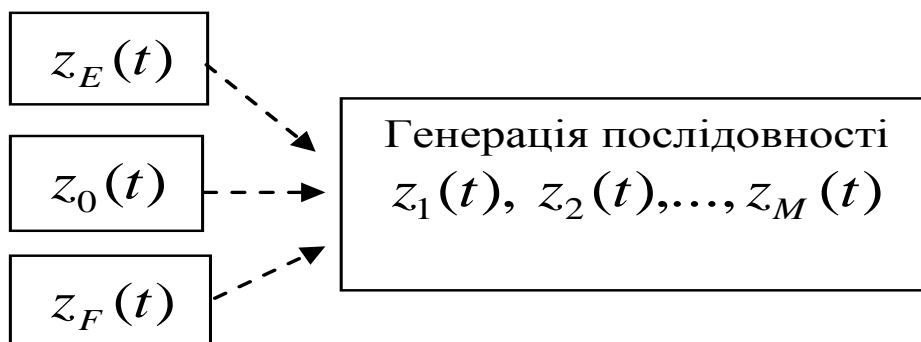


Рис. 2.11. Механізм генерації штучної ЕКГ із нетиповими циклами

На рис. 2.12 показано штучну ЕКГ реалістичної форми, що генерована за моделлю (2.35)–(2.38) за еталоном $z_0(t)$, параметри якого мали значення

$$A_P = 0,11; A_Q = -0,11; A_R = 1; A_S = -0,18; A_{ST} = 0; A_T = 0,2;$$

$$b_P^{(1)} = 0,025; b_Q^{(1)} = 0,025; b_R^{(1)} = 0,025; b_S^{(1)} = 0,015; b_{ST}^{(1)} = 0,04; b_T^{(1)} = 0,05;$$

$$b_P^{(2)} = 0,025; b_Q^{(2)} = 0,025; b_R^{(2)} = 0,025; b_S^{(2)} = 0,015; b_{ST}^{(2)} = 0,04; b_T^{(2)} = 0,03;$$

$$\mu_P = 0,399; \mu_Q = 0,47; \mu_R = 0,499; \mu_S = 0,534; \mu_{ST} = 0,6; \mu_T = 0,7;$$

з обмеженнями на внутрішні спотворення

$$\alpha_P^0 = 5\%, \alpha_R^0 = 10\%, \alpha_S^0 = 5\%, \alpha_T^0 = 15\%, \varepsilon_T^0 = 5\%, \delta_T^0 = 10\%,$$

зовнішніми спотвореннями $h_R(t)$ з амплітудою 7 % та дрейфом ізоелектричної лінії $h_D(t)$ з амплітудою 85 %.



Рис. 2.12. Штучна ЕКГ з імітацією екстрасистоли (1) та артефакту (2)

2.4. Генерація ЕКГ з альтернацією зубця T

Незначні модернізації генеративної моделі (2.35) дозволяють побудувати модель імітації електричної альтернативи зубця T , яку лікарі останнім часом вважають одним з предикторів загрозливих аритмій та раптової серцевої смерті.

Раптовою серцевою смертю вважають ситуацію, коли хворий гине фактично миттєво (від кількох секунд до години) після початку серцевого нападу.

Ефект електричної альтернативи іноді ототожнюється з мікроколиваннями (дисперсією) ЕКГ, що не відповідає початковому тлума-

ченню цього терміна. Насправді термін «альтернація» запозичений із лінгвістики й означає закономірне чергування звуків у словах із загальною основою.

Альтернацією елементів ЕКГ називають числовий параметр, який характеризує своєрідні чергування амплітудно-часових характеристик зубців P, Q, R, S, T , наприклад, зниження амплітуди зубців T на послідовних серцевих циклах (рис. 2.13).

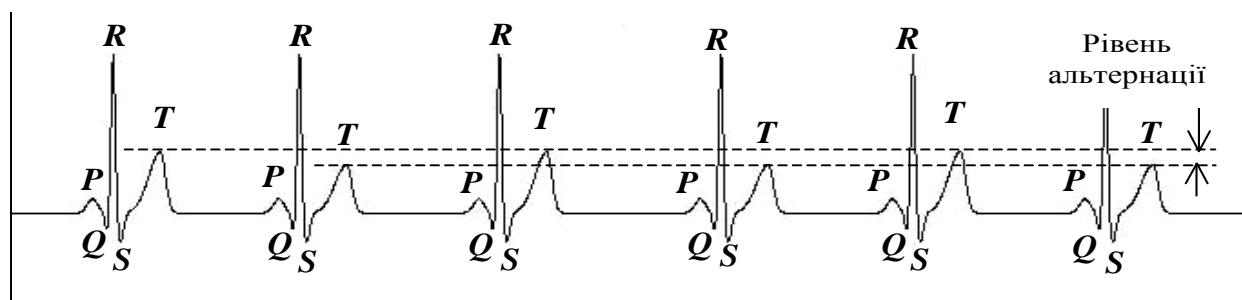


Рис. 2.13. Альтернація амплітуд зубця T

Мінімальний рівень альтернації зубця T , що підлягає аналізу, становить всього 5–10 мкВ. Однак на реальних ЕКГ унаслідок випадкових збурень амплітуда зубця T може зазнавати значно більших змін від циклу до циклу навіть за відсутності альтернативи.

У клінічній практиці трапляються не тільки найпростіший тип альтернації, показаний на рис. 2.13, а й складніші типи. Зокрема, альтернатива може виявлятися у чергуванні зубців T з різним ступенем симетрії чи різною тривалістю. Можливі також поєднання на одній ЕКГ різних типів альтернації.

На реальних ЕКГ може спостерігатися чітке чергування параметрів, коли на парних циклах спостерігаються зубці одного типу A , а на непарних – зубці іншого типу B , тобто припускається, що послідовність циклів можна умовно позначити «словом» $A B A B A B \dots$ з «буквами» A і B , що чітко чергуються (рис 2.14, а).

Водночас фаза чергування зубців може змінюватися у випадкові моменти часу, коли зубці одного типу трапляються на двох послідовних циклах (рис. 2.14, б).



Рис. 2.14. Чітке чергування зубців двох типів (а) та альтернація зі зміною фази чергування зубців (б)

Щоб побудувати математичну модель генерації ЕКГ з альтернаціями різного типу, сформулюємо такі основні гіпотези.

Гіпотеза 2.1. За відсутності альтернації амплітуда зубця T є реалізацією однієї випадкової величини Y , яка з обмеженою дисперсією змінюється щодо деякого середнього значення $M\{Y\}$.

Гіпотеза 2.2. За наявності альтернації амплітуди зубців T є реалізації двох випадкових величин Y_A і Y_B , які з обмеженою дисперсією змінюються щодо «своїх» середніх значень $M\{Y_A\} \neq M\{Y_B\}$, причому

$$Y_A \in [Y_A^{\min}, Y_A^{\max}], Y_B \in [Y_B^{\min}, Y_B^{\max}] \quad (2.39)$$

і

$$Y_A^{\max} < Y_B^{\min} \text{ або } Y_B^{\max} < Y_A^{\min}. \quad (2.40)$$

Гіпотеза 2.3. За наявності альтернації у випадкові моменти часу t_1, t_2, \dots, t_M порядок чергування зубців двох типів може бути порушено, проте число L таких моментів набагато менше загальної кількості M аналізованих циклів, причому

$$|t_i - t_j| > T_A \quad \forall i, j \in \{1, \dots, L\}, i \neq j, \quad (2.41)$$

де T_A – мінімальний інтервал часу, достатній для оцінювання альтернативи, який зазвичай відповідає не менше ніж 16 циклів.

Дотримуючись цих гіпотез, згенеруємо штучну ЕКГ на основі моделі (2.35), але для визначення параметрів \tilde{A}_{im} , що наявні у цій моделі, замість (2.36), скористаємося виразом

$$\tilde{A}_{im} = \begin{cases} A_i(1 + \alpha_{im}) & \forall i \in \{P, Q, R, S, ST\}, \\ A_T \lambda_m^{(A)} (1 + \alpha_{Tm}) & \text{коли } i = T, \end{cases} \quad (2.42)$$

в якому

$$\lambda_m^{(A)} = \begin{cases} 1 + \frac{\Delta_A}{A_T}, & \text{якщо } \lambda_{m-1}^{(T)} = 1, \\ 1, & \text{якщо } \lambda_{m-1}^T = 1 + \frac{\Delta_A}{A_T}, \end{cases} \quad (2.43)$$

за початкової умови $\lambda_0^{(A)} = 1$, при цьому параметри $\tilde{\mu}_{im}$, \tilde{b}_{im} , як і раніше, будемо обчислювати за формулами (2.37), (2.38).

З виразів (2.42), (2.43) з урахуванням (2.33) безпосередньо випливає, що для виконання обмежень (2.39), (2.40) величина Δ_A , яка визначає заданий рівень альтернативи амплітуд зубців T , має задовольняти умову

$$\Delta_A > 2\alpha_T^0 A_T, \quad (2.44)$$

де α_T^0 – обмеження на рівень випадкових спотворень α_{Tm} амплітуд зубців T .

За допомогою моделі (2.35) та співвідношень (2.37), (2.38), (2.42), (2.43), за якими обчислюються параметри спотворених циклів, генерується штучна ЕКГ з чіткою альтернативною амплітуд зубців T .

На рис. 2.15 зображено результат моделювання за умов $A_T = 0,2$ мВ, рівень альтернатії $\Delta_A = 50$ мкВ, $\alpha_T^0 = 10\%$ (20 мкВ) та відсутні зовнішні спотворення, тобто $h(t) \equiv 0$.

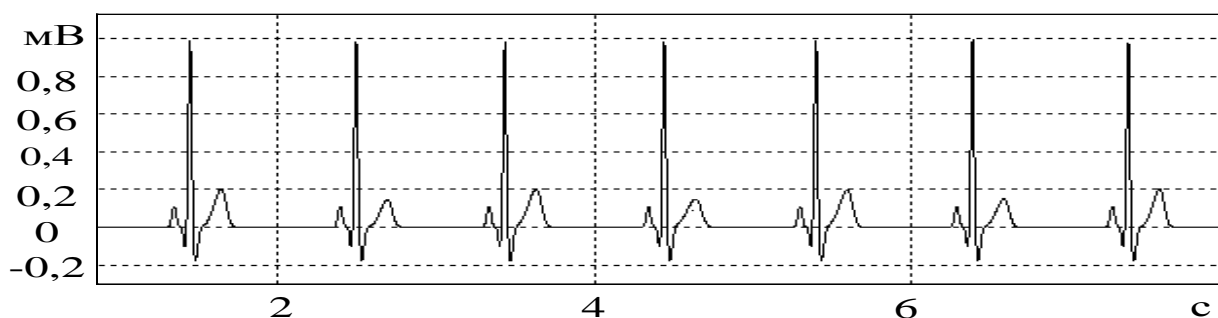


Рис. 2.15. Штучна ЕКГ із чітким чергуванням амплітуд зубця T

Зауважимо, що розглянута генеративна модель надає змогу імітувати різні типи альтернатії, які досить рідко трапляються у клінічній

практиці, наприклад, чергування зубців T з різною полярністю (рис. 2.16, а) або чергування виражених і невиражених зубців (рис. 2.16, б)

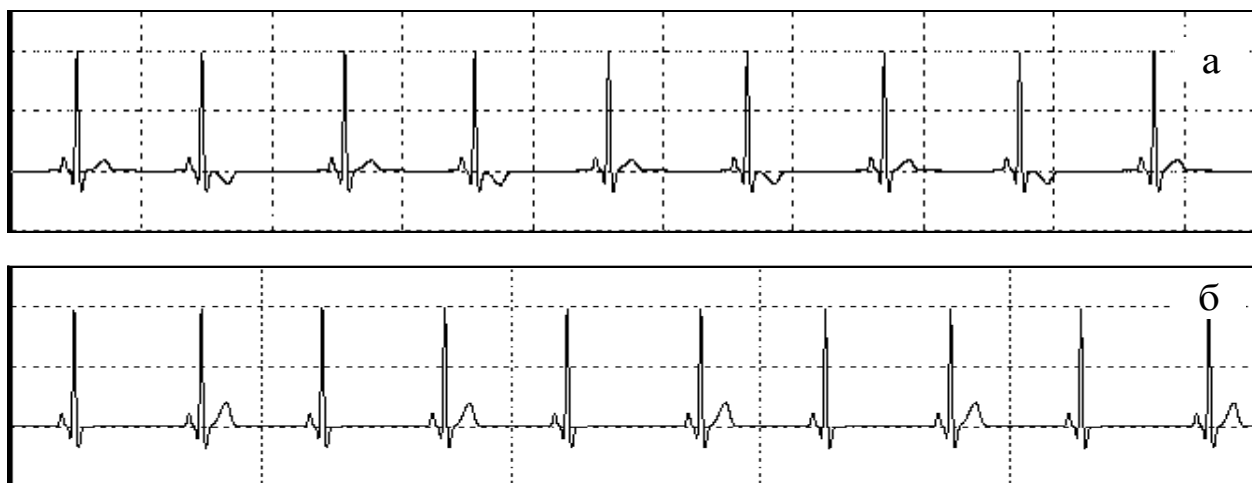


Рис. 2.16. Штучні ЕКГ з чергування зубців T різної полярності (а); виражених та невиражених зубців (б)

Для генерації штучної ЕКГ з L -кратною зміною фаз чергування зубців T сформуємо підмножину $I_L \subset \{1, \dots, L\}$ з L впорядкованих індексів, таких, що $I_L \cap (I_E \cup I_F) = \emptyset$, де $I_E \subset \{1, \dots, M\}$ і $I_F \subset \{1, \dots, M\}$ – підмножини упорядкованих індексів, що наявні у моделі (2.35), які визначають номери екстрасистол та артефактів.

Зважаючи на гіпотезу 2.3, потрібно, щоб усі елементи підмножини I_M задовольняли умови

$$|i - j| \geq M_0 \quad \forall i, j \in I_L, i \neq j, \quad (2.45)$$

де $M_0 = 16$ – мінімальна кількість циклів, за якими можна оцінювати рівень альтернативи.

Тоді процес генерації штучної ЕКГ знову буде реалізований співвідношеннями (2.35), (2.37), (2.38), (2.42), але замість (2.43) буде використано рекурентну формулу

$$\lambda_m^{(A)} = \begin{cases} 1 + \frac{\Delta_A}{A_T}, & \text{якщо } \lambda_{m-1}^{(A)} = 1 \quad \forall m \in \{1, \dots, M\} / I_L, \\ 1, & \text{якщо } \lambda_{m-1}^{(A)} = 1 + \frac{\Delta_A}{A_T} \quad \forall m \in \{1, \dots, M\} / I_L, \\ \lambda_{m-1}^{(A)}, & \forall m \in I_L. \end{cases} \quad (2.46)$$

З цього випливає, що співвідношення (2.43) є частковим випадком співвідношення (2.46), якщо $I_L = \emptyset$. Якщо ж $I_L \neq \emptyset$, то, згідно з (2.46), на кожному циклі $m \in I_L$ відбувається зміна фази чергування зубців, а далі, з урахуванням (2.45), порядок чергування зберігається принаймні на 16 наступних циклах.

На рис. 2.17 подано результат моделювання ЕКГ з екстрасистою та одноразовою зміною фази чергування зубців T різної амплітуди (момент зміни фази позначено стрілкою).

Запропоновані генеративні моделі природно узагальнюються у тому разі, коли альтернативу виражено в чергуванні інших характеристик зубців T , кожна з яких є реалізацією двох випадкових величин з різними математичними сподіваннями.

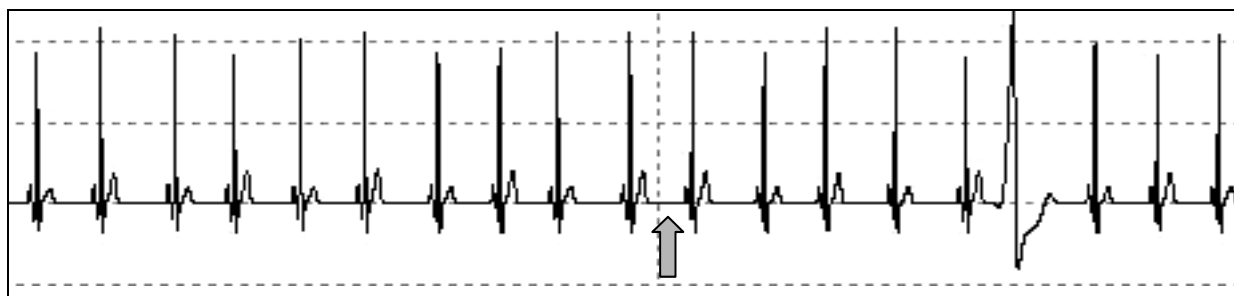


Рис. 2.17. Штучна ЕКГ з одноразовою зміною фази чергування зубців T

Так, наприклад, генерацію штучної ЕКГ із сумісною альтернативною амплітуд та ступеня симетрії зубців T реалізує модель (2.35), параметри \tilde{A}_{im} якої визначають рекурентні співвідношення (2.42), (2.46), параметри $\tilde{\mu}_{im}$ – формула (2.37), а для обчислення \tilde{b}_{im} використовують рекурентні співвідношення

$$\tilde{b}_{im} = \begin{cases} b_i^{(1)}(1 + \varepsilon_{im}^{(1)}) \quad \forall t \leq \mu_{im}, \quad \forall i \in \{P, Q, R, S, ST\}, \\ b_i^{(2)}(1 + \varepsilon_{im}^{(2)}) \quad \forall t > \mu_{im}, \quad \forall i \in \{P, Q, R, S, ST\}, \\ b_T^{(1)} \lambda_m^{(b1)} (1 + \varepsilon_{Tm}^{(1)}) \quad \forall t \leq \mu_{Tm}, \quad i = T, \\ b_T^{(2)} \lambda_m^{(b2)} (1 + \varepsilon_{Tm}^{(2)}) \quad \forall t > \mu_{Tm}, \quad i = T, \end{cases} \quad (2.47)$$

$$\lambda_m^{(b1)} = \begin{cases} 1 + \frac{\Delta_\beta}{b_T^{(1)}}, \quad \text{якщо } \lambda_{m-1}^{(b1)} = 1 \quad \forall m \in \{1, \dots, M\} / I_L, \\ 1, \quad \text{якщо } \lambda_{m-1}^{(b1)} = 1 + \frac{\Delta_\beta}{b_T^{(1)}} \quad \forall m \in \{1, \dots, M\} / I_L, \\ \lambda_{m-1}^{(b1)}, \quad \forall m \in I_L, \end{cases} \quad (2.48)$$

$$\lambda_m^{(b2)} = \begin{cases} 1 - \frac{\Delta_\beta}{b_T^{(2)}}, \quad \text{якщо } \lambda_{m-1}^{(b2)} = 1 \quad \forall m \in \{1, \dots, M\} / I_L, \\ 1, \quad \text{якщо } \lambda_{m-1}^{(b2)} = 1 - \frac{\Delta_\beta}{b_T^{(2)}} \quad \forall m \in \{1, \dots, M\} / I_L, \\ \lambda_{m-1}^{(b2)}, \quad \forall m \in I_L, \end{cases} \quad (2.49)$$

за початкових умов $\lambda_0^{(b1)} = \lambda_0^{(b2)} = 1$.

Величина Δ_β , що наявна у (2.48), (2.49), визначає заданий рівень альтернативи параметрів $b_T^{(1)}$ і $b_T^{(2)}$, що характеризують симетрію зубців T .

Зауважимо, що величина Δ_β , що наявна у співвідношеннях (2.48), (2.49), не впливає на загальну тривалість зубців, оскільки моменти початку $t_T^{(1)}$ та закінчення $t_T^{(2)}$ зубця T визначаються співвідношеннями $t_T^{(1)} = \mu_T - 3b_T^{(1)}$ і $t_T^{(2)} = \mu_T + 3b_T^{(2)}$, а у разі альтернації параметр $b_T^{(1)}$ збільшується на величину Δ_β , тоді як параметр $b_T^{(2)}$ зменшується на таку саму величину (рис. 2.18).



Рис. 2.18. Штучна ЕКГ з одночасною альтернацією амплітуд та симетрії зубців T

І, нарешті, для моделювання ефекту електричної альтернації серця в загальному випадку, коли спільно змінюються амплітуди, симетрії та тривалості зубців T , достатньо перейти від (2.48), (2.49) до рекурентних співвідношень

$$\lambda_m^{(b1)} = \begin{cases} 1 + \frac{\Delta_\beta}{b_T^{(1)}} + \frac{\Delta_\tau}{6b_T^{(1)}}, \text{ якщо } \lambda_{m-1}^{(b1)} = 1 & \forall m \in \{1, \dots, M\} / I_L, \\ 1, \text{ якщо } \lambda_{m-1}^{(b1)} = 1 + \frac{\Delta_\beta}{b_T^{(1)}} + \frac{\Delta_\tau}{6b_T^{(1)}} & \forall m \in \{1, \dots, M\} / I_L, \\ \lambda_{m-1}^{(b1)}, & \forall m \in I_L, \end{cases} \quad (2.50)$$

$$\lambda_m^{(b2)} = \begin{cases} 1 - \frac{\Delta_\beta}{b_T^{(2)}} + \frac{\Delta_\tau}{6b_T^{(2)}}, \text{ якщо } \lambda_{m-1}^{(b2)} = 1 & \forall m \in \{1, \dots, M\} / I_L, \\ 1, \text{ якщо } \lambda_{m-1}^{(b2)} = 1 - \frac{\Delta_\beta}{b_T^{(2)}} + \frac{\Delta_\tau}{6b_T^{(2)}} & \forall m \in \{1, \dots, M\} / I_L, \\ \lambda_{m-1}^{(b2)}, & \forall m \in I_L, \end{cases} \quad (2.51)$$

у яких величина Δ_τ визначає заданий рівень альтернації тривалості τ зубців T . В цьому випадку за умови альтернації кожен із параметрів $b_T^{(1)}$ і $b_T^{(2)}$ збільшується на величину $\frac{1}{6}\Delta_\tau$, а значить тривалість τ , яку визначають виразом $\tau = 3(b_T^{(1)} + b_T^{(2)})$, збільшиться на величину Δ_τ .

На рис. 2.19 подано результат моделювання штучної ЕКГ реалістичної форми з імітацією ефекту альтернації зубців T в загальному випадку на тлі тренда ізоелектричної лінії та випадкового шуму.



Рис. 2.19. Штучна ЕКГ з одночасною альтернацією амплітуд, тривалостей та симетрії зубців T

2.5. Порівняльний аналіз штучних та реальних ЕКГ

Сучасні засоби Інтернет-технологій надають змогу провести порівняння штучних ЕКГ із записами реальних ЕКГ, що зберігаються у спеціалізованих базах даних. Один із популярних ресурсів – Інтернет-портал **PhysioNet**, створений під керівництвом трьох інститутів США – National Institute of Biomedical Imaging and Bioengineering, National Institute of General Medical Sciences та National Institutes of Health.

PhysioNet містить великий архів цифрових записів фізіологічних сигналів (PhysioBank), зокрема 26 спеціалізованих баз ЕКГ.

Наведемо лише деякі з результатів експериментів, основна мета яких:

- **якісне** оцінювання можливостей запропонованих моделей створювати штучні сигнали реалістичної форми, візуальний аналіз яких лікарем-кардіологом ідентичний до інтерпретації реальних записів;

- **кількісне** оцінювання можливостей запропонованих моделей точно описувати форму та амплітудно-часові параметри окремих фрагментів циклів ЕКГ.

На рис. 2.20 подано результати порівняння реального запису ЕКГ № 201 бази MIT-BIH Arrhythmia Database, яка зареєстрована у чоловіка 68 років з діагнозом миготлива аритмія, та штучна ЕКГ, створена на основі моделі (2.20).

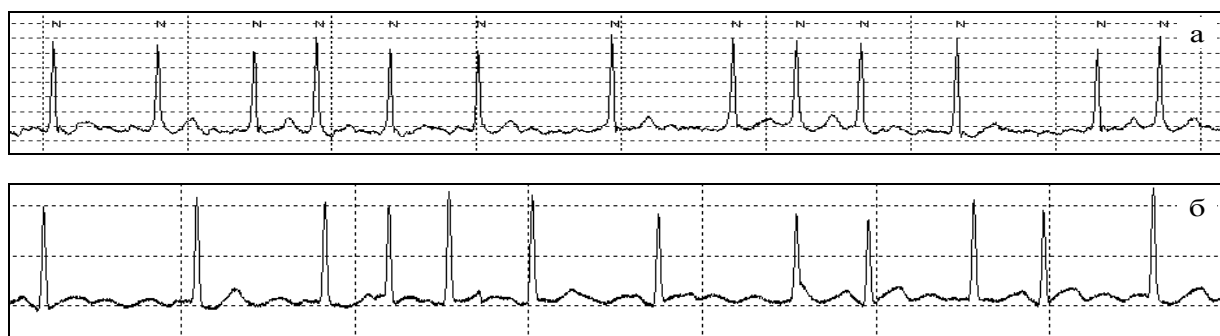


Рис. 2.20. ЕКГ хворого з діагнозом миготлива аритмія (а)
та модельна ЕКГ(б)

З рис. 2.20 можна побачити, що штучна ЕКГ за своєю формою досить близька до реального запису, а її візуальний аналіз однозначно призводить до встановлення правильного діагнозу – миготлива аритмія.

На рис. 2.21 подано фрагмент запису ЕКГ з екстрасистолами, що належить жінці 51 року із запису № 119 бази MIT-BIH Arrhythmia Database, а внизу – модельна ЕКГ, генерована на основі моделі (2.35).



Рис. 2.21. Реальна (а) та модельна (б) ЕКГ з екстрасистолами

Як видно і в цьому випадку, візуальний аналіз штучної ЕКГ аналогічний візуальному оцінюванню реального запису, що призводить до однозначного висновку про наявність шлуночкових екстрасистол.

Штучна ЕКГ із негативним зубцем *T* досить близька до реальної ЕКГ № 100 бази даних MIT-BIH Arrhythmia Database, зареєстрованої у чоловіка 69 років, що видно з порівняння цих графіків (рис. 2.22).

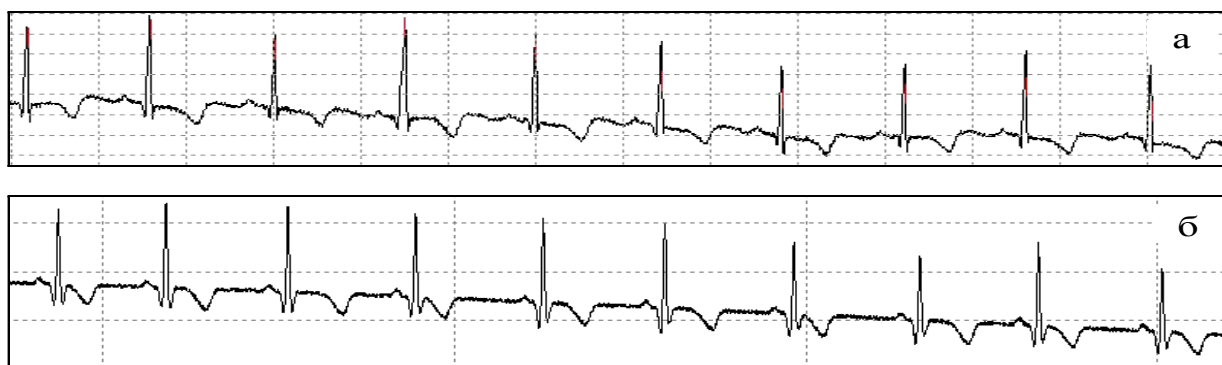


Рис. 2.22. Реальна (а) та модельна (б) ЕКГ з інверсним зубцем T

Генеративна модель (2.29) дає змогу синтезувати штучну ЕКГ із депресією сегмента ST , яка під час візуального аналізу фактично не відрізняється від реального запису ЕКГ, поданого кардіологами (рис. 2.23).



Рис. 2.23. Реальна (а) та модельна (б) ЕКГ з депресією сегмента ST

Аналогічні результати отримано під час візуального порівняння реального запису ЕКГ з альтернацією зубця T та штучної ЕКГ, генерованої моделлю (2.42) (рис. 2.24).



Рис. 2.24. Реальна (а) та модельна (б) ЕКГ з альтернацією зубця T

Розглянемо також деякі з результатів доказових експериментів, спрямованих на кількісне оцінювання точності реконструкції циклів реальних ЕКГ за сукупністю параметрів еталонів моделей.

Ідея таких експериментів полягала в тому, щоб за дискретними відліками циклу ЕКГ, що зберігається у спеціалізованих базах даних Інтернет-ресурсу PhysioNet, оцінювалися оптимальні значення вектора параметрів моделі (2.27)

$$G = (a_P, b_P^{(1)}, b_P^{(2)}, \mu_P, \dots, a_T, b_T^{(1)}, b_T^{(2)}, \mu_T, t_0), \quad (2.52)$$

що забезпечують найкраще наближення штучного циклу $z_0(k)$ до реального циклу $z_r(k)$ за критерієм

$$Cr = \sum_{k=0}^{K_r-1} [z_0(k) - z_r(k)]^2, \quad (2.53)$$

де K_r – кількість дискретних відліків $z_r(k)$, що визначається тривалістю T_r реального циклу та частотою дискретизації F_D (кроком квантування Δ за часом):

$$K_r = \frac{T_r}{\Delta}.$$

Тим самим формувалася «трійка» (рис. 2.25), що пов'язує реальний та штучний цикли ЕКГ за допомогою параметрів (2.52).

Експерименти показали, що за допомогою моделі (2.27) можна досить точно реконструювати реальні цикли різної форми (рис. 2.26).

Встановлено, що значення основних електрокардіографічних показників модельного та реального циклів, таких як депресія (рис. 2.26, а, б) або елевація (рис. 2.26, в) сегмента ST , тривалість розширеного комплексу QRS (рис. 2.26, б), амплітуди зубців, зокрема, негативного зубця T (рис. 2.26, д) перебували в межах, що дозволяють провести достовірну інтерпретацію сигналу.

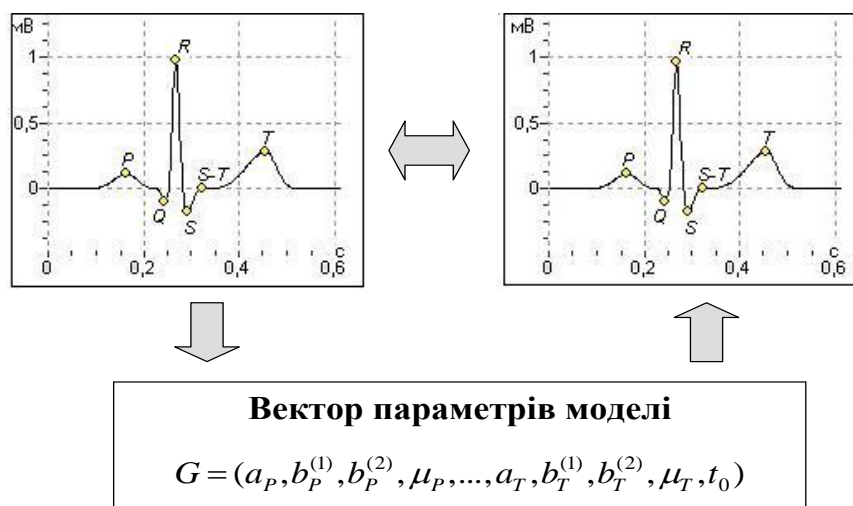


Рис. 2.25. Взаємозв'язок реального (ліворуч) та штучного (праворуч) циклів ЕКГ

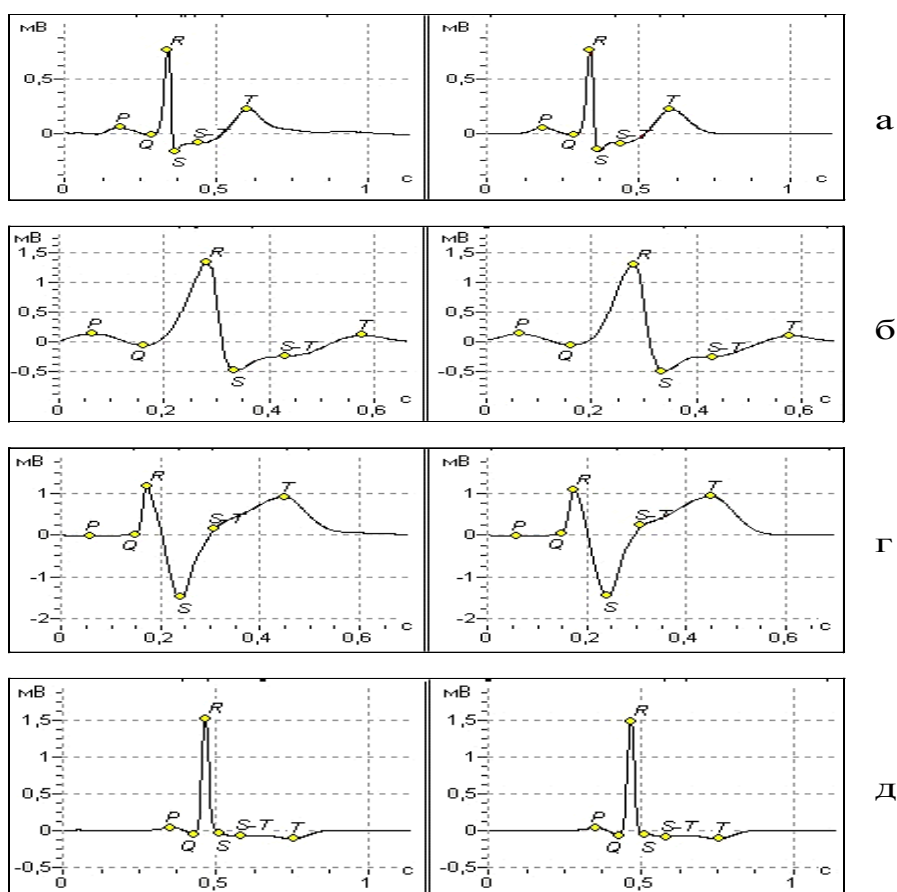


Рис. 2.26. Реальні (ліворуч) та модельні (праворуч) цикли різної форми

Експерименти підтвердили, що запропоновані моделі дають змогу генерувати штучні сигнали, форма яких збігається з формою реальних циклів у всій області визначення. Зокрема, під час реконструк-

ції циклу із негативним зубцем T (рис. 2.27) розбіжність $z_0(k)$ і $z_r(k)$ на сегменті ST було в межах 8 мкВ (менше 1 % діапазону зміни сигналу) в умовах депресії $ST = -0,125$ мВ.

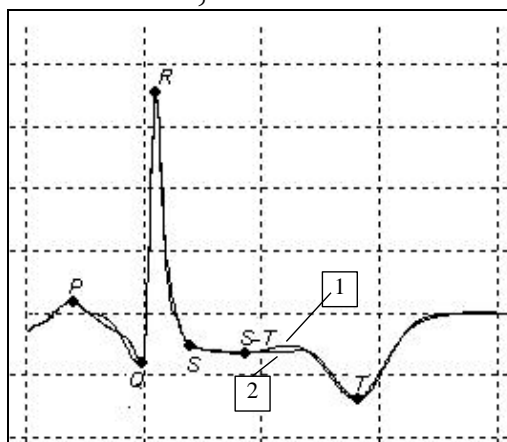


Рис. 2.27. Порівняння форми реального (1) та модельного (2) циклів

Незначна модифікація моделі (2.27)

$$z_0(t) = \sum_{i \in \{P, Q, R, R', S, ST, T\}} A_i \exp\left(-\frac{(t - \mu_i)^2}{2[b_i(t)]^2}\right) \quad (2.54)$$

дає змогу покращити якість реконструкції циклів більш складної форми (рис. 2.28), наприклад, моделювати реальний цикл із роздвоєним зубцем R (рис. 2.28, а) за рахунок його апроксимації не однією (рис. 2.28, б), а двома гаусовими функціями R і R' (рис. 2.28, в).

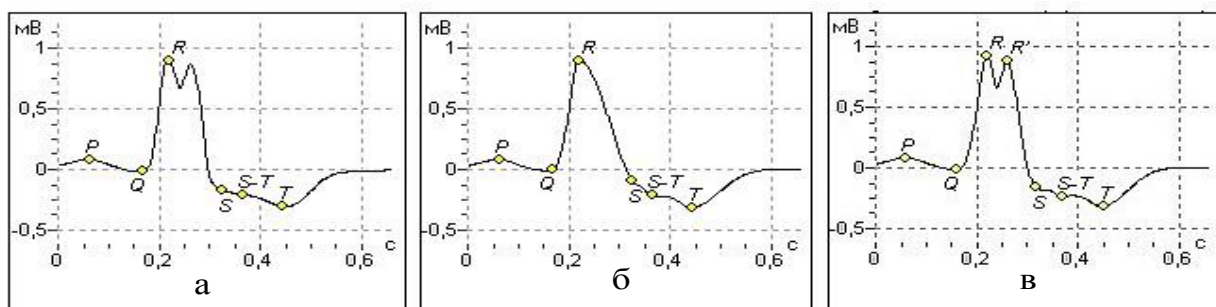


Рис. 2.28. Апроксимація реального циклу (а) основною (б) моделлю (2.27) та вдосконаленою (в) моделлю (2.54)

Розглянуті моделі штучних ЕКГ реалістичної форми можна використати для розв'язування різних наукових та прикладних задач.

Одна з таких задач, яка не втрачає актуальності, незважаючи на стрімкий розвиток засобів цифрової обчислювальної техніки та якос-

тей сучасних телекомунікаційних систем – економного зберігання та передачі ЕКГ (рис. 2.29)

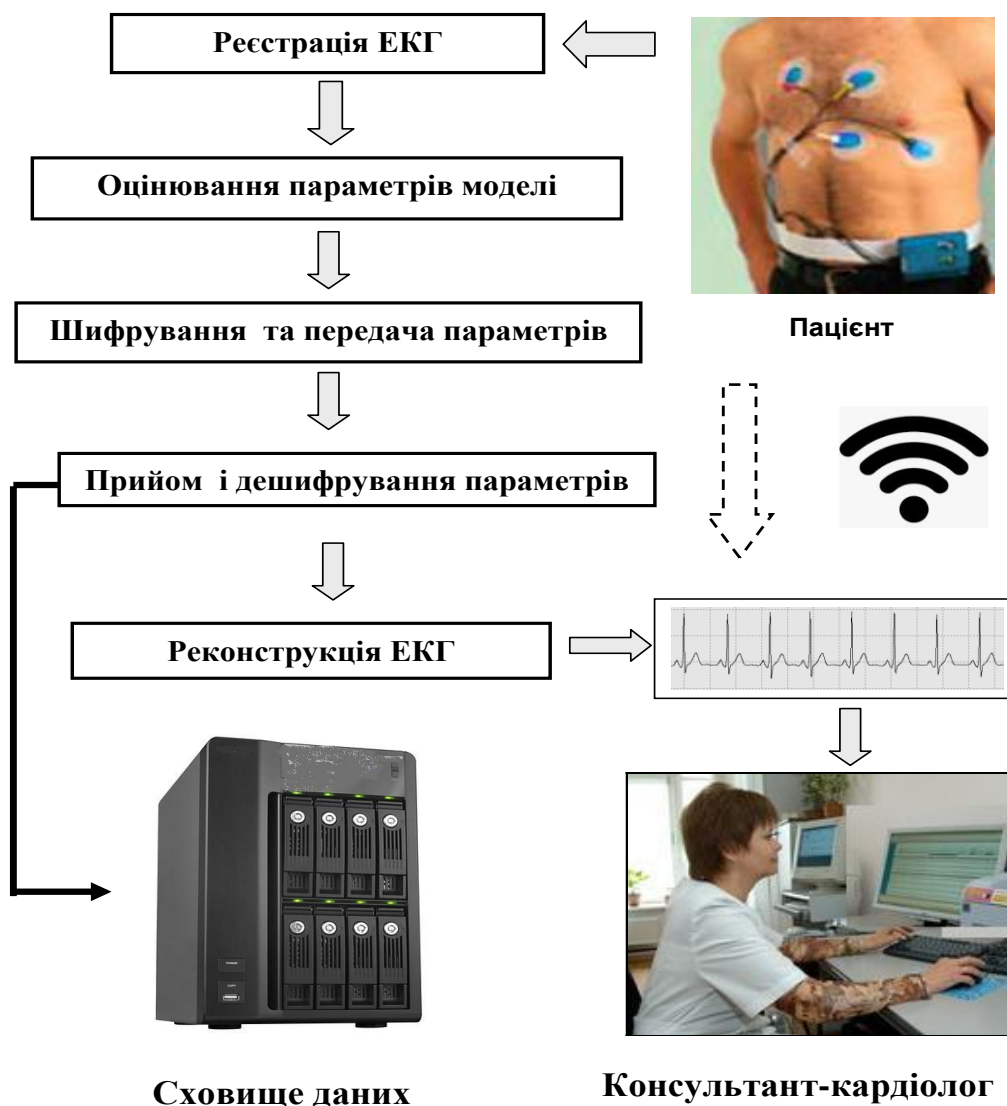


Рис. 2.29. Технологія економного зберігання та передачі ЕКГ

Замість передачі (зберігання) дискретних відліків ЕКГ пропонується передавати (зберігати) вектори оптимальних значень параметрів

$$G_m = (A_p[m], b_p^{(1)}[m], b_p^{(2)}[m], \mu_p[m], \dots, A_T[m], b_T^{(1)}[m], b_T^{(2)}[m], \mu_T[m], t_0[m]), \quad m = 1, 2, \dots \quad (2.55)$$

за якими відповідно до моделі можна реконструювати кожний m -й цикл ЕКГ.

Визначення векторів G_m , $m = 1, 2, \dots$ здійснюється на основі критерію (2.53) – мінімуму середнього квадрату відхилень дискретних значень реального та штучного циклів.

Така організація зберігання даних надає змогу не тільки суттєво скоротити обсяг інформації, що зберігається у сховищах даних, але й спростити доступ до потрібних записів для проведення різних досліджень, якщо як ключ використовувати значення будь-якого з параметрів A_P , $b_P^{(1)}$, $b_P^{(2)}$, μ_P, \dots, A_T , $b_T^{(1)}$, $b_T^{(2)}$, μ_T , що належить заданому діапазону значень.

Наприклад, якщо ЧСС = 60 уд/хв, то з частотою дискретизації $F_D = 500$ Гц кожен цикл ЕКГ має кодуватись 500 дискретними відліками, тоді як кодування вектором (2.22) містять лише 25 параметрів, тобто в 20 разів менше!

Зрозуміло, що для 24-годинного холтерівського запису, що містить не менше 100 тис. циклів ЕКГ (50 млн дискретних відліків), така економія на кожному циклі дуже суттєва, а якості реконструйованого сигналу достатньо для правильної діагностики (рис. 2.30).

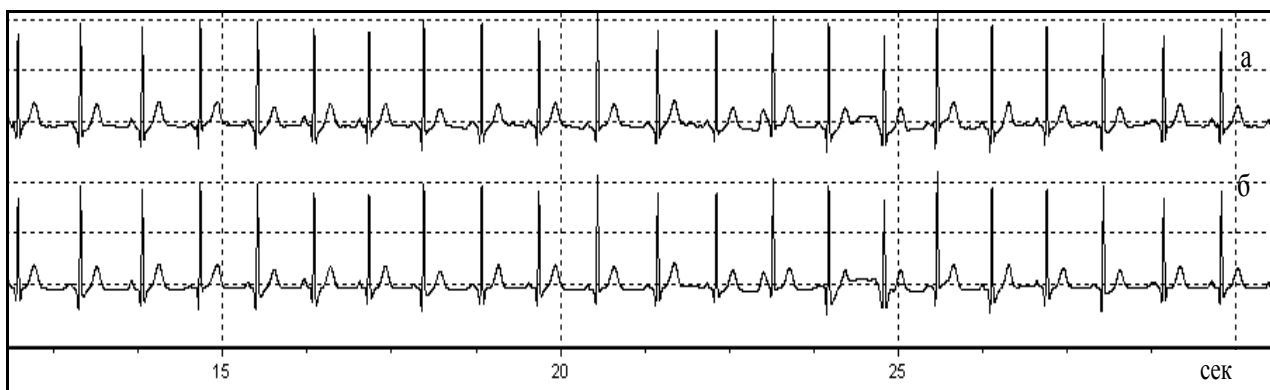


Рис. 2.30. Реальна (а) та реконструйована (б) ЕКГ

Завдання до комп'ютерних практикумів

Завдання 2.1. Необхідно розробити програму, яка надає змогу будувати цикли ЕКГ за формулою (2.27). Для спрощення виконання поставленої задачі припускаємо, що всі зубці ЕКГ симетричні за виключенням зубця T .

Робоче вікно програми може бути подібним до наведеного на рис. 2.31. Користувач має можливість вибрати тип зубця та за допомогою органів керування налаштувати параметри відповідних зубців.

Результатом роботи програми мають бути графіки циклів штучної ЕКГ, що реалізують різні варіанти форми, наведені на рис. 2.9.

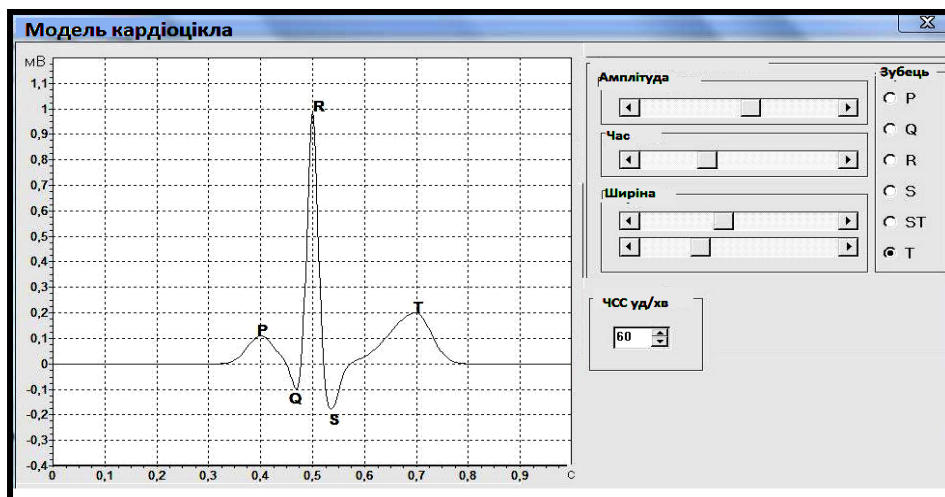


Рис. 2.31. Робоче вікно програми (завдання 2.1)

Завдання 2.2. Необхідно розробити програму, яка надає змогу будувати штучну ЕКГ з альтернацією зубця *T*. Генерація сигналу має здійснюватись за моделлю з чітким чергуванням амплітуд зубця *T*. Робоче вікно програми має включати елемент налаштування рівня альтернативі та кількості генерованих циклів.

Для спрощення виконання наступних практикумів на робочому вікні необхідно передбачити місце для розміщення елемента налаштування рівня шуму (рис. 2.32).



Рис. 2.31. Робоче вікно програми (завдання 2.1)

Запитання для самоконтролю

1. Назвіть основні функції серця, які забезпечують його роботу.
2. Що лежить в основі електричних явищ у серці?
3. Охарактеризуйте основні зубці та сегменти циклу ЕКГ.
4. Які традиційні відведення використовує електрокардіографія?
5. Які основні особливості реальних ЕКГ покладено в основу створення математичних моделей штучних сигналів?
6. Яка різниця між внутрішніми та зовнішніми збуреннями, що спотворюють цикли ЕКГ?
7. Поясніть особливість побудови інтерполяційної моделі породження ЕКГ реалістичної форми.
8. Яким чином базова інтерполяційна модель може бути узагальнена для опису процесу породження ЕКГ з нетиповими циклами?
9. Який основний недолік інтерполяційної моделі породження штучної ЕКГ?
10. Який механізм покладено в основу генеративної моделі породження штучної ЕКГ?
11. У чому полягає принцип спотворення циклів за генеративною моделлю?
12. Що таке електрична альтернація серця і чому важливо мати засоби виявлення такого явища?
13. Назвіть основні гіпотези, які покладено в основу генерації ЕКГ з електричною альтернацією серця?

РОЗДІЛ 3

ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ПРИГЛУШЕННЯ АДИТИВНИХ ЗАВАД

3.1. Загальні відомості з теорії сигналів

Повернемося до узагальненої моделі (1.3) породження сигналу з локально-зосередженими ознаками

$$z(t) = \Phi[z_0(t), \xi(t)] + h(t),$$

що спостерігається в умовах внутрішніх $\xi(t)$ та зовнішніх $h(t)$ збурень. Відповідно до цієї моделі для оцінювання корисного сигналу $z_0(t)$ за спостереженням $z(t)$ будемо на першому етапі пригнічувати зовнішні адитивні завади $h(t)$.

Покажемо, що для побудови ефективної ІТ оброблення сигналів з локально-зосередженими ознаками розв'язок такої відносно простої задачі потребує вдосконалення відомих алгоритмів частотно-виборчої фільтрації та згладжування випадкових шумів за допомогою реалізації спеціальних обчислювальних процедур.

Інтелектуальні властивості таких процедур дають змогу адаптуватися до конкретного сигналу $z(t)$, що обробляється, та визначати оптимальні значення параметрів налаштування цифрових фільтрів та алгоритмів згладжування. Порівняно з традиційними інтелектуальні процедури частотно-вибіркової фільтрації та згладжування підвищують ефективність пригнічення зовнішніх адитивних завад $h(t)$ без суттєвих спотворень корисного сигналу $z_0(t)$.

Під адаптацією розуміють здатність інтелектуальної системи пристосовуватися до невідомих та/або змінних ситуацій зовнішнього середовища. У біологічних системах завдяки фізіологічній адаптації забезпечується гомеостаз – здатність організму підтримувати оптимальні значення життєво важливих показників.

Один із яскравих прикладів адаптації – здатність живих організмів підтримувати температуру тіла у певних межах (терморегуляція) за суттєвих змін температури навколишнього середовища. В цьому випадку гомеостаз досягається завдяки додатковому теплоутворенню

та зниженню теплових втрат під час гіпотермії (зниження температури тіла нижче норми) або ж завдяки зменшенню теплоутворення та посиленню теплообміну з навколишнім середовищем під час гіпертермії (підвищення температури тіла вище за норму).

У загальному випадку задачу приглушення адитивних завад $h(t)$ можна сформулювати так. Необхідно з максимально можливою точністю відновити корисний сигнал $z_0(t)$ за спостереженою реалізацією

$$z(t) = z_0(t) + h(t) \quad (3.1)$$

в умовах неповної апріорної інформації про характеристики $h(t)$.

З теорії фільтрації відомі різні підходи до розв'язування такої класичної задачі.

Якщо під час реєстрації $z(t)$ можна синхронно вимірювати і перешкоду $h(t)$, то задача відновлення $z_0(t)$ суттєво спрощується: достатньо використовувати компенсаційну схему, на один вхід якої подавати спотворений сигнал $z(t)$, а на іншій – сигнал перешкоди $h(t)$.

Однак на практиці трапляються неконтрольовані перешкоди $h(t)$. У таких ситуаціях треба виділяти корисний сигнал $z_0(t)$ на фоні завади $h(t)$, використовуючи різні алгоритми фільтрації та згладжування.

Очевидно, що відокремити корисний сигнал $z_0(t)$ від перешкоди $h(t)$ можливо лише в тому разі, коли є хоча б мінімальні відмінності у характеристиках $z_0(t)$ та $h(t)$, наприклад, спектральні або статистичні характеристики.

Відповідно до загальної теорії сигналів оптимальним фільтром відновлення сигналу відомої форми на тлі білого гаусівського шуму є паралельний набір узгоджених фільтрів. В умовах протифазної передачі двох сигналів $z_{01}(t)$ і $z_{02}(t) = -z_{01}(t)$ достатньо мати лише один узгоджений фільтр. Якщо ж шуми гаусові, але не «білі», оптимальний фільтр стає значно складнішим і для точного визначення його структури потрібно вже розв'язувати відповідне інтегральне рівняння.

Оптимальний приймач у разі нескінченного часу спостереження це каскадне з'єднання попередньо «відбілюючого» й узгодженого фільтрів, причому частотна передатна функція «фільтра, що відбілює», інверсна функції спектральної щільності шумів, а узгоджений фільтр має бути попередньо налаштований на корисний сигнал.

У тих випадках, коли шум перестав бути гаусовим, ситуація ускладнюється: оптимальний (за критерієм максимуму правдоподібності) фільтр стає нелінійним та його структуру складно реалізувати.

Таким чином, навіть якщо форма корисного сигналу відома, побудова оптимального фільтра для виділення $z_0(t)$ на тлі адитивної перешкоди $h(t)$ є не простою задачею, яка має формальний розв'язок лише за певних передумов про модель $h(t)$, які не завжди виконуються на практиці.

Якщо ж функція $z_0(t)$ невідома, але може бути описана деякою параметричною моделлю $\hat{z}_0(t, \theta)$, то задачу фільтрації часто зводять до оцінювання вектора θ параметрів за спостережуваною реалізацією. Для дискретного сигналу $z[k], k = 1, 2, 3, \dots$ отримання такого оцінювання можливе з урахуванням рекурсивних процедур, які ґрунтуються на теорії оптимальної фільтрації Калмана. За певних передумов про сигнал $z_0(t)$ і адитивну перешкоду $h(t)$ такі процедури оптимальні у сенсі мінімуму середньоквадратичного відхилення $\hat{z}_0(t)$ від $z_0(t)$.

Однак такі процедури не гарантують збереження форми локальних фрагментів корисного сигналу і тому непридатні для оброблення сигналів складної форми з локально зосередженими ознаками.

Цілком очевидно, що в основу побудови ефективних обчислювальних процедур фільтрації сигналів з локально зосередженими ознаками має бути покладений інший критерій оптимальності, ніж середньоквадратичне відхилення сигналу за всією областю визначення. Наприклад, оцінювати мінімум середньоквадратичного відхилення сигналу на інформативних фрагментах або мінімум середньоквадратичного відхилення значень діагностичних ознак, зосереджених на таких фрагментах.

Проте універсальної теорії, що дозволяє формалізувати синтез фільтрів, які ґрунтуються на таких умовах, немає. Тому як для таких випадків потрібно запровадити деякі припущення про модель перешкоди та шукати прийнятні процедури, що забезпечують максимально можливе заглушення адитивної перешкоди за мінімальних спотвореннях значень діагностичних ознак, зосереджених на локальних фрагментах сигналу.

Природно, що підтвердженням ефективності (прагматичної «оптимальності») таких процедур будуть не формальні прийоми, а ре-

зультати чисельного моделювання на тестових і реальних даних відповідно до означення 2.1 та означення 2.2.

У реальних умовах реєстрації біомедичних сигналів, зокрема ЕКГ, трапляються зовнішні збурення різного типу (рис. 3.1).

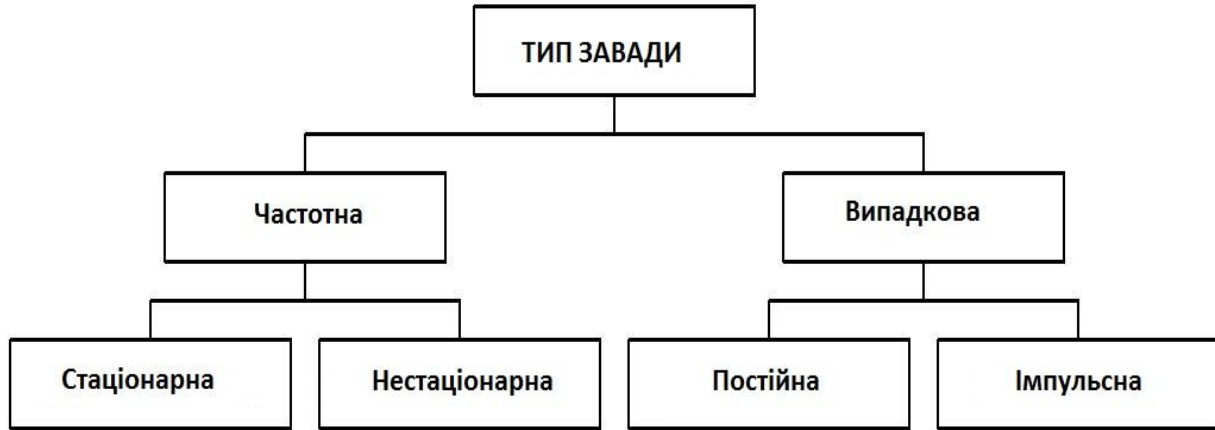


Рис. 3.1. Характерні типи зовнішніх збурень

До стаціонарної завади також належить мережева перешкода $h_1(t)$, зосереджена на 50 Гц та кратних частотах. До постійної випадкової (широкосмугової) завади належить, наприклад, перешкода $h_2(t)$, викликана м'язовим тремором у містах накладання електродів ЕКГ. Імпульсними називають короточасні високоамплітудні перешкоди $h_3(t)$, що викликаються електричним обладнанням, розміщеним поблизу електрокардіографа. Низькочастотний дрейф ізоелектричної лінії краще розглядати як нестаціонарну перешкоду $h_4(t)$.

Вважатимемо, що зазначені типи перешкод одночасно і незалежно спотворюють корисний сигнал, тобто будемо розглядати модель зовнішньої перешкоди як суму часткових моделей

$$h(t) = h_1(t) + h_2(t) + h_3(t) + h_F(t),$$

та будувати процедуру попереднього оброблення ЕКГ у вигляді послідовності фільтрів, кожний з яких спрямований на приглушення перешкоди певного типу.

Припустимо спочатку, що корисний сигнал спотворений лише частотною перешкодою $h_1(t)$. Покажемо, що під час оброблення ЕКГ та інших подібних сигналів з локально-зосередженими ознаками традиційні частотно-вибіркові фільтри не завжди забезпечують необхідну якість приглушення такої завади.

3.2. Адаптивна фільтрація частотних завад

Припустимо спочатку, що корисний сигнал спотворений лише частотною завадою, яка є сумою стаціонарних гармонічних коливань

$$h[k] = \sum_{g=1}^G a_g \cos(2\pi f_g k \Delta + \phi_g), \quad (3.2)$$

де a_g, f_g, ϕ_g – відповідно амплітуда, частота та початкова фаза g -ї гармоніки.

Вважатимемо, що величини a_g, ϕ_g заздалегідь невідомі, а частоти f_g гармонік відомі лише з точністю до обмежень f_g^{\min}, f_g^{\max} зверху та знизу, тобто

$$f_g^{\min} \leq f_g \leq f_g^{\max},$$

причому діапазони $\Delta_g = f_g^{\max} - f_g^{\min}$ можуть збігатися з областями навантажених частот корисного сигналу $z_0[\cdot]$.

Саме така складна ситуація виникає під час обробки реальних ЕКГ (рис. 3.2). Для прикладу побудуємо режекторний фільтр, який забезпечує ефективне приглушення гармонічної перешкоди (3.3) за мінімальних спотворень форми корисного сигналу $z_0[\cdot]$.

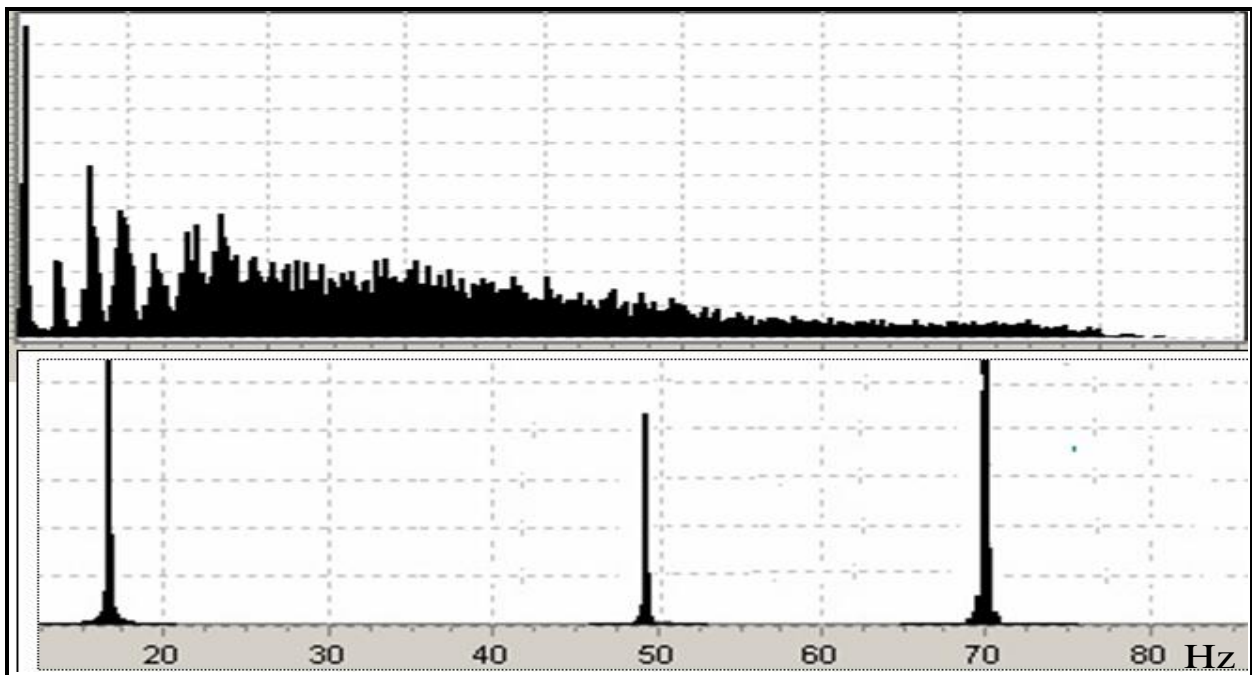


Рис. 3.2. Спектри ЕКГ (вгорі) та частотної завади (внизу)

У загальному випадку для приглушення гармонічних перешкод використовують частотно-виборчі фільтри, які пропускають сигнали певної смуги частот і затримують інші частоти.

Розрізняють такі частоти (рис. 3.3):

- фільтри нижніх частот (ФНЧ), які пропускають частоти, менші за частоту зрізу f_0 ;
- фільтри верхніх частот (ФВЧ), що пропускають частоти, вище за частоту зрізу f_0 ;
- смугові фільтри (СФ), що пропускають частоти в деякому інтервалі $[f_1, f_2]$;
- режекторні (загороджувальні) фільтри (РФ), що пропускають усі частоти, крім тих, що перебувають у певному інтервалі $[f_1, f_2]$.

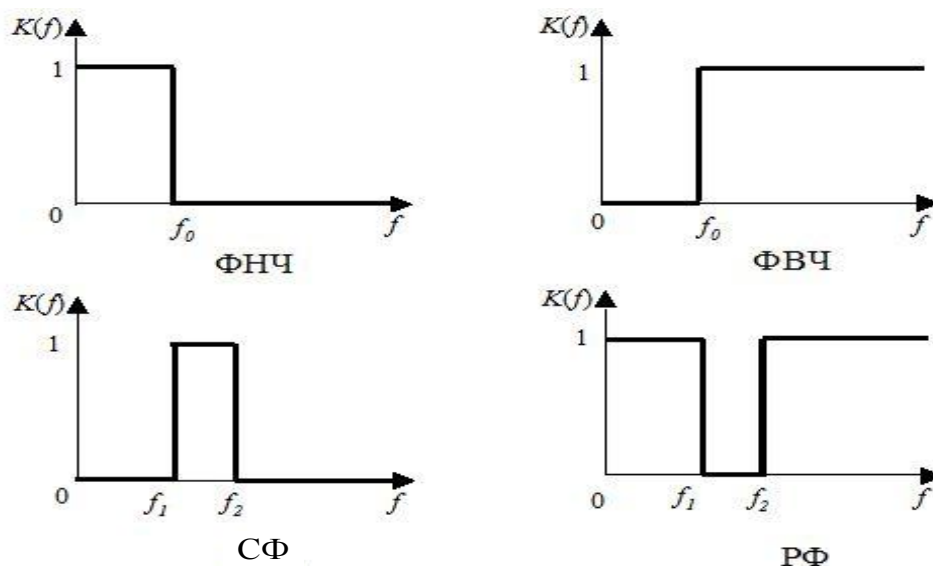


Рис. 3.3. Амплітудно-частотні характеристики фільтрів «ідеальних» фільтрів

Заглушення гармонічної перешкоди (3.3) можливе за допомогою каскаду режекторних фільтрів, налаштованих на загородження g -ї гармоніки. Для оброблення сигналу в реальному масштабі часу алгоритм фільтрації може бути фізично реалізований, якщо для отримання фільтрованого значення $z_{вих}[k]$ у k -й момент часу використовується інформація, отримана в попередні моменти часу $k, k-1, k-2$ і т. д.

У загальному випадку дискретний фільтр підсумовує (з ваговими коефіцієнтами) певну кількість вхідних значень $z_{вх}[\cdot]$ сигналу та певну кількість попередніх фільтрованих вихідних $z_{вих}[\cdot]$ значень:

$$z_{eux}[k] = b_0 z_{ex}[k] + b_1 z_{ex}[k-1] + \dots + b_m z_{ex}[k-m] - a_1 z_{eux}[k-1] - a_2 z_{eux}[k-2] - \dots - a_n z_{eux}[k-n]. \quad (3.3)$$

В окремому випадку, коли $a_1 = a_2 = \dots = a_n = 0$, тобто алгоритм не використовує попередні фільтровані значення, рівняння фільтра зводиться до дискретної згортки

$$z_{eux}[k] = \sum_{i=0}^m b_i z_{ex}[k-i]. \quad (3.4)$$

Фільтр, що реалізує алгоритм (3.4), називають нерекурсивним фільтром m -го порядку, який, на відміну від рекурсивного фільтра (3.3), завжди є стійким.

Задача проектування (синтезу) дискретного фільтра зводиться до вибору порядку та визначення значень параметрів вагових коефіцієнтів a_i, b_i . У тих випадках, коли фазові характеристики фільтра не важливі, така задача зводиться до апроксимації бажаної амплітудно-частотної характеристики, що доставляє мінімум p -норми відхилення, зокрема, середньоквадратичної помилки ($p = 2$), або максимальному відхиленню за модулем ($p \rightarrow \infty$).

Оскільки у загальному випадку така задача немає аналітичного розв'язку на практиці використовують різні чисельні алгоритми, зокрема метод, який ґрунтується на вирішенні системи рівнянь Юла-Уокера, метод експоненційного оцінювання Проні та низка інших субоптимальних методів синтезу фільтра.

Слід, однак, зауважити, що навіть у разі високого порядку m фільтра неможливо отримати «ідеальну» амплітудно-частотну характеристику (АЧХ) режекторного фільтра РФ прямокутної форми (рис. 3.3), яка забезпечує повне приглушення частот у смузі загородження без спотворення сигналу в смузі пропускання. В реальних умовах проектування загороджувального фільтра крім основних параметрів (частотних меж смуг пропускання f_p та затримання f_s) потребує враховувати додаткові вимоги – допустиму нерівномірності АЧХ у смузі пропускання R_p та мінімально необхідне згасання у смузі загородження R_s (рис. 3.4).

Синтез режекторного фільтра суттєво ускладнюється в ситуаціях, коли частотна завада зосереджена в діапазоні напружених частот корисного сигналу, оскільки зменшення різниці $\Delta_f = f_{2s} - f_{1s}$ неминуче призводить до збільшуються параметрів R_p, R_s . У результаті допус-

каються неприпустимі спотворення значень діагностичних ознак, зосереджених на локальних фрагментах сигналу.

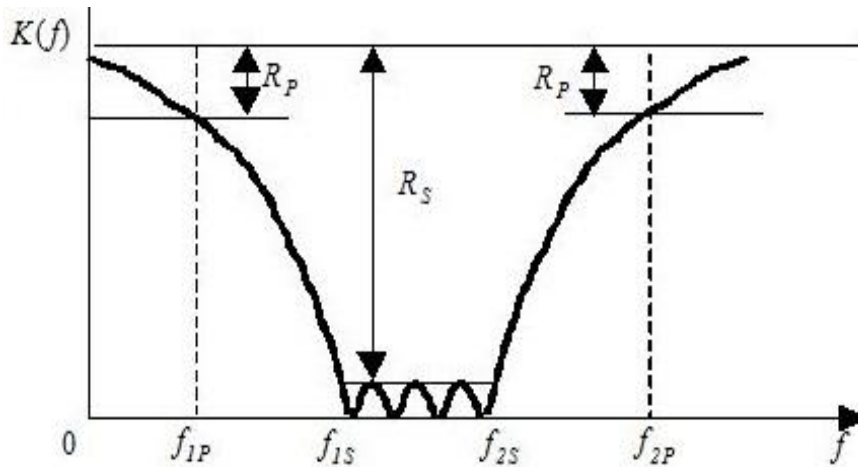


Рис. 3.4. Амплітудно-частотні характеристики реального режекторного фільтра

Розглянемо альтернативний підхід до приглушення гармонічних перешкод, який у науковій літературі називають методом фільтрації із зміною області перетворення сигналів. На відміну від розглянутих вище методів послідовного оброблення сигналу в часовій області такі фільтри проводять блокове оброблення всього масиву даних $z[1], \dots, z[K]$ у частотній області.

Алгоритм ґрунтується на використанні прямого *FFT* дискретного перетворення Фур'є

$$Z_n = \sum_{k=0}^{K-1} z[k] e^{-2\pi i n k / K}, \quad (3.5)$$

яке забезпечує перехід до частотних компонентів сигналу, та зворотного *IFFT* перетворення Фур'є, що виражається формулою

$$z[k] = \frac{1}{K} \sum_{n=0}^{K-1} Z_n e^{2\pi i n k / K}. \quad (3.6)$$

Суть алгоритму блокової фільтрації ґрунтується на двох відомих властивостях дискретного перетворення Фур'є (ДПФ):

а) ДПФ – лінійне відображення, отже якщо на сигнал діє адитивна перешкода, тобто

$$z(\cdot) = z_0(\cdot) + h(\cdot),$$

то спектр спостережуваного сигналу z дорівнює сумі спектрів корисного сигналу z_0 та перешкоди h , тобто

$$S\{z\} = S\{z_0\} + S\{h\}:$$

б) пряме FFT та зворотне $IFFT$ дискретні перетворення Фур'є оборотні, тобто з точністю до похибки округлення для будь-якого вектора $z = (z_1, \dots, z_N)$ виконується співвідношення

$$IFFT(FFT(z)) = z.$$

Звідси випливає, що для загородження стаціонарної гармонічної перешкоди (3.2), достатньо виконати послідовність лише трьох операцій:

- побудувати спектр $S\{z\} = S\{z_0\} + S\{h\}$ спостережуваного сигналу на основі прямого ДПФ;
- приглушити складові спектра $S\{z\}$, що відповідають частотам перешкоди $h(\cdot)$;
- відновити корисний сигнал $z_0[\cdot]$ у часовій області на основі зворотного ДПФ (рис. 3.5)

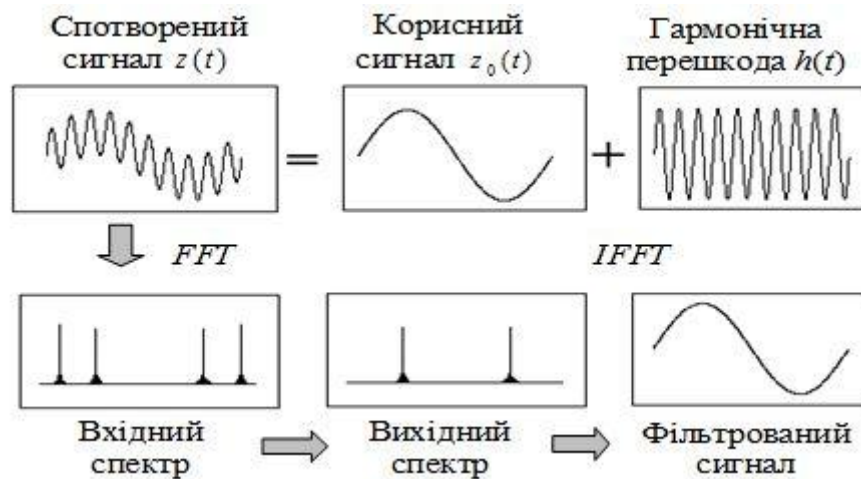


Рис. 3.5. Ідея режекторного фільтра на основі ДПФ

Однак такий привабливий прийом не завжди є ефективним, оскільки в загальному випадку спектр одиночного гармонічного коливання, побудований на підставі прямого ДПФ, розподілятиметься на цілу низку суміжних частот. Такий ефект у науковій літературі називають розходженням спектра дискретного гармонічного сигналу (**Spectrum leakage**) або ефектом Гіббса.

Таким чином, у загальному випадку для фільтрації зосередженої гармонічної перешкоди доводиться заглушувати всі складові ДПФ, на які в результаті розходження розподілилися спектральні компоненти перешкоди. Однак тоді доведеться торкнутися багатьох складових, що мають інформацію про корисний сигнал, і в результаті після зворотного перетворення Фур'є корисний сигнал буде суттєво спотворений.

Оскільки ефект розходження спектра впливає із самої природи ДПФ, опишемо цей ефект детальніше. Для цього розглянемо дискретний сигнал у вигляді найпростішого гармонічного коливання

$$h[k] = a \cos(\omega k \Delta + \varphi), \quad k = 0, \dots, K-1, \quad (3.7)$$

де a – амплітуда; Δ – крок квантування за часом; ω – колова частота; φ – початкова фаза.

Використовуючи процедуру (3.5) прямого ДПФ, подамо сигнал $h[k]$ у вигляді

$$Z_n = \sum_{k=0}^{K-1} h[k] e^{-2\pi i n k / K} = \sum_{k=0}^{K-1} a \cdot \cos(\omega k \Delta + \varphi) \cdot e^{-2\pi i n k / K}.$$

З урахуванням того, що $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$, маємо

$$\begin{aligned} Z_n &= \sum_{k=0}^{K-1} a \cdot \cos(\omega k \Delta + \varphi) \cdot e^{-2\pi i n k / K} = \sum_{k=0}^{K-1} \frac{a}{2} \cdot \left(e^{\omega k \Delta i + \varphi i - 2\pi i n k / K} + e^{-\omega k \Delta i - \varphi i - 2\pi i n k / K} \right) = \\ &= \sum_{k=0}^{K-1} \frac{a}{2} \cdot \left(e^{\varphi i} \cdot e^{ki(\omega \Delta - 2\pi n / K)} + e^{-\varphi i} \cdot e^{ki(-\omega \Delta - 2\pi n / K)} \right). \end{aligned}$$

Отже,

$$Z_n = \frac{a}{2} \cdot e^{\varphi i} \cdot \sum_{k=0}^{K-1} e^{ki(\omega \Delta - 2\pi n / K)} + \frac{a}{2} \cdot e^{-\varphi i} \cdot \sum_{k=0}^{K-1} e^{ki(-\omega \Delta - 2\pi n / K)}. \quad (3.8)$$

Оскільки величина $\sum_{k=0}^{K-1} e^{ki(\pm\omega \Delta - 2\pi n / K)}$ є сумою K членів геометричної прогресії зі знаменником $e^{i(\pm\omega \Delta - 2\pi n / K)}$, то

$$\sum_{k=0}^{K-1} e^{ki(\pm\omega \Delta - 2\pi n / K)} = \frac{1 - e^{i(\pm\omega \Delta - 2\pi n)}}{1 - e^{i(\pm\omega \Delta - 2\pi n / K)}}. \quad (3.9)$$

Отже,

$$Z_n = \frac{a}{2} e^{\phi i} \frac{1 - e^{i(K\omega\Delta - 2\pi n)}}{1 - e^{i(\omega\Delta - 2\pi n/K)}} + \frac{a}{2} \cdot e^{-\phi i} \frac{1 - e^{i(-K\omega\Delta - 2\pi n)}}{1 - e^{i(-\omega\Delta - 2\pi n/K)}}. \quad (3.10)$$

З виразу (3.10) видно, що $Y_n = 0$ тоді й тільки тоді, коли виконується умова

$$\frac{1 - e^{i(\pm K\omega\Delta - 2\pi n)}}{1 - e^{i(\pm\omega\Delta - 2\pi n/K)}} = 0. \quad (3.11)$$

Оскільки

$$e^{i(\pm K\omega\Delta - 2\pi n)} = \cos(\pm K\omega\Delta - 2\pi n) + i \sin(\pm K\omega\Delta - 2\pi n),$$

то чисельник виразу в лівій частині (3.11) перетворюється на нуль у тому разі, коли $(\pm K\omega\Delta - 2\pi n) = 2\pi M$, або, що те саме,

$$K\omega\Delta / 2\pi = M,$$

де M – ціле число.

Звідси випливає, що $Z_n = 0$ для тих значень n , для яких

$$e^{i(\pm\omega\Delta - 2\pi n/K)} \neq 1,$$

а значить знаменник виразу в лівій частині (3.11) не прямує до нуля.

У тому самому випадку, коли

$$e^{i(\pm\omega\Delta - 2\pi n/K)} = 1,$$

на підставі рівності (3.9) можна розкрити невизначеність вигляду $0/0$:

$$\frac{1 - e^{i(\pm K\omega\Delta - 2\pi n)}}{1 - e^{i(\pm\omega\Delta - 2\pi n/K)}} = \sum_{k=0}^{K-1} e^{ki(\pm\omega\Delta - 2\pi n/K)} = K.$$

Таким чином, якщо $K\omega\Delta / 2\pi$ є цілим числом, то

$$Z_n = \frac{aK}{2} e^{\pm i\phi}$$

для тих значень n , для яких виконано умову $e^{i(\pm\omega\Delta - 2\pi n/K)} = 1$, і $Z_n = 0$ для інших значень n , для яких не виконується умова $e^{i(\pm\omega\Delta - 2\pi n/K)} = 1$.

Очевидно, що умова $e^{i(\pm\omega\Delta - 2\pi n/K)} = 1$ виконується тоді і лише тоді, коли величина $\pm\omega\Delta - 2\pi n/K$ кратна числу 2π , тобто

$$n = K \left(\frac{\Delta \cdot \omega}{2\pi} - m \right),$$

або

$$n = K \left(m - \frac{\Delta \cdot \omega}{2\pi} \right),$$

де m – ціле число.

З урахуванням очевидного обмеження $0 < n \leq K$ величина m може приймати лише два значення, а саме $m = 0$ у першому випадку та $m = 1$ у другому випадку. Інакше кажучи, якщо відношення $K\omega\Delta/2\pi$ є цілим числом, то ненульові амплітуди Z_n у виразі (3.10) мають місце лише для значень

$$n = K \frac{\omega \cdot \Delta}{2\pi}$$

(у цьому випадку перший доданок виразу (3.10) не дорівнює нулю) і для значень

$$n = K \left(1 - \frac{\omega \cdot \Delta}{2\pi} \right)$$

(у цьому випадку другий доданок виразу (3.10) не дорівнює нулю).

У результаті з формули (3.10) остаточно отримуємо, що за умови, коли $K\omega\Delta/2\pi$ є цілим числом, обчислене ДПФ містить лише два спектральні відліки, відмінні від нуля:

$$Z_n = \begin{cases} \frac{aK}{2} e^{i\varphi}, & \text{коли } n = \frac{\omega\Delta}{2\pi} K, \\ \frac{aK}{2} e^{-i\varphi}, & \text{коли } n = \left(1 - \frac{\omega\Delta}{2\pi}\right) K, \\ 0, & \text{в інших випадках.} \end{cases} \quad (3.12)$$

Якщо відношення $K\omega\Delta/2\pi$ не є цілим числом, то $Z_n \neq 0$ для будь-якого значення n . В цьому і полягає ефект розходження спектра одиночного гармонічного коливання (3.7), яке подано кінцевим числом дискретних значень (рис. 3.6).

Зрозуміло, що для побудови ефективної процедури заглушення зосередженої гармонічної перешкоди потрібно усунути або послабити ефект розходження спектра дискретного гармонічного сигналу.

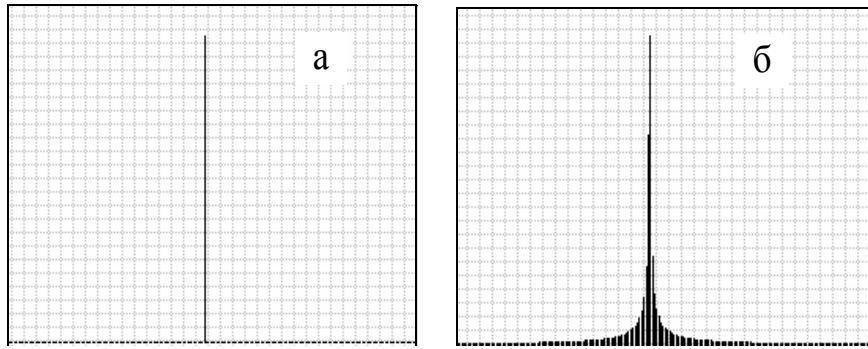


Рис. 3.6. Спектри одиночного гармонічного коливання (3.7), коли $K\omega\Delta / 2\pi$ – ціле число (а) та неціле (б) число

Відомим прийомом ослаблення ефекту розходження спектра є множення сигналу на вагову функцію $W(k)$, що має максимум у центральній точці сигналу (за $k = K/2$) і плавно спадну до країв для значень $k = 0$ і $k = K - 1$.

У цьому випадку формула ДПФ набуває вигляду

$$Z_n(W) = \sum_{k=0}^{K-1} W[k]z[k]e^{-i\frac{2\pi nk}{K}}.$$

Вважається, що за вдалого вибору $W(k)$ можна дещо послабити ефекти, пов'язані з виникненням стрибків сигналу під час періодичного продовження аналізованої кінцевої послідовності, і цим зменшити ефект розходження спектра. Однак використання вагової функції $W(k)$ призводить до спотворення спектра корисного сигналу.

У зв'язку з цим виникає природне бажання знайти спосіб боротьби з самим явищем, що породило ефект розходження спектра, а не з його наслідком. Інакше кажучи, варто спробувати отримати «адекватний» спектр одиночного дискретного гармонічного коливання на основі традиційного алгоритму ДПФ без використання вагової функції $W(k)$.

Скористаємося виразом (3.10) та подамо амплітуди спектральної щільності як функцію дійсної змінної k :

$$Z_n(k) = \frac{a}{2} \cdot e^{\phi i} \cdot \frac{1 - e^{i(k\omega\Delta - 2\pi n)}}{1 - e^{i(\omega\Delta - 2\pi n/k)}} + \frac{a}{2} \cdot e^{-\phi i} \cdot \frac{1 - e^{i(-k\omega\Delta - 2\pi n)}}{1 - e^{i(-\omega\Delta - 2\pi n/k)}}.$$

Припустимо, що $K \cdot (\omega\Delta / 2\pi) = R$, де R не є цілим числом. За таких умов, як було показано вище, обов'язково спостерігається ефект розходження спектра. Однак очевидно, що коли $R > 1$, то є таке зна-

чення $k < K$ (не обов'язково ціле), що $k \cdot (\omega\Delta/2\pi) = \lfloor R \rfloor$, де $\lfloor \eta \rfloor$ тут і далі означає цілу частину числа η .

Тоді, якщо визначити число K_0 таке, що

$$K_0 = \sup\{k : k < K, k \cdot (\omega\Delta/2\pi) - \text{ціле}\} = \lfloor R \rfloor \cdot 2\pi/\omega\Delta, \quad (3.13)$$

то враховуючи неперервність функції $Z_n(k)$ на інтервалі $[K_0, K]$, маємо

$$\lim_{k \rightarrow K_0} Z_n(k) = Z_n(K_0).$$

Оскільки $K_0 \cdot (\omega\Delta/2\pi) = \lfloor R \rfloor$ – ціле число, то за $k = K_0$ дотримується раніше отримана умова відсутності ефекту розходження спектра.

Отже, якщо K_0 – ціле число, то за рахунок зменшення довжини інтервалу спостережень від K до K_0 відбувається поступове зменшення ефекту розходження спектра та його повне зникнення у разі значення довжини масиву K_0 , тобто

$$\lim_{k \rightarrow K_0} Z_n(k) = Z_n(K_0) = \begin{cases} \frac{aK_0}{2} e^{i\varphi}, & \text{коли } n = \frac{\omega\Delta}{2\pi} K_0, \\ \frac{aK_0}{2} e^{-i\varphi}, & \text{коли } n = (1 - \frac{\omega\Delta}{2\pi}) K_0, \\ 0, & \text{в інших випадках} \end{cases}$$

Однак випадок, коли K_0 , яке задовольняє умову (3.13), є цілим числом швидше рідкісний виняток із правила, ніж правило. Зрозуміло, що коли K_0 не є цілим, повністю усунути ефект розходження спектра не вдається.

Водночас, коли довжина масиву зменшується від K та наближається до $K_0 < K$, то ефект розходження спектра буде меншим, ніж під час початкової довжини K .

Зауважимо, що сформульована вище умова відсутності ефекту розходження спектра еквівалентна тому, що $\omega\Delta/2\pi = M/K$, де M – ціле число. Отже, можна сформулювати таке твердження.

Твердження 3.1. Якщо невідома колова частота ω дискретного гармонічного сигналу та крок квантування за часом Δ такі, що

$$\omega\Delta/2\pi = p/q,$$

де p/q – деяке раціональне число, то для довжини K оброблюваного сигналу, кратної знаменнику цього числа, спектральне розкладання сигналу на основі ДПФ матиме всього дві спектральні компоненти відмінні від нуля.

Справді, якщо $\omega\Delta/2\pi = p/q$, то за умови, що K кратне q , тобто $K = m \cdot q$, маємо $K\omega\Delta/2\pi = mp$ – ціле число і, як було показано вище, ефекту розходження спектра немає.

Наслідок 3.1. Якщо невідома колова частота ω дискретно заданого одиничного гармонічного сигналу та крок квантування за часом Δ такі, що

$$\omega\Delta/2\pi = p/q,$$

де p/q – деяке раціональне число, знаменник якого q менше довжини оброблюваного сигналу K , то за умови зменшення кількості оброблюваних точок відбувається періодичне з періодом q ослаблення ефекту розходження спектра до його повного усунення.

Справді, якщо

$$K = nq + m \quad (m < q),$$

то для значення $K_1 = K - m = nq$ буде виконано умову:

$$K_1\omega\Delta/2\pi = nq \cdot p/q = np,$$

де np – ціле. Отже, для такої довжини масиву ефекту розходження спектра не буде.

Під час подальшого зменшення довжини сигналу для значень $K_2 = K - m - q, \dots, K_i = K - m - (i-1)q = (n+1-i)q$ буде виконано умову

$$K_i\omega\Delta/2\pi = (n+1-i) \cdot q \cdot p/q = (n+1-i)p - \text{ціле число},$$

і тоді періодично з періодом q відбуватиметься усунення ефекту розходження спектра.

Приклад 3.1. Нехай $\omega\Delta/2\pi = 1/7$, $K = 1000$. У цьому випадку умову твердження 3.1 не виконано.

Однак якщо $K_1 = 994$, $K_2 = 987$, $K_3 = 980, \dots, K_i = 1001 - 7i$ довжина інтервалу спостережень стає кратною 7, тобто знаменнику чис-

ла $\omega\Delta/2\pi$, і отже, на таких інтервалах відбувається повне усунення ефекту розходження спектра.

Цілком очевидно, що випадок, коли $\omega\Delta/2\pi$ є раціональним числом, також є рідкісним винятком: у загальному випадку число $\alpha = \omega\Delta/2\pi$ швидше за все буде ірраціональним. Тому можна спробувати наблизити величину $\omega\Delta/2\pi$ деяким раціональним числом.

У теорії чисел є ефективний апарат наближення дійсних чисел раціональними дробами, зокрема, ланцюговими дробами і відповідними дробами.

Відповідно до теореми Діріхле для будь-якого дійсного числа α та довільного $\tau > 1$ можна знайти раціональний дріб $\frac{a}{b}$ такий, що

$$\left| \alpha - \frac{a}{b} \right| < \frac{1}{b\tau}, \quad b \leq \tau. \quad (3.14)$$

Як дріб, що задовольняє умову теореми Діріхле, використовують n -й підхідний дріб $\frac{P_n}{Q_n}$ розкладання α у ланцюговий дріб, де n – найбільший номер такий, що $Q_n < \tau$.

Нагадаємо, що ланцюговий дріб $[a_0, a_1, a_2, \dots, a_n, \dots]$ є «багатоступінчастим» кінцевим або нескінченним дробом вигляду:

$$[a_0, a_1, a_2, \dots, a_n, \dots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_{n-1} + \frac{1}{a_n + \dots}}}}$$

Усічення ланцюгового дроби $[a_0, a_1, a_2, \dots, a_n, \dots]$ до $k < n$ компонент є k -й відповідний дріб $[a_0; a_1, \dots, a_k]$.

Приклад 3.2. Потрібно наблизити ірраціональне число $\sqrt{5}$ раціональним дробом $\frac{a}{b}$ з точністю до $\frac{1}{1000b}$. Для цього розкладемо $\sqrt{5}$ у ланцюговий дріб $\sqrt{5} = [2, 4, 4, 4, 4, \dots]$ та знайдемо відповідні дроби для цього розкладання (табл. 3.1).

Таблиця 3.1. Дробі розкладання числа $\sqrt{5}$

a_n	2	4	4	4	4	4	...
P_n	2	9	38	161	682
Q_n	1	4	17	72	305	1292	...

Найбільшим знаменником дробу, який залишається меншим від числа 1000, є $Q_n = 305$. Тоді шуканий дріб дорівнює $\frac{682}{305}$, причому

$$\left| \sqrt{5} - \frac{682}{305} \right| < \frac{1}{1000 \cdot 305}.$$

Таким чином, згідно з теоремою Діріхле, величину $\alpha = \omega \cdot \Delta / (2\pi)$ можна наблизити деяким раціональним числом $\frac{p}{q}$ так, що

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q\tau} \text{ для будь-якого } \tau > 1.$$

Тоді, якщо взяти $K = Mq$, де $M \ll \tau$, то

$$|K \cdot \alpha - M \cdot p| = \left| M \cdot q \cdot \alpha - M \cdot q \cdot \frac{p}{q} \right| = M \cdot q \cdot \left| \alpha - \frac{p}{q} \right| < M \cdot q \cdot \frac{1}{q \cdot \tau} = \frac{M}{\tau}.$$

Отже, якщо $\omega \cdot \Delta / (2\pi) = \sqrt{5}$, то для інтервалу завдовжки $K = 1000$ та заданою точністю порядку $1/1000$ із зменшенням довжини масиву до значень $K_1 = 915$, $K = 610$, $K = 305$ величина $K_i \cdot \omega \cdot \Delta / (2\pi)$ буде фактично цілим числом з точністю $0,001 i$, ($i = 1, 2, 3$), наслідком чого буде значне зменшення ефекту розходження спектра.

Експерименти показують, що у разі зменшення довжини масиву усунення ефекту розходження спектра відбувається з меншим періодом. Справа в тому, що відповідний дріб – не єдина можливість наближення дійсного числа раціональними дробами.

Відповідно до теореми Гурвіца, для будь-якого дійсного числа α існує множина раціональних дробів $\frac{a}{b}$ таких, що

$$\left| \alpha - \frac{a}{b} \right| < \frac{1}{\sqrt{5} \cdot b^2}. \quad (3.15)$$

Більше того, згідно з дослідженнями Маркова за винятком низки чисел певного виду оцінювання (3.15) може бути покращене:

$$\left| \alpha - \frac{a}{b} \right| < \frac{1}{\sqrt{8} \cdot b^2}. \quad (3.16)$$

Таким чином, завжди можна з достатнім ступенем точності наблизити в загальному випадку ірраціональне число $\alpha = \omega \cdot \Delta / (2\pi)$ деяким раціональним числом і цим зменшити ефект розходження спектра.

За відомих ω і Δ у деяких випадках можна аналітично визначити оптимальне значення K_{opt} , за яким розходження спектра буде відсутнє.

Твердження 3.2. Якщо колова частота ω одиночної гармонічної перешкоди та крок квантування за часом Δ відомі, причому число $\alpha = \omega \cdot \Delta / (2\pi)$ є раціональним, тобто $\alpha = p/q$, де $q < K$, то оптимальне значення K_{opt} , за яким розходження спектра буде відсутнім, визначається співвідношенням:

$$K_{opt} = \lfloor K/q \rfloor \cdot q.$$

Справді, нехай $K = aq + b$, де $b < q$. Тоді

$$K_{opt} = \lfloor K/q \rfloor \cdot q = aq,$$

та число

$$K_{opt} \omega \Delta_t / 2\pi = K_{opt} \cdot p/q = aq \cdot p/q = ap$$

є цілим, а значить за умови, що довжина сигналу дорівнює K_{opt} згідно з твердженням 3.1, ефект розходження спектра буде відсутній.

Якщо ж за відомих ω і Δ число $\alpha = \omega \cdot \Delta / (2\pi)$ не є раціональним, то для визначення K_{opt} слід розкласти в ланцюговий дріб числа $\alpha = \omega \cdot \Delta / (2\pi)$ та знайти оптимальне раціональне наближення $\alpha_{opt} = p_{opt} / q_{opt}$ числа α , після чого визначити значення K_{opt} із співвідношення

$$K_{opt} = \lfloor K / q_{opt} \rfloor \cdot q_{opt}. \quad (3.17)$$

Оскільки процес розкладання в ланцюговий дріб досить громіздкий, то в більшості випадків достатньо знайти субоптимальне значення \tilde{K}_{opt} , за яким розходження спектра буде якщо не мінімальним, але досить малим. Таке число \tilde{K}_{opt} можна знайти із співвідношення

$$\tilde{K}_{opt} = \lfloor \lfloor K\alpha \rfloor / \alpha \rfloor + 1, \quad (3.18)$$

де $\alpha = \omega\Delta / (2\pi)$.

Зрозуміло, що за невідомої частоти ω неможливо скористатися виразами (3.17) або (3.18) та аналітично визначити K_{opt} або \tilde{K}_{opt} , тому наведені формули є лише теоретичним обґрунтуванням принципу роботи запропонованої пошукової процедури, суть якої полягає в такому.

У загальному випадку передбачається, що перешкода $h_g[\cdot]$ зосереджена на певній фіксованій частоті $f_g \in [f_g^{\min}, f_g^{\max}]$, але в результаті розходження спектра її потужність розподілена на цілу низку суміжних частот. Тому, якщо для фіксованої частоти f_g ослаблювати ефект розходження її спектра, то це неминуче призводить до збільшення амплітуди спектральної компоненти ДПФ, частота якої близька до f_g .

Звідси випливає, що якщо побудувати сімейство спектральних щільностей $S_K\{\cdot\}$, $S_{K-1}\{\cdot\}$, $S_{K-2}\{\cdot\}$,... для різних довжин масиву K , $K-1$, $K-2$,..., що поступово зменшуються від початкової довжини K , то серед спектральних щільностей $S_K\{\cdot\}$, $S_{K-1}\{\cdot\}$, $S_{K-2}\{\cdot\}$,... знайдеться така, за якою спектральна складова, що відповідає частоті f_g перешкоди, буде мати найбільше значення. Знайдена таким чином спектральна щільність і визначить оптимальне значення K_{opt} .

На рис. 3.7 показано приклади спектральних щільностей, побудованих на основі ДПФ для гармонічного коливання

$$h[k] = \cos(2\pi f_g k \Delta), \quad k = 0, \dots, K-1 \quad (3.19)$$

із частотою $f_g = 20,127$ Гц ($\Delta = 0,001$ с) для різних значень довжини K оброблюваного масиву.

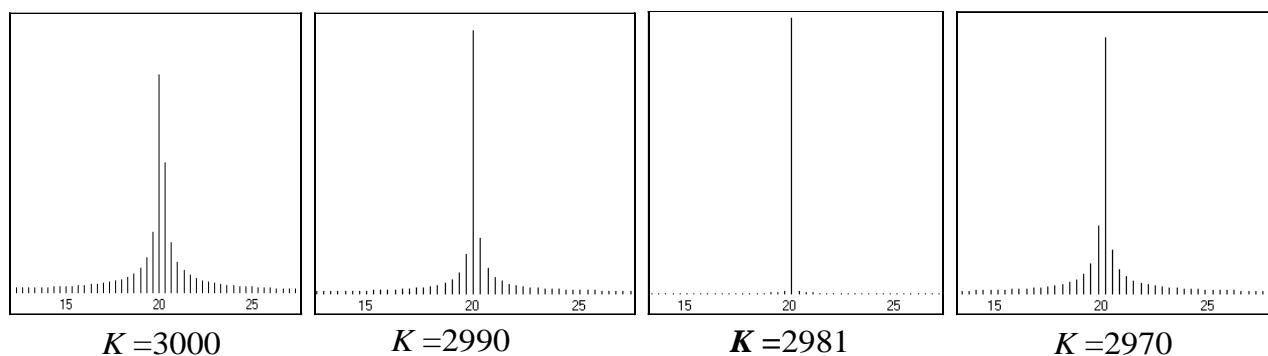


Рис. 3.7. Спектральні щільності сигналу (3.19) для різних K

Можна бачити, що за поступового зменшення кількості дискретних значень від $K = 3000$ до $K = 2970$ ефект розходження спектра спочатку слабшає, а потім знову посилюється. Найбільшого значення амплітуда спектральної щільності набуває для оптимального значення $K_{opt} = 2981$.

Оскільки найчастіше на практиці частота гармонічної перешкоди, або, що те саме, колова частота ω , невідома, то для досягнення поставленої мети пропонується використовувати пошукову процедуру автоматичного підбору оптимального числа $K_{opt} \leq K$.

Для прискорення пошуку K_{opt} слід врахувати обмеження f_g^{\min} , f_g^{\max} інтервалу частот, на якому зосереджена гармонічна перешкода, а також задати деяке прийнятне зменшення кількості точок δ_K в оброблюваному масиві. В результаті отримуємо такий алгоритм.

Крок 1. Послідовно скорочуємо довжину масиву z , зменшуючи число K_j оброблюваних точок від початкового значення K до $K - \delta_K$.

Крок 2. Використовуючи процедуру (3.5) прямого ДПФ, будуємо сімейство спектральних щільностей $S_K\{\cdot\}$, $S_{K-1}\{\cdot\}$, $S_{K-2}\{\cdot\}$, ... для різних значень K_j , тобто обчислюємо спектральні компоненти

$$C_n(K_j), n = -K_j / 2, \dots, 0, \dots, K_j / 2$$

для кожного $K_j \in [K - \delta_K, K]$.

Крок 3. За сімейством побудованих спектральних щільностей визначаємо оптимальне значення K_{opt} , що задовольняє умову

$$K_{opt} = \arg \max_{K_j \in [K-\delta_K, K]} \left\{ \frac{\max_{n \in \Omega_h} C_n(K_j)}{\sum_{n \in \Omega_h} C_n(K_j)} \right\}, \quad (3.20)$$

де Ω_h – множина номерів спектральних компонент із частотами, що належать інтервалу $[f_g^{\min}, f_g^{\max}]$.

Крок 4. Як оцінку частоти \hat{f}_g гармонічної перешкоди беремо значення максимальної спектральної компоненти на інтервалі $[f_g^{\min}, f_g^{\max}]$, обчислену за K_{opt} .

Крок 5. Модифікуємо спектр, прирівнюючи нулю компоненти спектральної щільності, що відповідають знайденій частоті \hat{f}_g та симетричній («негативній») частоті $-\hat{f}_g$.

Крок 6. Використовуючи процедуру (3.6) зворотного ДПФ, за модифікованим спектром відновлюємо корисний сигнал у часовій області.

Зауважимо, що розглянутий алгоритм приглушення зосередженої гармонічної перешкоди можна реалізувати дещо іншим способом, коли дискретний масив $z[1], z[2], \dots, z[K]$ доповнюється нулями до деякої більшої довжини $K + \delta_K$ та проводиться аналогічний аналіз сімейства спектральних щільностей $S_K\{\cdot\}, S_{K+1}\{\cdot\}, S_{K+2}\{\cdot\}, \dots, S_{K+\delta_K}\{\cdot\}$.

Покажемо ефективність запропонованого алгоритму (рис. 3.8) на прикладі фільтрації реальної ЕКГ (рис. 3.8, а). На рисунку зображено ЕКГ, яка суттєво спотворена в результаті накладання гармонічної перешкоди з частотою 16,68 Гц з амплітудою, що становить 50 % діапазону зміни (рис. 3.8, б). Проте запропонований алгоритм фільтрації дозволив ефективно приглушити перешкоду, повністю зберігши форму корисного сигналу (рис. 3.8, в).

Пояснимо запропонований алгоритм, проаналізувавши спектральні щільності сигналу та перешкоди у частотній області (рис. 3.9).

Характерна риса корисного сигналу (ЕКГ) у тому, що його енергію розподілено на цілу низку гармонічних компонент (рис. 3.9, а). Інакше кажучи, гармонічне розкладання корисного сигналу є сумою

великої кількості гармонік, які перебувають фактично на всій області частот від 0 до 85 Гц.

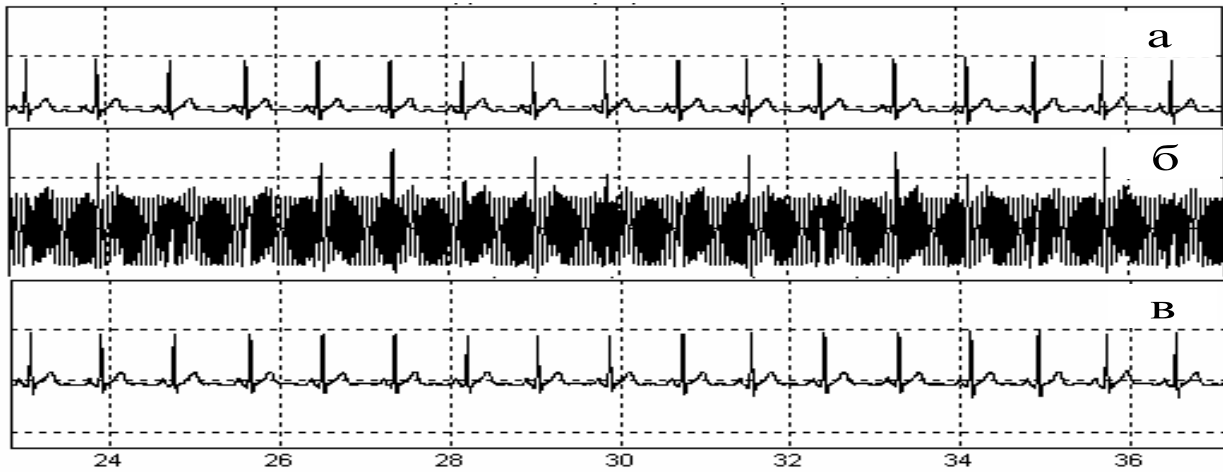


Рис. 3.8. Реальна ЕКГ (а), спотворений сигнал (б) та результат фільтрації (в)

У цій області розміщено і гармонічну перешкоду з частотою 16,68 Гц. Внаслідок розходження спектра основну потужність перешкоди розподілено на інтервалі 12–23 Гц (рис. 3.9, б), причому у цьому діапазоні амплітуда гармонічної перешкоди значно вища, ніж амплітуда гармонічних компонент корисного сигналу.

Вихідна довжина масиву, що обробляється, становить $K = 30$ тис. дискретних значень. Згідно з формулою (3.5) зменшення довжини оброблюваного масиву від K до $K - \delta_K$ точок призводить до зміни кроку Δ_f базових частот f_n спектральних компонентів ДПФ, які апроксимують оброблюваний сигнал.

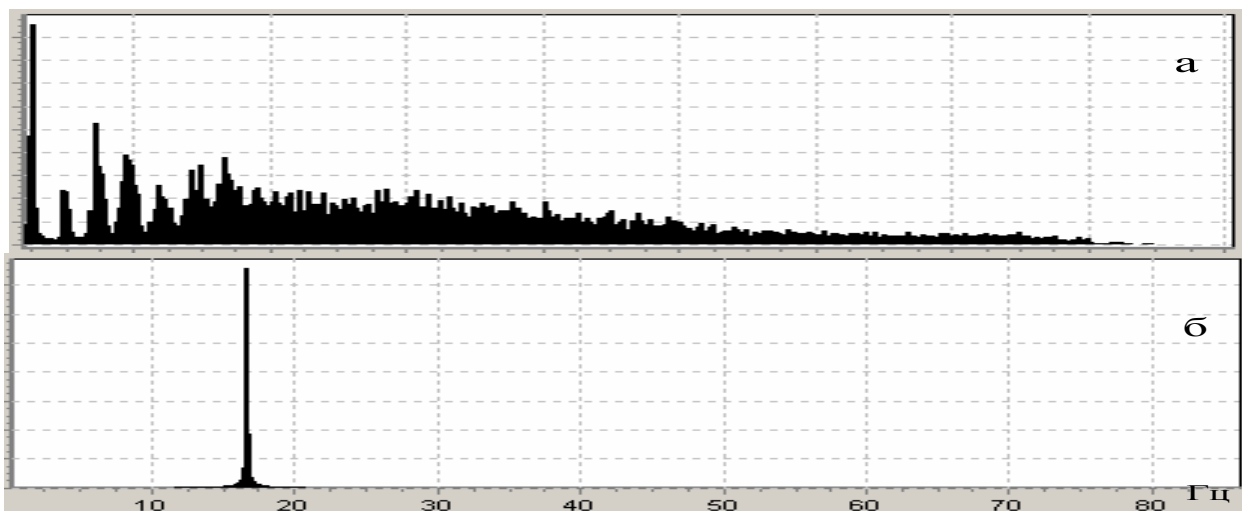


Рис. 3.9. Спектри ЕКГ (а) та гармонічної перешкоди 16,48 Гц (б)

Оскільки гармонічна перешкода є адитивною, а перетворення Фур'є – лінійною, можна окремо розглядати спектри корисного сигналу та перешкоди. Зрозуміло, що поступове зменшенні довжини масиву призводить до перерозподілу навантажених частот як перешкоди, так і корисного сигналу (рис. 3.10).

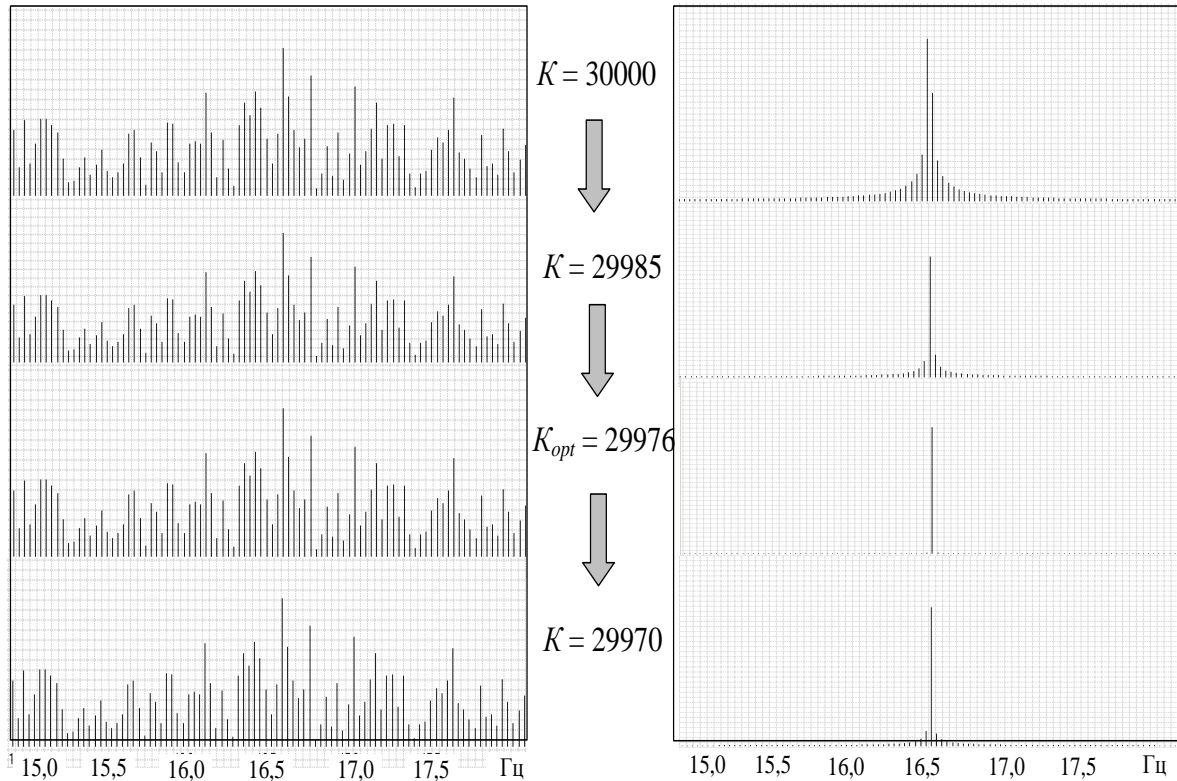


Рис. 3.10. Сімейство спектрів ЕКГ корисного сигналу (ліворуч)

Перерозподіл навантажених частот спектра корисного сигналу (рис. 3.10, ліворуч) не призводить до яких-небудь непередбачуваних наслідків, оскільки через оборотність ДПФ завжди можна точно відновити корисний сигнал за скороченим масивом. Головний ефект запропонованого алгоритму зводиться до зменшення ефекту розходження спектра гармонічної перешкоди.

Як видно, довжина масиву зменшилася до значення $K_{opt} = 29976$ точок, ефект розходження виявився мінімальним (рис. 3.10 праворуч). Це дає можливість скоротити необхідну смугу редуції (аж до заглушення лише однієї спектральної компоненти ДПФ, що збігається з частотою перешкоди) і тим самим зменшити кількість компонентів корисного сигналу, що пригнічуються, а значить і його спотворення.

Нагадаємо, що для використання запропонованого алгоритму достатньо вказати лише інтервал $[f_g^{\min}, f_g^{\max}]$ пошуку гармонічної перешкоди, дійсна частота якої визначається автоматично.

Кількісне оцінювання ефективності запропонованого алгоритму фільтрації проводилася на модельних сигналах різної стандартної форми.

Для ілюстрації наведемо результат фільтрації меандра, що генерувався у 30 тис. точках з частотою дискретизації $F_D=1000$ Гц і був спотворений гармонічною перешкодою з частотою 18,1 Гц різних рівнів – 20 %, 50 % і 100 % діапазону зміни корисного сигналу (рис. 3.11).

Експерименти показали, що навіть під час суттєвої перешкоди вдалося майже без спотворень відновити форму настільки складного сигналу. Так, зокрема, після фільтрації перешкоди 100 % рівня розбіжність фільтрованого сигналу відносно еталону не перевищувало 0,22 % за всією областю визначення.

Ефективність запропонованого алгоритму експериментально підтверджено і за фільтрації інших сигналів, зокрема, для розділу двох близьких за частотою гармонічних коливань.

На рис. 3.12 наведено результат одного з таких модельних експериментів. Модельний сигнал генерувався в 30 тис. точок з частотою дискретизації 1000 Гц і був сумою двох гармонічних коливань з частотами 16,0123 Гц (корисний сигнал) і 16,5123 Гц (перешкода). Обидві гармоніки мали однакові амплітуди та нульові початкові фази. Зауважимо, що різниця частот корисного сигналу і перешкоди відповідає 0,1 % частоти Найквіста.

Порівняння фільтрованого та корисного (еталонного) сигналу показало, що майже по всій області визначення відмінності їх значень не перевищували 2 % діапазону зміни сигналів. Фактично такий самий результат отримано і під час поділу гармонічних коливань з близькими високими частотами 250,123456789 Гц і 251,123456789 Гц, що наближалися до частоти Найквіста.

Для реалізації описаного пошукового алгоритму вузькосмугової режекторної фільтрації використано процедуру швидкого перетворення Фур'є М. Фріго (**M. Frigo**) та С. Джонсона (**S. Johnson**), яка визнана однією з найкращих.

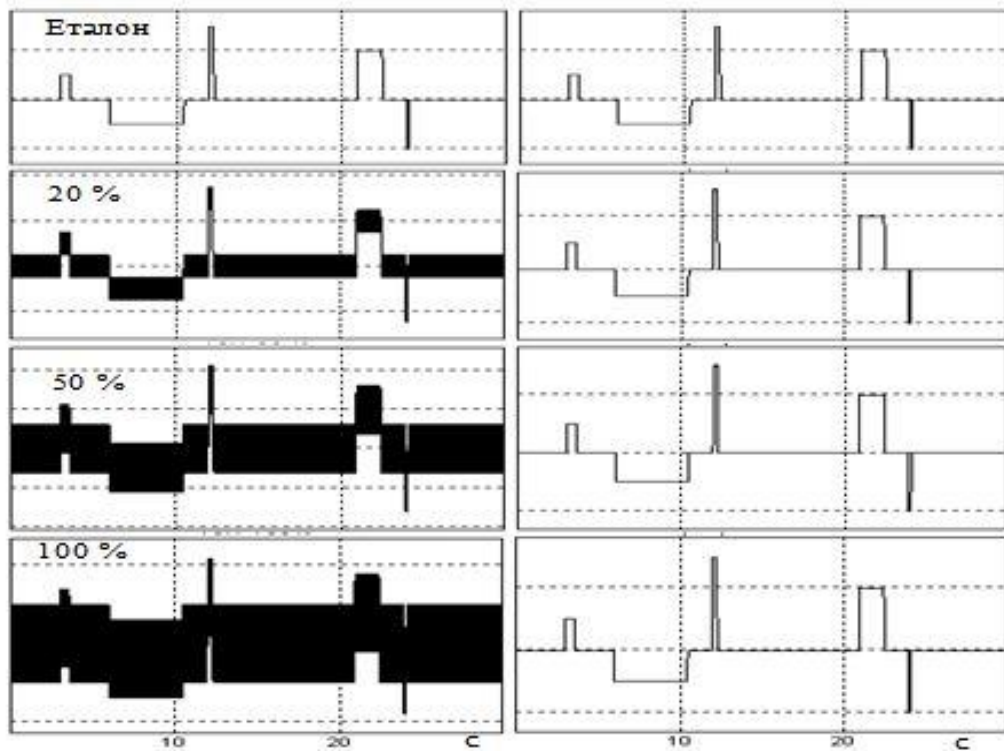


Рис. 3.11. Результати фільтрації меандра до фільтрації (ліворуч) та після фільтрації (праворуч)

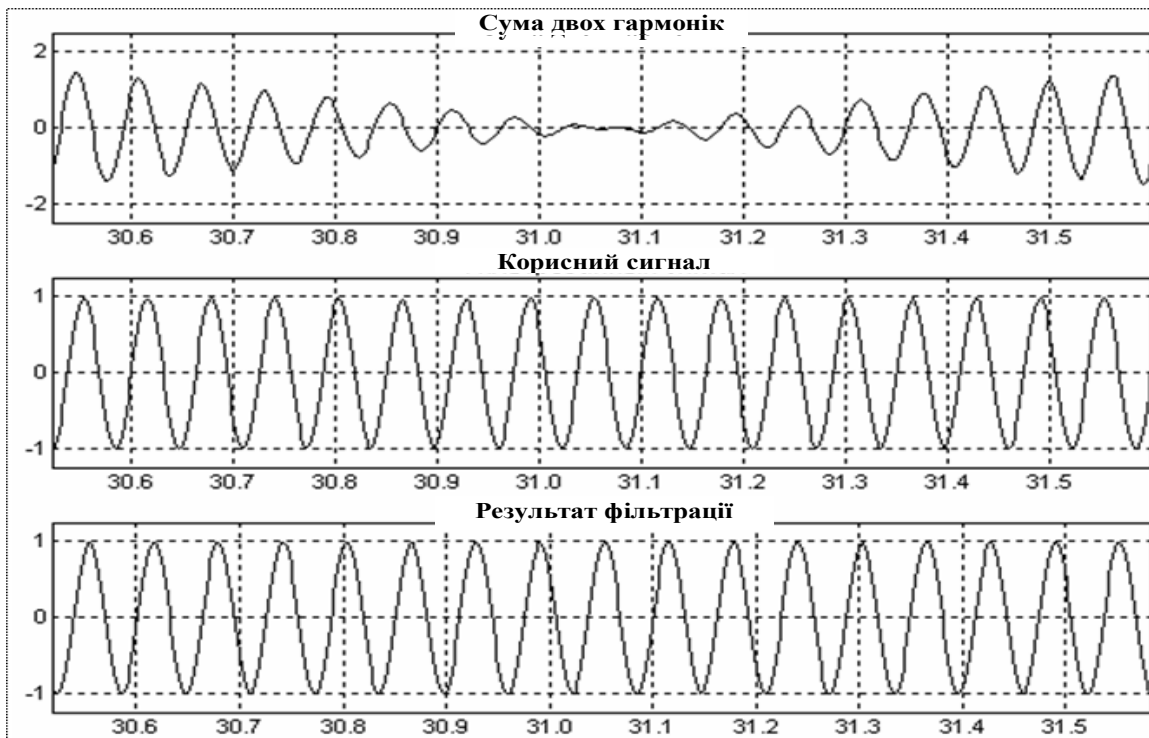


Рис. 3.12. Результат поділу двох гармонічних сигналів із близькими частотами

3.3. Адаптивна фільтрація випадкових завад

Розглянемо тепер випадок, коли адитивна перешкода має випадкову природу. У цьому випадку модель перешкоди подамо послідовністю $h[k]$, $k = 1, 2, \dots, K$ незалежних випадкових величин, які за невідомим законом розподілені на фіксованому інтервалі $[-h_0, h_0]$, причому $M\{h\} = 0$, $M\{h[i]h[j]\} = 0$, $i \neq j$, де $M\{\cdot\}$ – знак математичного сподівання, а $h_0 > 0$ – константа, що обмежує рівень перешкоди.

Зрозуміло, що для заглушення випадкового шуму неефективно застосовувати частотно-виборчі фільтри, оскільки в загальному випадку спектр такої завади $h[k]$ розподілений за всією смугою частот від нуля до частоти Найквіста.

Один з відомих підходів для оцінювання корисного сигналу $\tilde{z}_0[k]$ за спостереженням $z[k] = z_0[k] + h[k]$ полягає у використанні алгоритмів згладжування даних, серед яких використовують метод експоненційного згладжування та метод ковзного середнього.

Найпростіший метод експоненційного згладжування (**Simple Exponential Smoothing**) дає оцінку $\tilde{z}_0[k]$ у такому вигляді

$$\tilde{z}_0[k] = \alpha z[k] + (1 - \alpha) \tilde{z}_0[k - 1], \quad (3.21)$$

де $\alpha = \text{const}$, $0 < \alpha < 1$ – параметр згладжування.

Рекурентне співвідношення (3.21) можна подати в еквівалентній формі

$$\tilde{z}_0[k] = \tilde{z}_0[k - 1] + \alpha(z[k] - \tilde{z}_0[k - 1]).$$

Звідки випливає, що оцінка згладженого значення k -й момент визначається оцінкою, отриманою в $k - 1$ -й момент часу, додавання частка α різниці між поточним спостереженням $z[k]$ та оцінкою $\tilde{z}_0[k - 1]$ в попередній момент часу.

Якщо послідовно використовувати рекурентне співвідношення (3.21), то оцінку $\tilde{z}_0[k]$ можна висловити через значення часового ряду:

$$\begin{aligned} \tilde{z}_0[k] = & \alpha z[k] + \alpha(1 - \alpha)z[k - 1] + \alpha(1 - \alpha)^2 z[k - 2] + \dots + \\ & + (1 - \alpha)^i z[k - i] + \dots + (1 - \alpha)^{W_0} \tilde{z}_{00} = \alpha \sum_{i=0}^{W_0-1} (1 - \alpha)^i z[k - i] + (1 - \alpha)^{W_0} \tilde{z}_{00}, \end{aligned}$$

де W_0 – кількість членів ряду, а \tilde{z}_{00} – величина, що характеризує початкові умови для першого застосування формули (3.21), коли $k = 1$.

Оскільки $0 < \alpha < 1$, то $(1 - \alpha)^{W_0} \rightarrow 0$, якщо $W_0 \rightarrow \infty$, а значить

$$\tilde{z}_0[k] = \alpha \sum_{i=0}^{\infty} (1 - \alpha)^i z[k - i]. \quad (3.22)$$

Таким чином, оцінка $\tilde{z}_0[k]$ – це зважена сума всіх попередніх членів часового ряду з вагами, що експоненційно зменшуються.

Якщо припустити, що $z_0[k] \equiv z_0 = \text{const}$, а $h[k]$ – послідовність незалежних випадкових величин з нульовим середнім та дисперсією σ^2 , то з (3.22) випливає, що дисперсія оцінки визначається співвідношенням

$$D\{\tilde{z}_0\} = M\{(\tilde{z}_0 - z_0)^2\} = M\left\{\left(\alpha \sum_{i=0}^{\infty} (1 - \alpha)^i h[k - i]\right)^2\right\} = \frac{\alpha}{2 - \alpha} \sigma. \quad (3.23)$$

Оскільки $0 < \alpha < 1$, то на підставі (3.23) можна зробити висновок, що чим менше α , тим більше скорочується дисперсія оцінки \tilde{z}_0 .

Відомо кілька різновидів методу експоненційного згладжування, зокрема двопараметричний метод Холта (**Holt's method**):

$$\begin{aligned} \tilde{z}_0[k] &= \alpha z[k] + (1 - \alpha)(\tilde{z}_0[k - 1] + t[k - 1]), \\ t[k] &= \beta(y[k] - z[k - 1]) + (1 - \beta)t[k - 1], \end{aligned}$$

де β – параметр згладжування тренду, а також трипараметричний метод Вінтерса (**Winter's method**):

$$\begin{aligned} \tilde{z}_0[k] &= \alpha(z[k] / s[k - p]) + (1 - \alpha)(\tilde{z}_0[k - 1] + t[k - 1]), \\ t[k] &= \beta(\tilde{z}_0[k] - \tilde{z}_0[k - 1]) + (1 - \beta)t[k - 1], \\ s[k] &= \gamma(z[k] / \tilde{z}_0[k]) + (1 - \gamma)t[k - 1], \end{aligned}$$

у якому є додаткова константа γ згладжування циклів із періодом p .

Інший популярний метод – алгоритм ковзного середнього (**Moving Average**), який дає змогу згладжування спотворених даних відповідно до формули

$$\tilde{z}_0[k] = \frac{1}{W_0} \sum_{j=0}^{W_0-1} z[k - j], \quad (3.24)$$

де $W_0 = \text{const}$ – ширина вікна згладжування. Можна побачити, що ступінь згладжування збільшується із збільшенням W_0 , причому $D\{\tilde{z}_0\} = \sigma^2 / W_0$.

Оскільки

$$\sum_{j=0}^{W_0-1} z[k-j] = \sum_{j=0}^{W_0-1} z[k-1-j] + z[k] - z[k-W_0],$$

то формулу (3.24) можна перетворити до рекурентної форми

$$\tilde{z}_0[k] = \tilde{z}_0[k-1] + \lambda(y[k]) - z[k-W_0], \quad (3.25)$$

де $\lambda = 1/W_0$.

Перевага формули (3.5) порівняно з (3.24) очевидна – швидкість роботи алгоритму не залежить від ширини вікна згладжування W_0 .

Експерименти показали, що, незважаючи на відмінність формул (3.21) і (3.24), завдяки налаштуванню параметрів ці процедури можуть забезпечувати майже однаковий ступінь згладжування спотворених даних.

Зауважимо, що формули (3.24) та (3.25) використовують лише попередні дискретні значення сигналу, тому ними можна скористатися в системах реального часу, що є їхньою безперечною перевагою.

У тих самих випадках, коли допустимо оброблювати відразу весь масив $z[1], \dots, z[K]$ для оцінювання згладженого значення $\tilde{z}[k]$ доцільно використовувати як попередні, так і наступні значення сигналу. У цьому випадку процедуру ковзного середнього можна записати у вигляді

$$\tilde{z}_0[k] = \frac{1}{2W_0 + 1} \sum_{j=-W_0}^{W_0} z[k-j]. \quad (3.26)$$

Після очевидних перетворень формулу (3.26) можна подати у рекурентній формі

$$\tilde{z}_0[k] = \tilde{z}_0[k-1] + \mu(z[k+W_0]) - z[k-1-W_0], \quad (3.27)$$

де $\mu = 1/(1 + 2W_0)$. Зрозуміло, що процедура (3.26) дає змогу оцінювати згладжені значення сигналу, коли

$$W_0 \leq k \leq K - W_0.$$

Зауважимо, що під час згладжування згідно з (3.26), усі дискретні значення сигналу у вікні $1 + 2W_0$ усереднюються рівномірно. Таку процедуру можна розглядати як дискретну згортку сигналу, що обробляється з ядром прямокутної форми.

Однак можна скористатися й іншими симетричними ядрами, зокрема, відомими вікнами Бартлетта (**Bartlett window**), Гауса (**Gaussian window**), Хеммінга (**Hamming window**), Блакмана (**Blackman window**) та іншими. У цьому випадку вага дискретних значень більш далеких від центру вікна поступово зменшується, наближаючись до нуля на межах вікна.

Покажемо, що аналогічного ефекту можна досягти багаторазовим застосуванням процедури (3.26). Для простоти припустимо, що вікно згладжування $1 + 2W_0$ дорівнює трьом точкам. Тоді, згідно з (3.26), на першому етапі ітерації згладжені значення сигналу в моменти часу $k - 1$, k і $k + 1$ визначаються виразами

$$\tilde{z}_0^{(1)}[k - 1] = \frac{1}{3} z[k - 2] + \frac{1}{3} z[k - 1] + \frac{1}{3} z[k],$$

$$\tilde{z}_0^{(1)}[k] = \frac{1}{3} z[k - 1] + \frac{1}{3} z[k] + \frac{1}{3} z[k + 1],$$

$$\tilde{z}_0^{(1)}[k + 1] = \frac{1}{3} z[k] + \frac{1}{3} z[k + 1] + \frac{1}{3} z[k + 2].$$

Застосовуючи повторно процедуру (3.26) до отриманих величин $\tilde{z}_0^{(1)}[k - 1]$, $\tilde{z}_0^{(1)}[k]$, $\tilde{z}_0^{(1)}[k + 1]$, оцінимо згладжене значення $\tilde{z}_0^{(2)}[k]$ на другому етапі ітерації:

$$\begin{aligned} \tilde{z}_0^{(2)}[k] &= \frac{1}{3} \tilde{z}_0^{(1)}[k - 1] + \frac{1}{3} \tilde{z}_0^{(1)}[k] + \frac{1}{3} \tilde{z}_0^{(1)}[k + 1] = \\ &= \frac{1}{9} z[k - 2] + \frac{2}{9} z[k - 1] + \frac{1}{3} z[k] + \frac{2}{9} z[k + 1] + \frac{1}{9} z[k + 2]. \end{aligned}$$

Звідси видно, що тепер уже вагова функція стає трикутною.

Слід звернути увагу на один загальний недолік розглянутих вище традиційних процедур згладжування, який продемонструємо модельним прикладом.

Нехай сигнал, що спостерігається (ЕКГ), спотворено 20 % випадковою перешкодою з нульовим математичним сподіванням (рис. 3.13, а). Для такого сигналу процедура (3.26) з вікном $W_0 = 7$ не дає змогу отримати необхідне заглушення шумового компонента на інформативному фрагменті ST (рис. 3.13, б).

Тільки якщо ширина вікна $W_0 = 30$ згладжування призводить до бажаного скорочення шумового компонента на фрагменті ST (рис. 3.13, в). Проте в цьому випадку відбувається неприпустиме зниження амплітуди іншого інформативного фрагмента – комплексу QRS , аналіз якого також має діагностичну цінність.

Таким чином, під час використання традиційного алгоритму ковзного середнього виникає проблема протиріч вибору оптимального значення параметра згладжування W_0 : якщо зменшувати W_0 , то не забезпечується бажаний ефект заглушення шумового компонента, а якщо збільшувати W_0 , то відбуваються неприпустимі спотворення корисного сигналу.

Аналогічна суперечність виникає і з вибором оптимального значення параметра α алгоритму експоненційного згладжування (3.21).

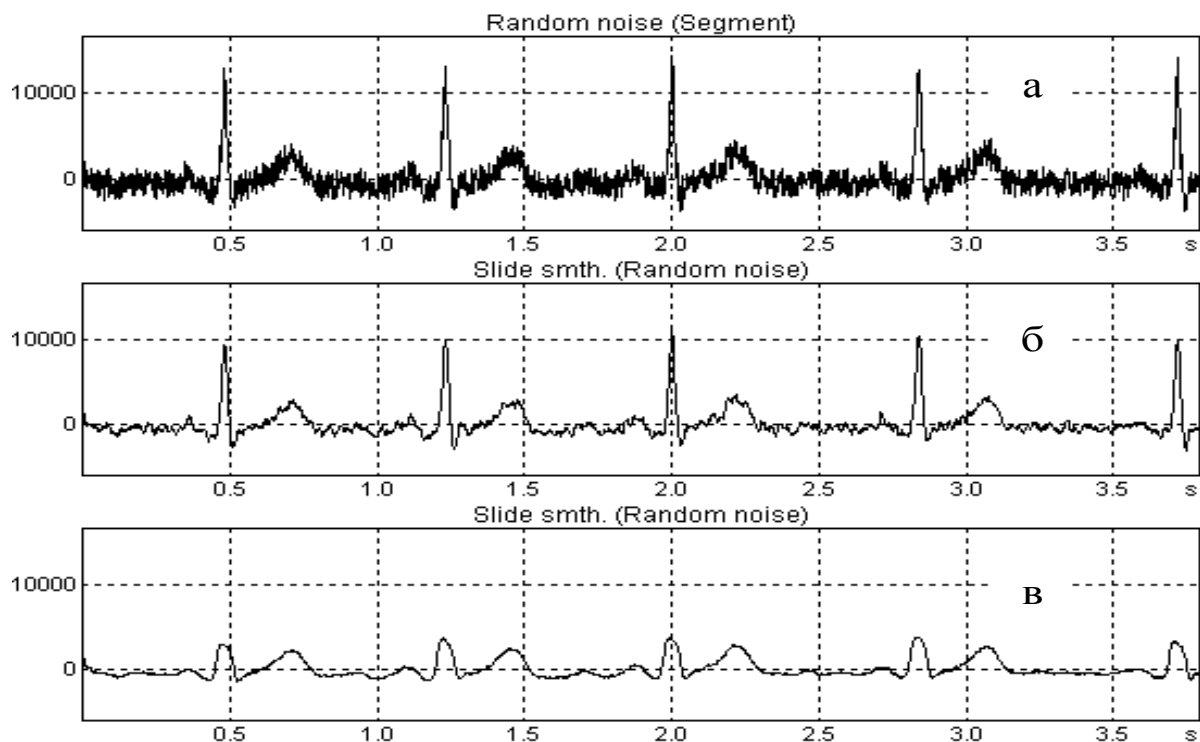


Рис. 3.13. Результати згладжування спотвореної ЕКГ:

а – вихідний сигнал; б – вікно $W_0 = 7$ точок; в – вікно $W_0 = 30$ точок

Таким чином, під час оброблення сигналів з локально зосередженими ознаками традиційні алгоритми згладжування можна використовувати лише за дуже малих рівнів шумової компоненти, що обмежує сферу їх практичного використання.

Для усунення зазначеного недоліку запропоновано адаптивний алгоритм ковзного середнього, основна ідея якого полягає в такому.

Оскільки передбачається, що рівень випадкової перешкоди $h(k)$ обмежений за модулем величиною h_0 , тобто

$$h(k) \in [-h_0, h_0] \quad \forall k = 1, \dots, K, \quad (3.28)$$

то будь-яке відхилення згладженого значення $\hat{z}_0[k]$ від спостережуваного $z[k]$, що перевищує h_0 , можна пояснити лише тим, що сигнал надмірно згладжений.

Тому пропонується пристосувати (адаптувати) параметр W до оброблюваного сигналу таким чином, щоб, з одного боку, забезпечити максимальний ступінь згладжування, а з другого боку, не допустити спотворення сигналу, що перевищує заданий поріг h_0 .

Формально запропонований алгоритм реалізується формулою

$$\hat{z}_0[k] = \frac{1}{2W_k + 1} \sum_{j=-W_k}^{W_k} z[k-j], \quad (3.29)$$

яка відрізняється від (3.26) лише тим, що кількість точок у вікні, що усереднюються, залежить від номера k оброблюваної точки.

Значення параметра W_k , що використано в (3.29), визначається за умов

$$W_k \leq W_0, \quad (3.30)$$

$$|\hat{z}[k] - z[k]| \leq h_0, \quad (3.31)$$

$$|W_k - W_{k-1}| \leq 1, \quad (3.32)$$

де W_0 – обмеження на параметр згладжування, що задається користувачем.

Алгоритм передбачає два етапи оброблення масиву дискретних значень спотвореного сигналу. На першому етапі визначаються оптимальні значення W_1, \dots, W_K для кожної точки сигналу $z[k]$, $k = 1, \dots, K$, а на другому – згладжування $z[k]$ з використанням знайдених оптимальних параметрів W_1, \dots, W_K .

Процедура визначення W_k , $k = 1, \dots, K$ реалізується в такий спосіб. Для заданого значення W_0 обирається перший блок $1 + 2W_0$ оброблю-

ваних точок з порядковими номерами $j = 1, 2, \dots, 1 + 2W_0$ та оцінюється згладжене значення у точці з порядковим номером $i = W_0$ за формулою

$$\hat{z}[i] = \frac{1}{2W_j + 1} \sum_{j=-W_i}^{W_i} z[i + j], \quad (3.33)$$

з початковим значенням параметра згладжування $W_i = W_0$.

Якщо виконується умова

$$|z[i] - \hat{z}[i]| > h_0, \quad (3.34)$$

то значення W_i зменшується на одиницю, тобто $W_i = W_0 - 1$ і формула (3.29) застосовується до масиву точок із порядковими номерами $j = 1, 2, \dots, 2W_0$, тобто перший блок зменшується на одну точку.

Описаний ітераційний процес скорочення довжини оброблюваного блоку триває доти, доки вперше не порушиться умова (3.34). Як тільки це трапиться, обирається наступний блок $1 + 2W_0$ точок із порядковими номерами $j = 2, \dots, 2 + 2W_0$, які обробляються аналогічно і т. д. до кінця масиву.

У результаті для кожної k -ї точки обирається значення $W_k^{(1)}$, яке задовольняє умови (3.30) та (3.31).

Остаточним кроком адаптації параметрів згладжування є модифікація знайдених значень

$$W_k^{(1)} \rightarrow W_k^{(2)} \rightarrow W_k^{(3)},$$

що забезпечує виконання умови (3.32) «плавності» зміни вікон у процесі оброблення сигналу. Розглянемо механізм такої модифікації на прикладі.

Припустимо, наприклад, що $W_0 = 30$ і на першому етапі отримано таку послідовність параметрів згладжування (вікон):

$$W_k^{(1)} = \{25, 28, 30, 15, 18, 23, 27, 26, 30, 29, 18, 24, 30, \dots, 24, 28, 26\},$$

які задовольняють умови (3.26) та (3.27). Тоді перехід від $W_k^{(1)}$ до $W_k^{(2)}$, що забезпечує «правосторонню плавність» значень, здійснюється так:

$$W_{i+1}^{(2)} = \begin{cases} W_i^{(1)} + 1, & \text{коли } W_{i+1}^{(1)} > W_i^{(1)} \\ W_{i+1}^{(1)} & \text{в іншому випадку} \end{cases}.$$

У результаті масив параметрів набуває вигляду

$$W_k^{(2)} = \{25, 26, 27, 15, 16, 17, 18, 19, 20, 21, 18, 19, 20, \dots, 24, 25, 26\}.$$

Далі здійснюється модифікація $W_k^{(2)}$, що забезпечує «лівосторонню плавність»:

$$W_{i-1}^{(3)} = \begin{cases} W_i^{(2)} + 1, & \text{коли } W_{i-1}^{(2)} > W_i^{(2)} \\ W_{i-1}^{(2)} & \text{в іншому випадку} \end{cases}.$$

У результаті формується кінцевий масив $W_k^{(3)}$ оптимальних значень параметрів згладжування з «двосторонньою» плавністю:

$$W_k^{(3)} = \{18, 17, 16, 15, 16, 17, 18, 19, 20, 19, 18, 19, 20, \dots, 24, 25, 26\},$$

який далі використовується для згладжування спотвореного сигналу.

Для оцінювання ефективності запропонованого алгоритму проводились експерименти на модельних та реальних сигналах згідно зі схемою, показаною на рис. 3.14.

Як еталони використовувалися одиночні типові сигнали стандартної форми (прямокутної, трикутної, пилкоподібної тощо) у вигляді одиночних імпульсів та їх послідовностей з параметрами, що змінюються в часі. Перешкода генерувалася як послідовність незалежних випадкових величин, рівномірно розподіленої на інтервалі $[-h_0, h_0]$, де h_0 задавали як відсоток від діапазону Δ_0 стандартного сигналу.

Результати порівнювались за максимальними $\Delta_{\max}^{(1)}$, $\Delta_{\max}^{(2)}$ та середньоквадратичними $\sigma^{(1)}$, $\sigma^{(2)}$ відхиленнями згладжених сигналів щодо еталонного в умовах використання традиційного та адаптивного алгоритмів, які наведено в табл. 3.2.



Рис. 3.14. Схема експерименту

Для ілюстрації на рис. 3.15 подано результати згладжування прямокутного імпульсу одиничної амплітуди, ширина якого становила $1/3$ області визначення сигналу. Сигнал генерувався в 1000 точках (рис. 3.15, а) і спотворювався 20-відсотковою перешкодою (рис. 3.15, б). Проводилося порівняння результатів застосування традиційного (рис. 3.15, в) та адаптивного алгоритмів згладжування (рис. 3.15, г).

Під час згладжування вузького прямокутного імпульсу (рис. 3.16, а), ширина якого становила всього 6 точок (0,6 % області визначення), рівень перешкоди, як і раніше, становив 20 % (рис. 3.16, б).

У цьому випадку переваги запропонованого алгоритму виявилися ще більш переконливішими: після згладжування традиційним алгоритмом (рис. 3.16, в) амплітуда імпульсу зменшилася більш ніж у п'ять разів, тоді як під час згладжування адаптивним алгоритмом (рис. 3.16, г) амплітуда короткого імпульсу фактично не змінилась.

Застосування процедури адаптивного згладжування до ЕКГ, що спостерігається на тлі 10-відсоткового шуму (рис. 3.17, а), дозволило отримати сигнал (рис. 3.17, б), форма і амплітудно-часові параметри якого фактично не спотворювалися. Це підтверджує суттєві переваги запропонованого алгоритму порівняно з традиційним, результати якого було наведено на рис. 3.13.

Таблиця 3.2. Результати експериментів

№	Еталон z_0			Перешкода h		Алгоритм згладжування			
	Модель	Δ_0	σ_0	%	σ_h	Традиційний		Адаптивний	
						$\Delta_{\max}^{(1)}$	$\sigma^{(1)}$	$\Delta_{\max}^{(2)}$	$\sigma^{(2)}$
1	Прямокутний (широкий)	1	0,47	20	0,058	0,49	0,08	0,07	0,01
2	Прямокутний (вузький)	1	0,08	20	0,056	0,84	0,08	0,09	0,01
3	Складний 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
4	Пилоподібне коливання	3,14	0,89	20	0,182	1,52	0,09	0,26	0,03
5	Трикутне коливання	3,14	0,86	20	0,181	0,15	0,03	0,11	0,03
6	Згасаюча синусоїда	1,57	0,22	5	0,023	0,033	0,008	0,021	0,007
7	Гаусові імпульси	1,27	0,18	25	0,121	0,22	0,06	0,11	0,03
8	Загасаючі гаусові імпульси	1,7	0,13	10	0,049	0,38	0,08	0,23	0,03

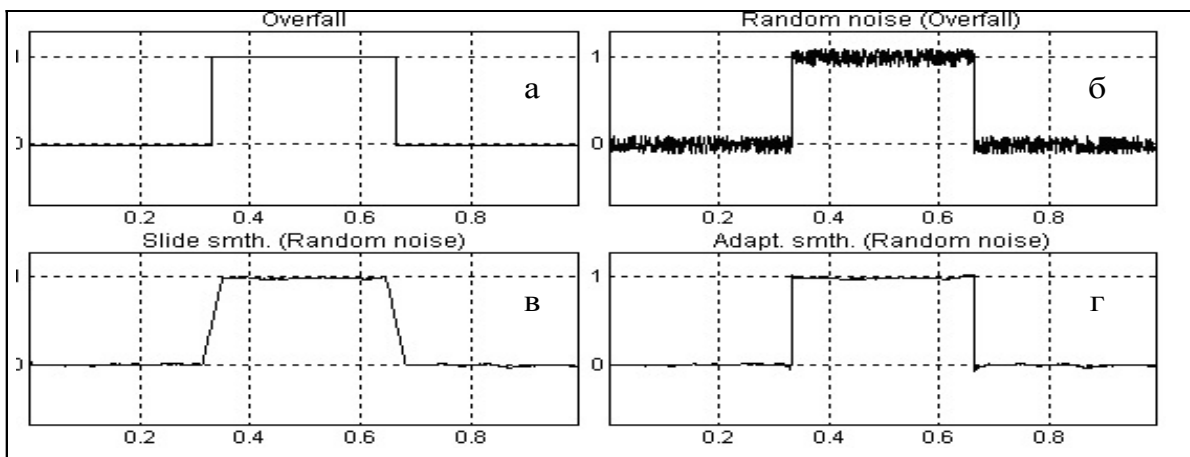


Рис. 3.15. Згладжування широкого прямокутного імпульсу

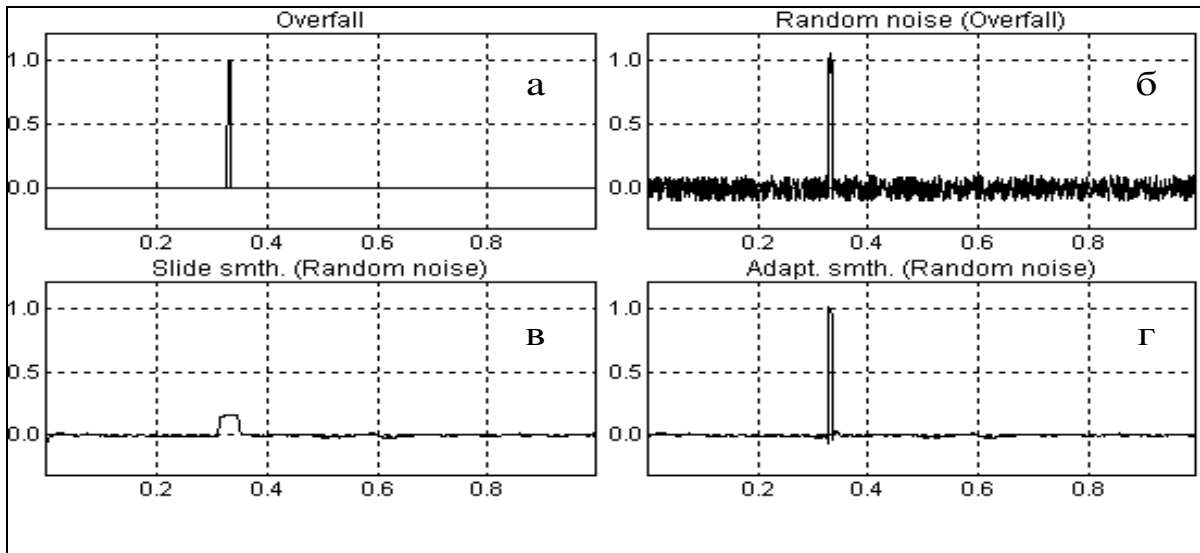


Рис. 3.16. Згладжування вузького прямокутного імпульсу

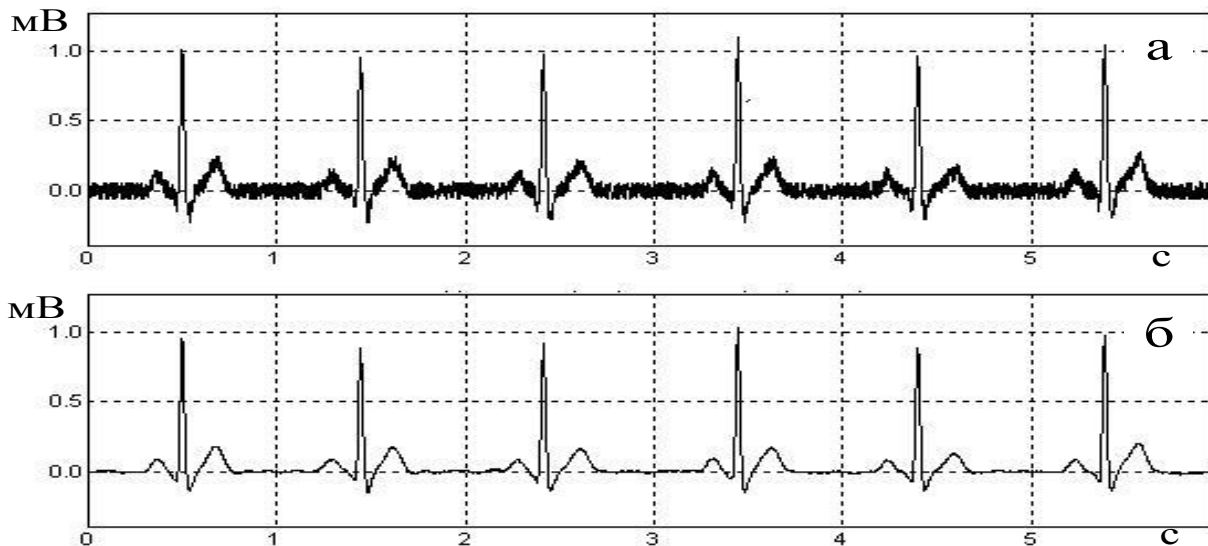


Рис. 3.17. Адаптивне згладжування спотвореної ЕКГ

3.4. Фільтрація завад іншого типу

Як зазначалося, в реальних умовах крім гармонічних (вузькосмугових) і випадкових (широкосмугових) завад трапляються перешкоди іншого типу – дрейф ізоелектричної лінії та імпульсні завади.

У першому наближенні дрейф ізоелектричної лінії можна вважати низькочастотною адитивною завадою. Однак видаляти дрейф на основі фільтрації верхніх частот неефективно, оскільки такий спосіб неминуче призводить до неприпустимо великих спотворень форм інформативних фрагментів сигналу, зокрема, сегмента *ST* та зубця *T*, які мають діагностичну цінність.

Зазвичай для заглушення дрейфу ізоелектричної лінії використовують інший спосіб, який полягає у виконанні таких кроків:

- автоматичне виявлення QRS -комплексів на сигналі;
- пошук вузлових точок на ділянці між зубцем R та початком комплексу QRS , які визначають рівень дрейфу ізоелектричної лінії;
- за зазначеними точками формується кусково-неперервна функція $\Psi(k)$ – кубічний сплайн, який оцінює тренд сигналу;
- побудовану функцію $\Psi(k)$ віднімають від сигналу $z(k)$, внаслідок чого сигнал $\tilde{z}(k) = z(k) - \Psi(k)$ очищується від дрейфу ізоелектричної лінії.

Незважаючи на те, що зазначений алгоритм набув досить широкої популярності, він має низку обмежень. Тому навіть така, на перший погляд, проста задача, як автоматичне виявлення комплексів QRS на реальних ЕКГ, викликає труднощі. Це пояснюється тим, що на практиці трапляється не менше 55 типів комплексу, які можуть суттєво відрізнитися формою (рис. 3.18).

Додаткові проблеми виникають тоді, коли амплітуди зубців R та T мають близькі значення, які змінюються від циклу до циклу. У таких випадках вузлові точки для побудови сплайну $\Psi(k)$ будуть знайдені неправильно, що знижує точність оцінювання дрейфу ізоелектричної лінії.

Також відомо, що процедура побудови сплайну сама по собі не така вже проста і може призводити до суттєвих обчислювальних помилок.

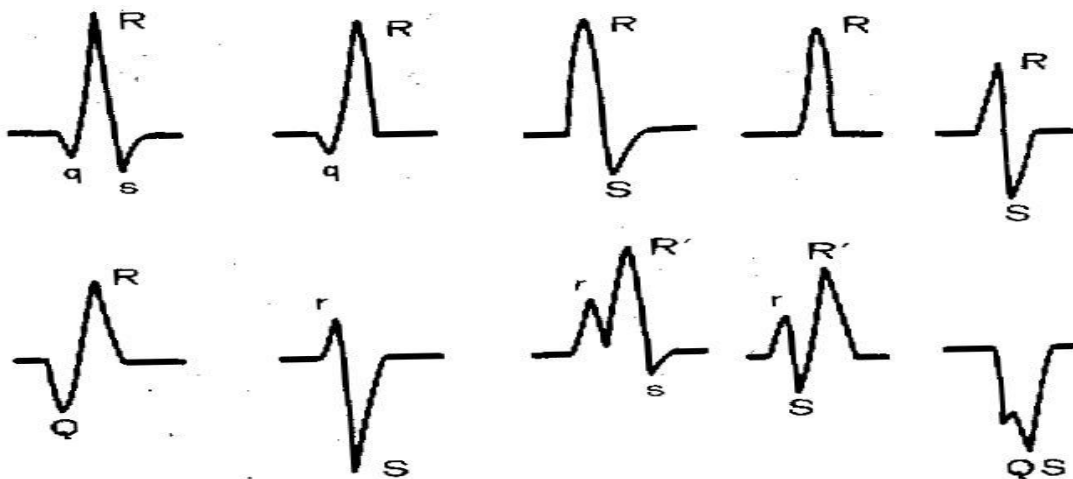


Рис. 3.18. Деякі типові форми комплексу QRS

Розглянемо інший підхід до розв'язку задачі усунення дрейфу ізоелектричної лінії, який є більш простим та, як показала практика, достатньо ефективним.

Будемо оцінювати дрейф ізоелінії безпосередньо за дискретними значеннями сигналу, що обробляється, на основі алгоритму ковзного середнього з великим вікном $W_B = \text{const}$. Рекурентна процедура

$$\psi[k] = \psi[k-1] + \frac{1}{W_B} (z[k] - z[k - W_B]), \quad k \in [W_B + 1, K] \quad (3.35)$$

у реальному часі оцінює тренд послідовності $z[W_B + 1], \dots, z[K]$, автоматично адаптуючись до цієї послідовності.

Використовуючи значення $\psi[k]$, отримані за формулою (3.35), послідовність $z[W_B + 1], \dots, z[K]$ модифікується таким чином:

$$\tilde{z}[k] = z[k] - \psi[k], \quad k \in [W_B + 1, K], \quad (3.36)$$

де

$$\psi[W_B] = \frac{1}{W_B} \sum_{k=1}^{W_B} z[W_B - k]. \quad (3.37)$$

На рис. 3.19 показано один із результатів тестування процедури (3.35)–(3.37).

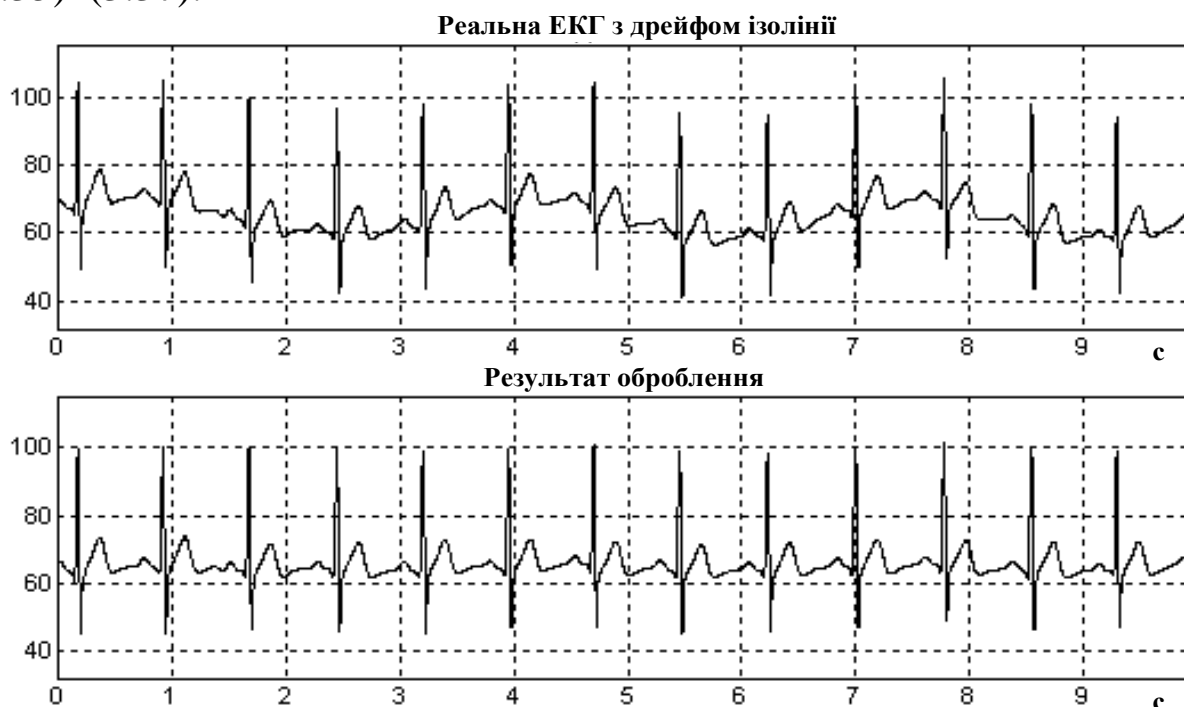


Рис. 3.19. Результат видалення дрейфу ізоелінії з ЕКГ

Зрозуміло, що ефективність процедури залежить від правильності вибору параметра W_B , який має забезпечити усереднення точок у вікні, ширина якого заздалегідь перевищує тривалість окремих циклів сигналу, що обробляється. Тому налаштування параметра W_B проводиться експериментально з використанням графічних засобів інструментальної системи, яку було розглянуто в розділі 1.

Імпульсна завада – ще один тип спотворення, яке часто трапляється під час реєстрації сигналів у реальних умовах. Така завада призводить до викидів (диких значень) сигналу у випадкові моменти часу $\tau = \tau_1, \tau_2, \dots, \tau_N$. Оскільки викиди можуть мати досить високу амплітуду, то дискретне значення сигналу $z[\tau] = z_0[\tau] + h[\tau]$ істотно відрізняється від значення корисного сигналу $z_0[\tau]$ в момент дії завади $h[\tau]$.

Вочевидь, нераціонально використовувати розглянуті вище алгоритми згладжування для заглушення імпульсних завад, тому що процедура усереднення даних не є робасною для такого типу завад, що призводить до спотворення корисного сигналу навколо моментів $\tau = \tau_1, \tau_2, \dots, \tau_N$ дії завади.

Найбільш проста та ефективна процедура заглушення поодиноких імпульсних завад необмеженого рівня – рангова фільтрація на основі обчислення медіани дискретних значень сигналу в ковзному вікні фіксованої ширини W_0 .

Фільтроване значення сигналу $\hat{z}[k]$ у k -й момент часу визначається як центральне значення (медіана) упорядкованого масиву $z[k-m], \dots, z[k], \dots, z[k+m]$ у вікні з непарною кількістю точок, яке рухається вздовж значень сигналу.

Для заглушення одиночних викидів, коли між моментами часу $\tau_{j-1}, \tau_j, \tau_{j+1}$ перешкоди немає, достатньо послідовно переглядати лише по три точки сигналу, поклавши $m = 1$. Тоді, після впорядкування масиву $z[k-1], z[k], z[k+1]$, викид «переміститься» в першу або третю точку.

Для ілюстрації на рис. 3.20 зображено один з результатів заглушення таких завад на ЕКГ сигналі.

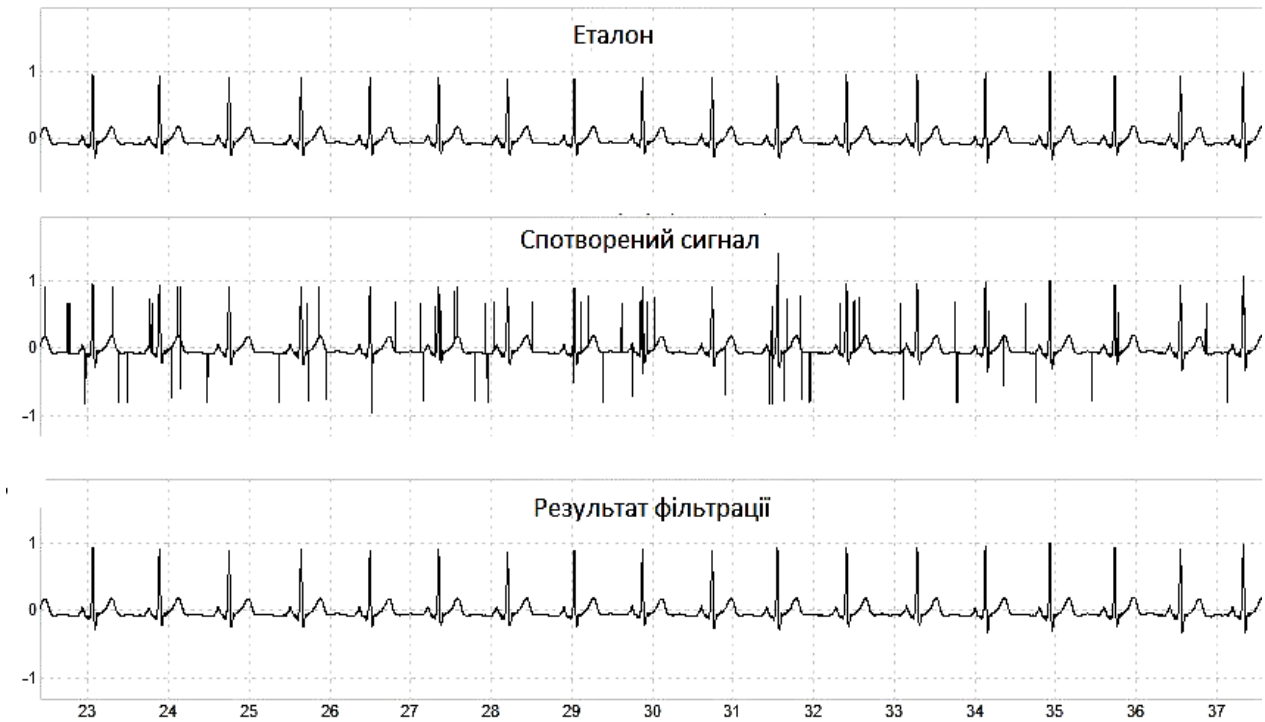


Рис. 3.20. Результат рангової фільтрації імпульсних завад на ЕКГ

З рисунка видно, що навіть за великої кількості поодиноких імпульсних завад високого рівня медіанний фільтр з вікном $W_0 = 3$ є ефективним.

Якщо ж імпульсні завади будуть з'являтися на послідовних точках сигналу, то параметр m вікна треба збільшувати. Однак експерименти показали, що за $m > 5$, коли ширина ковзного вікна сягає 11, застосування процедури стає неефективним через можливі спотворення корисного сигналу.

Так, наприклад, в умовах оброблення складного сигналу прямокутної форми (рис 3.21, а), який спотворено імпульсними завадами (рис 3.21, б), медіанним фільтром з вікном $W_0 = 5$ (рис. 3.21, в) повністю збережено особливості еталону, а з вікном $W_0 = 11$ (рис. 3.21, г) втрачено (позначено стрілкою) вузький імпульс, що належав еталону.

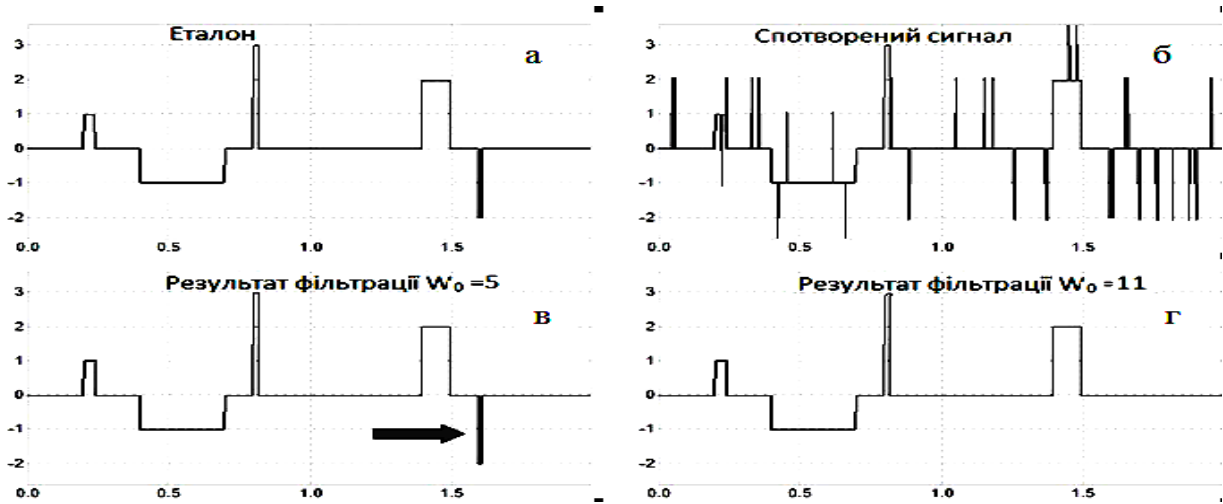


Рис. 3.21. Результат фільтрації сигналу складної форми

Розглянуті процедури належать до програмного компоненту попереднього оброблення сигналів, який реалізує послідовне оброблення завдань різного типу (рис. 3.22).

Спотворений сигнал $z[k]$

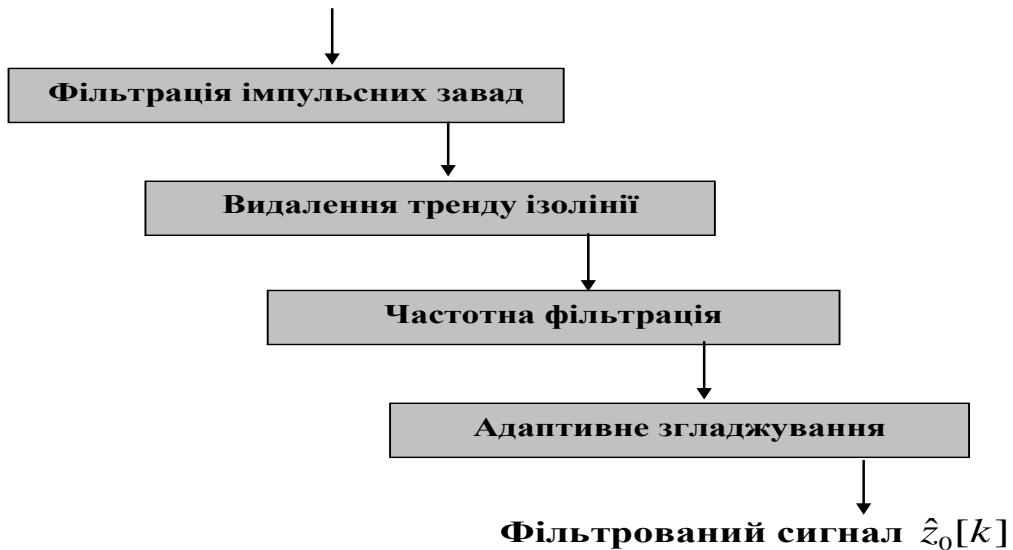


Рис. 3.22. Каскадне заглушення завдань різного типу

Завдання до комп'ютерного практикуму

Завдання 3.1. Необхідно розробити програму, яка реалізує процедури заглушення випадкових перешкод на спотворених сигналах складної форми. На вхід процедур надходить сигнал, побудований за попереднім комп'ютерним практикумом (див. завдання 2.2) у вигляді ЕКГ з альтернацією зубця T .

За вибором користувача програма має реалізувати два алгоритми згладжування з різними значеннями параметрів налаштування:

- експоненційного згладжування за формулою (3.21);
- ковзного середнього за формулою (3.26) або (3.27).

Робоче вікно програми може бути побудовано так, як показано на рис. 3.23.

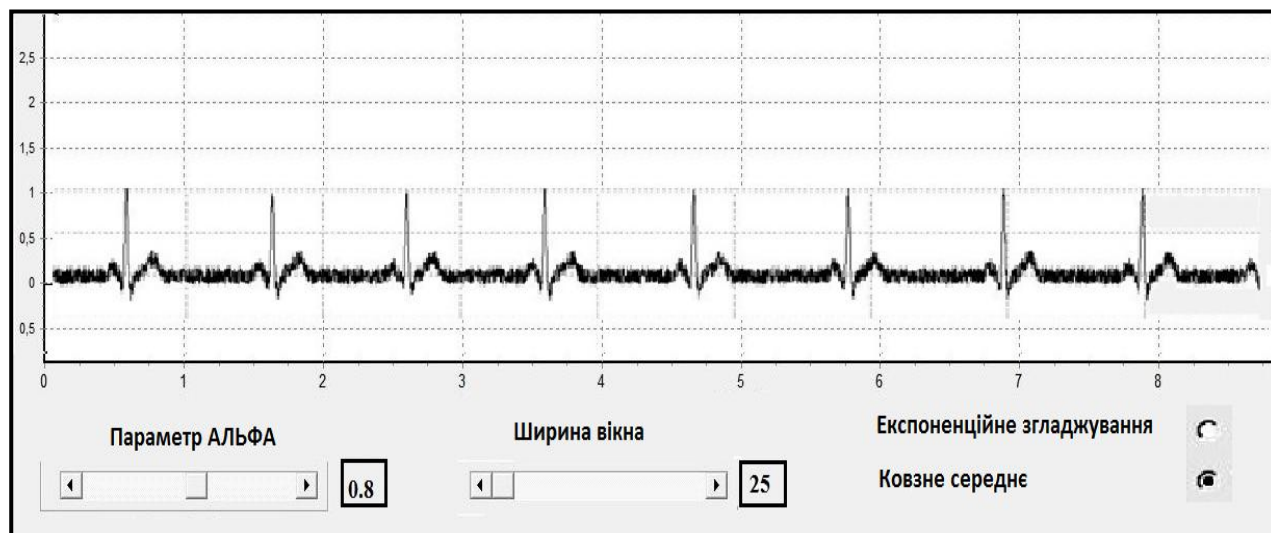


Рис. 3.23. Робоче вікно програми (завдання 3.1)

Якщо користувач обрав метод експоненційного згладжування, то повзунок, за допомогою якого налаштовується параметр алгоритму ковзного середнього (ширина вікна W), стає неактивним та навпаки, якщо користувач обрав метод ковзного середнього, то неактивним стає повзунок зміни параметра α .

Під час зміни параметрів W та α рівень спотворень сигналу має зменшуватись або збільшуватись, а у відповідних частинах вікна має відображатися поточне значення α або W .

Запитання для самоконтролю

1. Дайте загальне визначення інтелектуальної властивості «адаптація». Наведіть приклад цієї властивості в організмі людини.
2. Назвіть основні типи завад, що спотворюють біомедичні сигнали в реальних умовах.
3. Опишіть амплітудно-частотну характеристику (АЧХ) ідеального фільтра нижніх частот (ФНЧ).

4. Яку амплітудно-частотну характеристику (АЧХ) має ідеальний фільтр верхніх частот (ФВЧ)?
5. Яку амплітудно-частотну характеристику (АЧХ) має ідеальний смуговий фільтр (СФ)?
6. Яку амплітудно-частотну характеристику (АЧХ) має ідеальний режекторний фільтр (РФ)?
7. У чому полягає проблема синтезу реального режекторного фільтра?
8. Опишіть принцип частотно-вибіркової фільтрації на основі прямого та оберненого дискретних перетворень Фур'є.
9. У чому полягає ефект розходження спектра дискретного гармонічного сигналу (ефект Гіббса)?
10. За рахунок чого зменшується ефект Гіббса під час адаптивної фільтрації?
11. У чому полягає принцип заглушення випадкових завад на основі алгоритму експоненціального згладжування?
12. У чому полягає принцип заглушення випадкових завад на основі традиційного алгоритму ковзного середнього?
13. Як можна прискорити швидкодію традиційного алгоритму ковзного середнього?
14. Назвіть основні недоліки традиційних алгоритмів згладжування.
15. У чому полягає основна відмінність адаптивного алгоритму згладжування?
16. Яким чином можна видалити з сигналу імпульсні завади?
17. Що передбачає каскадна фільтрація адитивних завад?

РОЗДІЛ 4

ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ОБРОБЛЕННЯ ЦИКЛІЧНИХ СИГНАЛІВ

4.1. Оцінювання форми усередненого циклу

Фундаментальні та прикладні дослідження в техніці, біології, економіці та інших галузях часто пов'язані з аналізом спостережуваних процесів, які мають циклічний (повторюваний у часі) характер. Такі процеси породжують специфічні сигнали, які у науковій літературі називають циклічними. Типові приклади циклічних сигналів – електрокардіограми, реограми, фотоплетизмограми та інші біомедичні сигнали, що відбивають циклічний характер роботи систем кровообігу та дихання живого організму.

Проблема підвищення ефективності вилучення діагностичної інформації із циклічних сигналів, спотворених неадитивними збуреннями, досі залишається актуальною. Розглянемо деякі нові підходи до вилучення локалізованої діагностичної інформації із циклічних сигналів складної форми та на прикладі оброблення реальних сигналів продемонструємо їхню ефективність.

Одна з задач комп'ютерного оброблення циклічних сигналів – оцінювання форми усередненого циклу. Усереднення циклів дає змогу виявити приховані збуреннями зміни форми сигналу, які мають діагностичну цінність. Наприклад, метод SAECG (Signal Averaged ECG) дозволяє виявити на ЕКГ пізні потенціали (Late Potentials), які є предиктором небезпечних аритмій та навіть раптової серцевої смерті.

Тривіальні процедури усереднення циклів у часовій області, які застосовують у існуючих цифрових системах, мають низку недоліків. Щоб продемонструвати це і показати переваги запропонованого підходу, скористаємося моделлю (2.29) породження ЕКГ реалістичної форми, спростивши деякі розрахунки.

Будемо вважати, що зовнішніх завад немає, тобто $h(t) \equiv 0$, а завдяки внутрішнім збуренням на послідовності циклів змінюється лише форма одного з інформативних фрагментів – зубця T . У цьому

випадку процес породження послідовності циклів можна подати у вигляді спрощеної моделі (2.27):

$$z_m(t) = \sum_{i \in \{P, Q, R, S, S-T, T\}} \tilde{A}_{im} \exp \left[-\frac{(t - \tilde{\mu}_{im})^2}{2\tilde{b}_{im}^2} \right], \quad m = 1, 2, \dots, \quad (4.1)$$

де

$$\tilde{A}_{im} = \begin{cases} A_i(1 + \alpha_{im}), & \text{якщо } i = T, \\ A_i, & \text{якщо } i \neq T, \end{cases} \quad (4.2)$$

$$\tilde{\mu}_{im} = \begin{cases} \mu_i(1 + \delta_{im}), & \text{якщо } i = T, \\ \mu_i, & \text{якщо } i \neq T, \end{cases} \quad (4.3)$$

$$\tilde{b}_{im} = \begin{cases} b_i^{(1)}(1 + \varepsilon_{im}^{(1)}) \quad \forall t \leq \mu_{im}, & \text{якщо } i = T, \\ b_i^{(2)}(1 + \varepsilon_{im}^{(2)}) \quad \forall t > \mu_{im}, & \text{якщо } i = T, \\ b_i^{(1)} \quad \forall t \leq \mu_{im}, & \text{якщо } i \neq T, \\ b_i^{(2)} \quad \forall t > \mu_{im}, & \text{якщо } i \neq T. \end{cases} \quad (4.4)$$

У (4.2)–(4.4) – параметри, які визначають форми інформативних фрагментів еталонного циклу $z_0(t)$ корисного сигналу, який не спостерігається та має бути оцінений на основі усереднення спотворених циклів $z_m(t)$, $m = 1, 2, \dots$

На рис. 4.1 зображено результат усереднення в часовій області трьох циклів (рис. 4.1, а), породжених моделлю (4.1)–(4.4) за допомогою синхронізації цих циклів (рис. 4.1, б) по максимуму зубців R . Звідси видно, що форма зубця T на усередненому циклі (рис. 4.1, в) суттєво спотворена. За рахунок того, що моменти появи зубців T не синхронні, відбувається «розмивання» форми зубця T на усередненому циклі $z_0(t)$. Це може призвести до помилок у визначенні діагностичних ознак, що характеризують форму зубця T .

Зрозуміло, що синхронізація циклів за максимумами зубців T призведе до того, що на усередненому циклі $z_0(t)$ будуть спотворені вже інші фрагменти – форма зубців P, Q, R, S .

Здається, описану проблему можна вирішити, якщо синхронізувати інтервали, на яких зосереджені окремі фрагменти. Однак у ре-

льних ситуаціях між зубцями P, Q, R, S, T немає чітких границь і тому складно синхронізувати зазначені інтервали.

Зазначена проблема, яка спостерігається у методі усереднення циклів у часовій області, ще більше посилюється під час спотворення амплітудно-часових параметрів інших фрагментів ЕКГ.

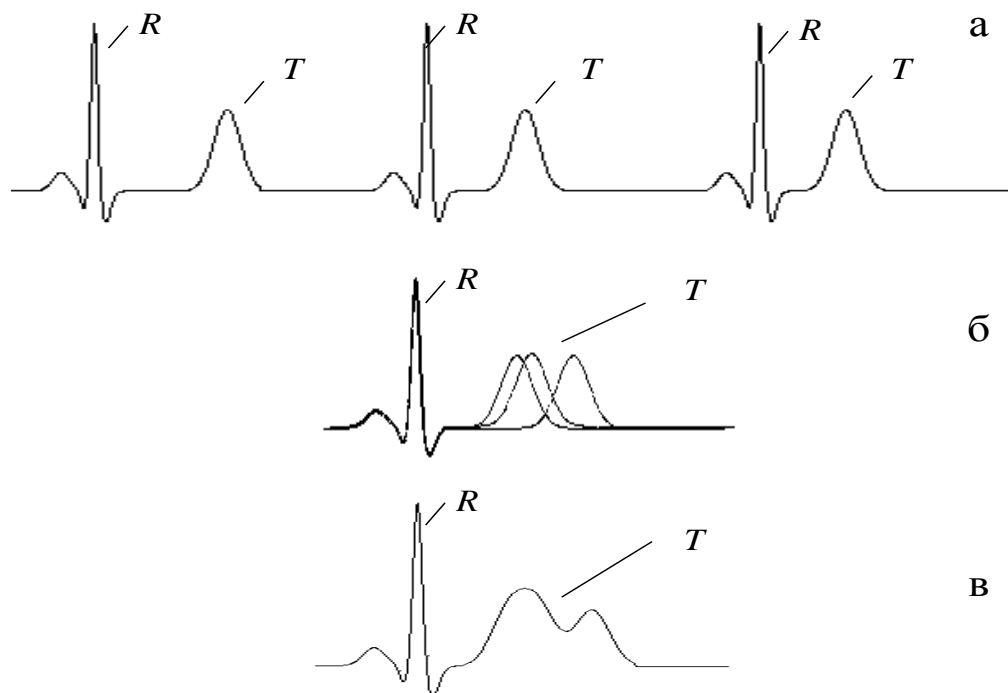


Рис. 4.1. Оцінювання усередненого циклу в часовій області

Розглянемо альтернативний підхід усереднення циклів, який зводиться до переходу від скалярного сигналу $z(t)$ до його відображення на площині з фазовими координатами $z(t), \dot{z}(t)$, де $\dot{z}(t)$ – перша похідна сигналу. Оскільки диференціювання спотворених функцій є некоректно поставленою математичною задачею, то до практичної реалізації методу потрібно залучати спеціальні обчислювальні процедури, які буде розглянуто далі.

Очевидно, що похідна $\dot{z}(t)$ будь-якої диференційованої функції $z(t)$ інваріантна щодо зсуву аргументу, тобто виконується умова

$$\dot{z}(t) \equiv \dot{z}(t - \tau), \quad \forall \tau = \text{const}. \quad (4.5)$$

Такий перехід до фазового портрета в координатах $z(t), \dot{z}(t)$ дозволяє автоматично синхронізувати зубці T на різних циклах не зважаючи на їх можливі зсуви.

Якщо для тих самих трьох спотворених циклів (рис. 4.2, а), побудувати фазовий портрет, то він має вигляд трьох траєкторій, які збігаються (рис. 4.2, б). Усереднення цих траєкторій з наступним переходом у часову область дає змогу відновити еталонний цикл (рис. 4.2, в) без будь-яких спотворень зубця T та інших інформативних фрагментів.

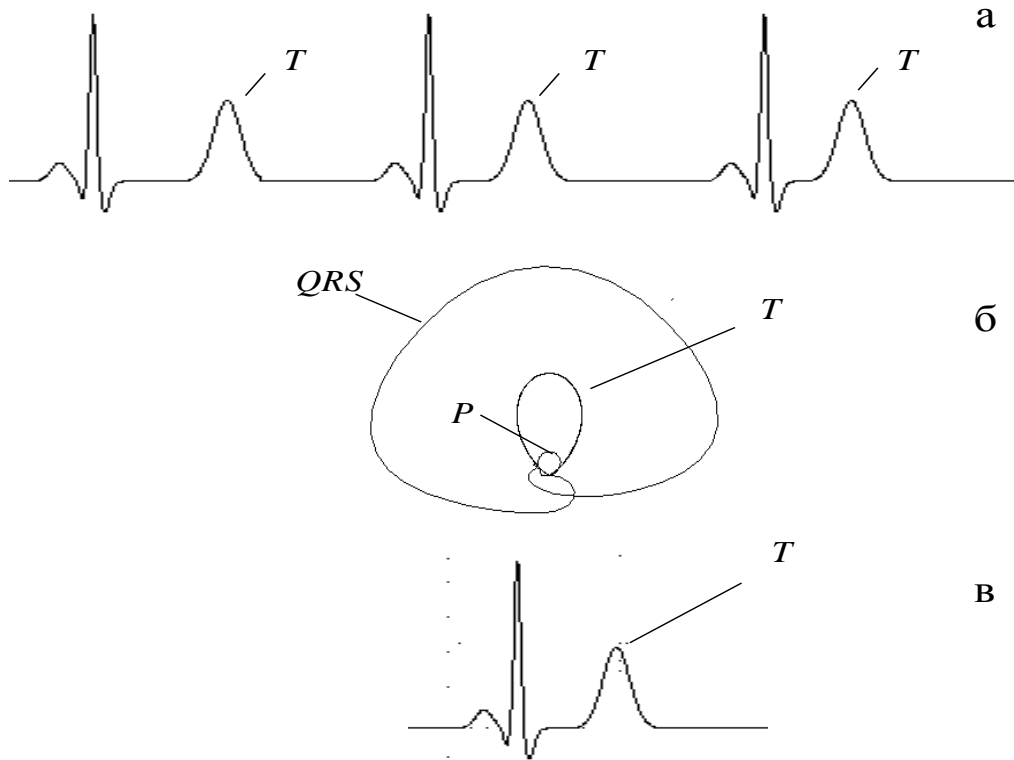


Рис. 4.2. Оцінювання усередненого циклу на фазовій площині

Нагадаємо, що строго періодична функція $f(t) \equiv f(t - T_0)$, де T_0 – період, завжди породжує траєкторії у фазовому просторі, які збігаються. Однак зворотне твердження неправильне: траєкторії, які збігаються у фазовому просторі, не обов’язково породжені строго періодичною функцією, як видно з наведеного прикладу.

У загальному випадку кожному m -му циклу, що породжується моделлю (2.29), завдяки дії випадкових збурень $\alpha_{im}, \varepsilon_{im}^{(1)}, \varepsilon_{im}^{(2)}$ відповідає своя траєкторія на фазовій площині $z(t), \dot{z}(t)$, яка відхиляється від траєкторії корисного сигналу, що не спостерігається $z_0(\cdot)$. Однак можна довести, що оскільки $\alpha_{im}, \varepsilon_{im}^{(1)}, \varepsilon_{im}^{(2)}$ – незалежні випадкові величини та мають нульові математичні сподівання, то для будь-якої фіксованої точки $\theta = \theta_x$ області визначення корисного сигналу $z_0(\cdot)$ координати $z_m[\theta_{mx}], \dot{z}_m[\theta_{mx}]$ відповідних точок $\theta_{1x}, \dots, \theta_{Mx}$ сигналу $z(\cdot)$,

що спостерігається, групуватимуться в локальній області фазового простору з центром у точці $Z = z_0^*[\theta_x], \dot{z}_0^*[\theta_x]$ та обмеженим радіусом ρ_z .

Фазовий портрет можна будувати і в координатах $z(t), z(t - \tau)$, де τ – затримка в часі. Таке відображення називають псевдо фазовим портретом (рис. 4.3). Однак, як буде показано далі, фазовий портрет у координатах $z(t), \dot{z}(t)$ має низку переваг, зокрема, надає практичну можливість оцінювати додаткові ознаки ЕКГ, які мають діагностичну цінність.

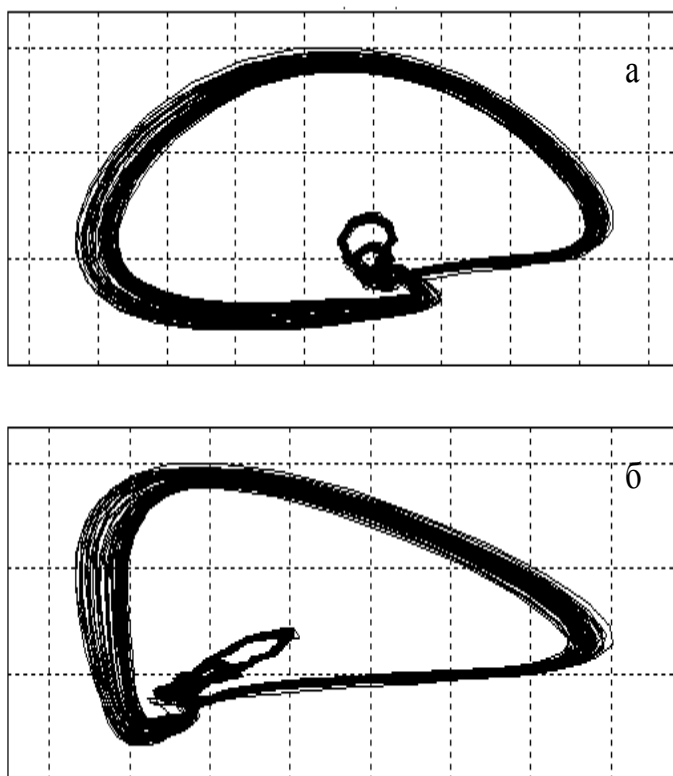


Рис. 4.3. Фазовий (а) та псевдофазовий (б) портрети одного сигналу

Відображення сигналу в координатах $z(t), \dot{z}(t)$ можна пояснити як графоаналітичний метод дослідження динамічної системи, стан якої можна описати двома диференціальними рівняннями

$$\dot{x}_1 = x_2, \quad (4.6)$$

$$\dot{x}_2 = F(x_1, x_2), \quad (4.7)$$

де $x_1 = z(t)$ – вихідна координата системи, наприклад, значення сигналу ЕКГ, що вимірюють у момент часу t , $x_2 = \dot{z}(t)$ – перша похідна сигналу, а $F(x_1, x_2)$ – деяка нелінійна функція.

Розділивши (4.7) на (4.6), отримаємо рівняння, в якому немає часу в явному вигляді:

$$\frac{dx_2}{dx_1} = \frac{F(x_1, x_2)}{x_2}. \quad (4.8)$$

Розв'язок рівняння (4.8)

$$x_2 = \Psi(x_1) \quad (4.9)$$

з урахуванням позначень $x_1 = z(t)$ і $x_2 = \dot{z}(t)$ можна подати у вигляді залежності

$$\dot{z}(t) = \Psi(z(t)), \quad (4.10)$$

яка визначає фазову траєкторію на площині $z(t), \dot{z}(t)$.

Оскільки невідомі аналітичні вирази функцій $x_1 = z(t)$ і $F(\cdot)$, будуватимемо фазовий портрет ЕКГ експериментально, використовуючи цифрову послідовність $z(t_k)$ дискретних значень сигналу, що спостерігається в моменти часу $t_k \equiv k\Delta$, $k = 0, 1, \dots, K-1$, де Δ – крок квантування.

Для переходу від скалярного сигналу $z(t)$ до векторів у просторі $z(t), \dot{z}(t)$ необхідно оцінювати першу похідну сигналу. Відомо, що обчислення похідної функції, що спотворено завадою, є некоректною математичною задачею. Покажемо це на такому наочному прикладі.

Нехай сигнал, що спостерігається,

$$z(t) = z_0(t) + h(t), \quad (4.11)$$

є сумою корисного сигналу $z_0(t)$ та завади

$$h(t) = h_0 \sin(\omega t), \quad (4.12)$$

де $h_0 = \text{const}$, $\omega = \text{const}$.

З виразу (4.11), з урахуванням (4.12), випливає, що

$$|z(t) - z_0(t)| \leq h_0. \quad (4.13)$$

З другого боку, відповідно до (4.11) маємо

$$\dot{z}(t) = \dot{z}_0(t) + \omega h_0 \cos(\omega t), \quad (4.14)$$

тобто

$$|\dot{z}(t) - \dot{z}_0(t)| \leq \omega h_0. \quad (4.15)$$

Зважаючи на порівняння (4.13) та (4.15), можна зробити висновок, що за будь-яким невеликим відхиленням Δ_z функцій $z(t)$ і $z_0(t)$, що відбувається завдяки дії завади, можливе будь-яке велике відхилення Δ_z похідних $\dot{z}(t)$ і $\dot{z}_0(t)$, що залежить від частоти ω . Саме це свідчить про некоректність математичної постановки задачі оцінювання похідною за спотвореною реалізацією.

Разом з тим за допомогою відомих чисельних методів можна обчислювати наближені значення похідної дискретно заданої функції $z[k]$, $k = 0, \dots, K - 1$.

Один з відомих методів чисельного оцінювання значення похідної $\dot{z}[k]$ у k -й точці дискретного сигналу $z[k]$, $k = 0, \dots, K - 1$ ґрунтується на диференціюванні інтерполяційного полінома Лагранжа, побудованого за n вузлами в околі k -ї точки. В такому випадку помилка апроксимації стає найменшою, якщо точка k є центральною серед вузлів інтерполяції.

Використовуючи загальну формулу, можна показати, що для кількості вузлів $n = 7$ та одиночного кроку квантування оцінювання похідної в k -й точці можна визначити за формулою

$$\dot{z}[k] = 1/60(z[k+3] - 9z[k+2] + 45z[k+1] - 45z[k] + 9z[k-1] - z[k-2]). \quad (4.16)$$

Зауважимо, що для побудови фазових портретів важливі не кількісні характеристики оцінювання (4.16), а лише збереження «форми» функції, що апроксимує похідну $\dot{z}[k]$ у тому розумінні, щоб на апроксимуючій функції адекватно відображалися особливі точки $z[k]$, $k = 0, \dots, K - 1$, які мають діагностичну цінність, зокрема, екстремальні значення зубців ЕКГ.

Отже, можна передбачити оброблення сигналу у два етапи: оцінювання похідної попередньо фільтрованого сигналу за формулою (4.16) і адаптивне згладжування отриманих даних на основі алгоритму (3.29), який забезпечує регуляцію процедури чисельного оцінювання похідної (рис. 4.4).

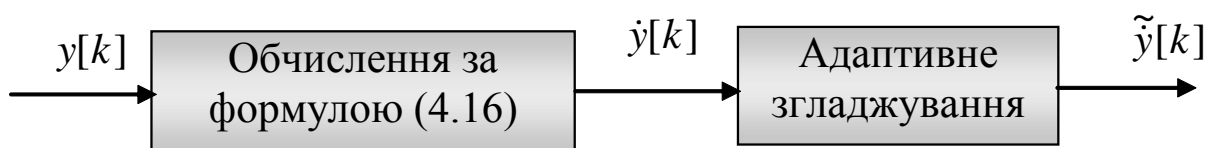


Рис. 4.4. Підхід до чисельного диференціювання сигналу

Для прикладу на рис. 4.5 наведено результат чисельного оцінювання першої похідної модельного сигналу – гармонічного коливання, який спотворено 10-відсотковою адитивною перешкодою.

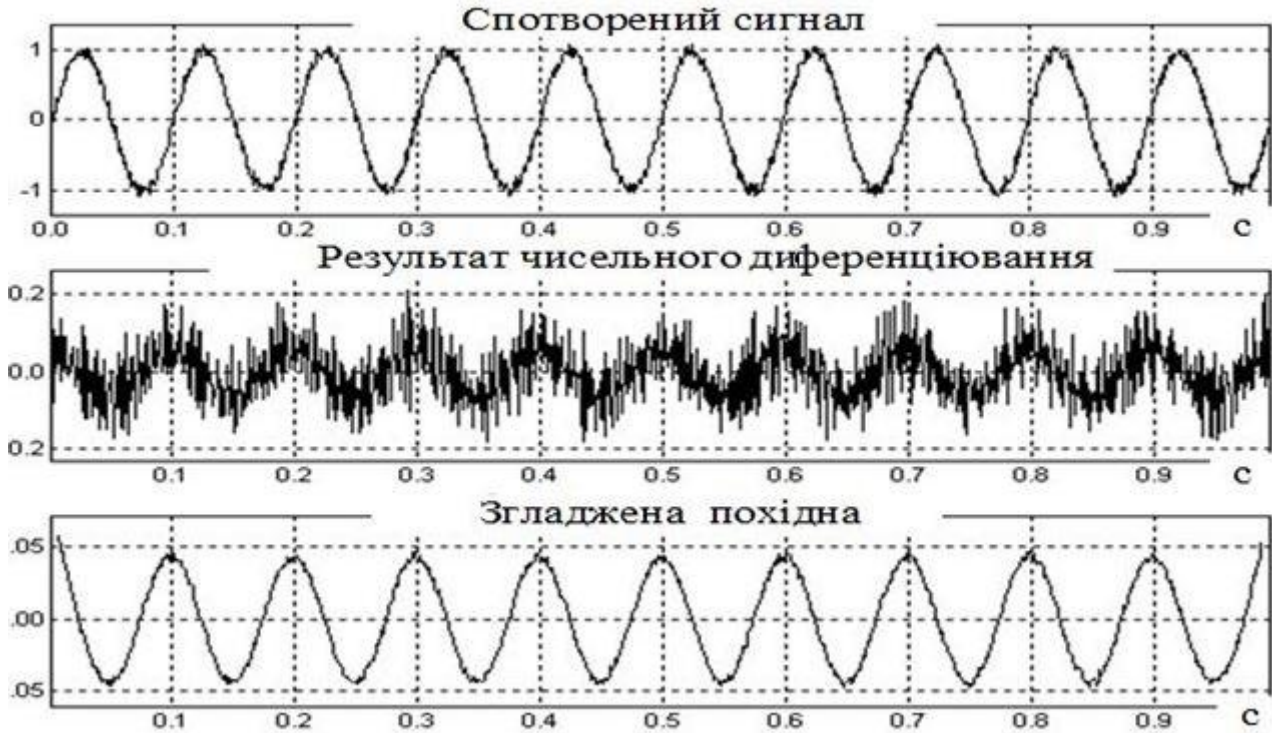


Рис. 4.5. Результат оцінювання похідної модельного сигналу

Розглянемо безпосередньо усереднення циклів на фазовій площині. Для цього послідовно будемо розв'язувати такі задачі:

Задача 4.1. Поділ сигналу, що спостерігається, на окремі цикли.

Задача 4.2. Виділення найбільш представницького (домінантного) циклу.

Задача 4.3. Селекція циклів, що підлягають усередненню.

Задача 4.4. Оцінювання еталонного циклу через усереднення фазових траєкторій.

Задача 4.5. Повернення до еталонного циклу у часовій області.

Оскільки значення сигналу та похідної суттєво відрізняються за рівнем, виконаємо нормування

$$z_m^*[k] = \frac{z_m[k] - \min_{0 \leq z \leq K-1} z_m[k]}{\max_{0 \leq k \leq K-1} z_m[k] - \min_{0 \leq k \leq K-1} z_m[k]}, \quad k = 0, \dots, K-1,$$

$$\dot{z}_m^*[k] = \frac{\dot{z}_m[k] - \min_{0 \leq k \leq K-1} \dot{z}_m[k]}{\max_{0 \leq k \leq K-1} \dot{z}_m[k] - \min_{0 \leq k \leq K-1} \dot{z}_m[k]}, \quad k = 0, \dots, K-1,$$

та перейдемо до послідовності нормованих векторів $z[k] \in [0,1]$ і $\dot{z}[k] \in [0,1]$:

$$(z^*[0], \dot{z}^*[0]), (z^*[1], \dot{z}^*[1]), \dots, (z^*[K-1], \dot{z}^*[K-1]). \quad (4.17)$$

Відомі алгоритми поділу сигналу на окремі цикли в часовій області не завжди є ефективними, оскільки іноді амплітуди зубців T можуть перевершувати амплітуди зубців R , що ускладнює розв'язок задачі пошук QRS -комплексів (рис. 4.6).

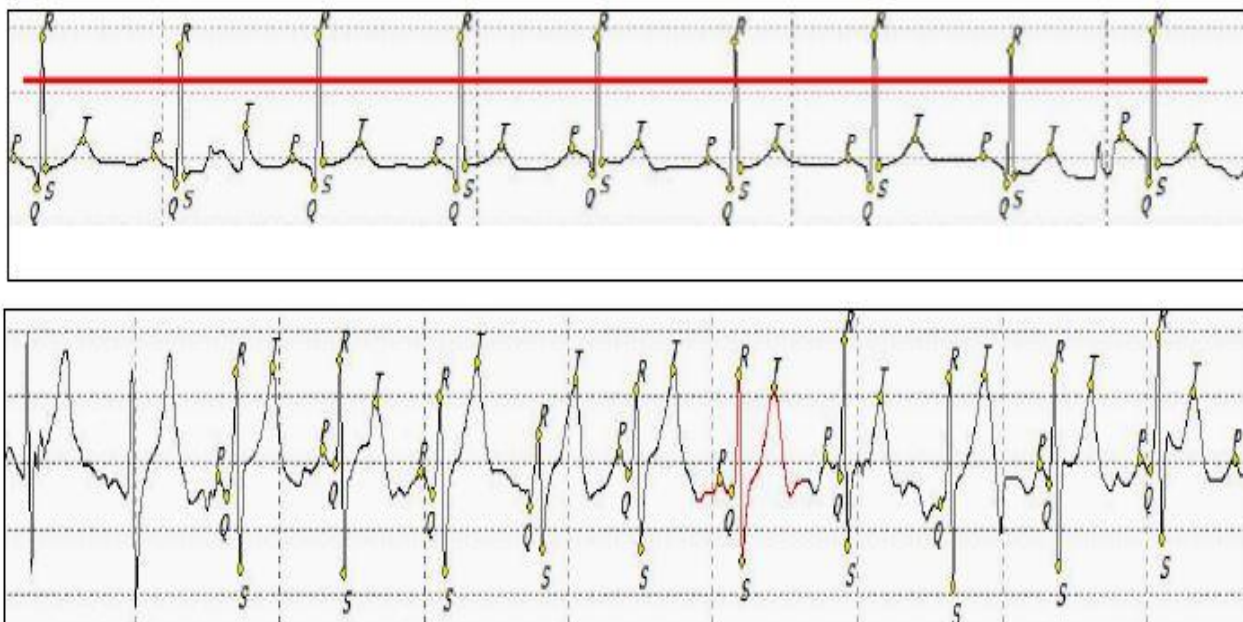


Рис. 4.6. Складність поділу сигналу на цикли в часовій області

Покажемо, що можна побудувати алгоритм розподілу реальних ЕКГ на серцеві цикли, використовуючи фазовий портрет ЕКГ.

Обчислимо квадрати відстаней між нормованими фазовими координатами кожної k -ї точки і деякою фіксованою (базовою) точкою $k_0 \in [0, K-1]$:

$$r_0^2[k] = (z^*[k] - z^*[k_0])^2 + (\dot{z}^*[k] - \dot{z}^*[k_0])^2, \quad k = 0, \dots, K-1. \quad (4.18)$$

Оскільки зазвичай зубець R відрізняють високі значення амплітуди і похідної, то у разі вдалого вибору точки k_0 в моменти появи

зубців R на профілі відстаней $r_0^2[k]$ з'являтимуться виражені максимуми навіть тоді, коли ЕКГ має складну форму, зокрема, високий зубець T (рис. 4.7).

Якщо виділити підмножини Ω_m точок $k \in [0, K - 1]$, таких, що

$$r_0^2[k] \geq r_{QRS} \quad \forall k \in \Omega_m, \quad (4.19)$$

де r_{QRS} – поріг, і для кожної підмножини Ω_m визначити номери точок k_m , $m = 1, \dots, M$, які задовольняють умову

$$k_m = \operatorname{argmax}_{k \in \Omega_m} r_0^2[k], \quad (4.20)$$

то отримана послідовність номерів k_m , $m = 1, \dots, M$ визначить межі окремих циклів сигналу.

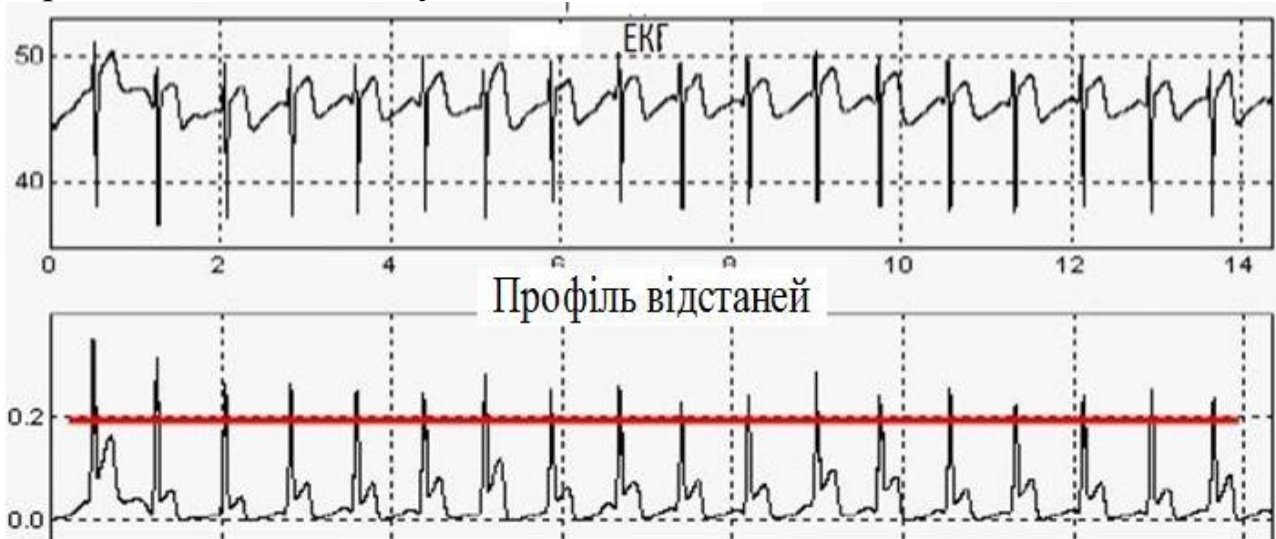


Рис. 4.7. Поділ сигналу на окремі цикли за профілем відстаней

У результаті фазовий портрет буде розділений на M траєкторій

$$Q_m = \{(z_m^*[k_m], \dot{z}_m^*[k_m]), k_m = 0, \dots, K^{(m)} - 1\}, \quad m = 1, \dots, M. \quad (4.21)$$

Оптимальне значення базової точки k_0 , вибір якої впливає на форму профілю $r_0^2[k]$, визначається за умови

$$k_0 = \operatorname{argmax}_{0 \leq k_j \leq K-1} \left\{ \frac{\max_{0 \leq k \leq K-1} r[k, k_j]}{\frac{1}{K} \sum_{k=0}^{K-1} r[k, k_j]} \right\}. \quad (4.22)$$

Для визначення оптимального значення порога r_{QRS} , що використано у виразі (4.19), необхідно забезпечити певну рівновагу між можливими помилками пропуску цілі та помилковою тривоною. Експерименти показали, на сучасних комп'ютерах оптимальний поріг фактично миттєво можна визначити методом послідовного перебору впорядкованих значень.

Процедуру виділення QRS -комплексів можна зробити ще більш ефективною, якщо попередньо перейти від початкової послідовності спостережень $z[k]$, $k = 0, \dots, K - 1$ до послідовності величин

$$\tilde{z}[k] = z[k] - \frac{1}{2W} \sum_{i=-W}^W z[k-i], \quad k = W, \dots, K-1-W, \quad (4.23)$$

де W – вікно ковзного згладжування.

Оскільки тривалість QRS -комплексу істотно менша за тривалість зубців P і T , такий спосіб дає змогу отримати більш виражені пікові значення на профілі квадратів відстаней (рис. 4.8)

$$\tilde{r}_0^2[k] = (\tilde{z}^*[k] - \tilde{z}^*[k_0])^2 + \tilde{z}^*[k] - \tilde{z}^*[k_0])^2 \quad (4.24)$$

порівняно з профілем, що обчислюється за (4.18).

Наступний етап обробки сигналу передбачає селекцію циклів, які підлягають усередненню. Тому необхідно класифікувати типові і нетипові цикли, тобто реалізувати інтелектуальну властивість «узагальнення».

Один з відомих підходів до розв'язування задачі класифікації ґрунтується на пошуку узагальнюючих ознак, що характеризують об'єкти одного класу. Часто узагальнюючі ознаки формують методами машинного навчання (**Machine Learning**) за прецедентами – представникам окремих класів, зібраних у навчальну вибірку спостережень.

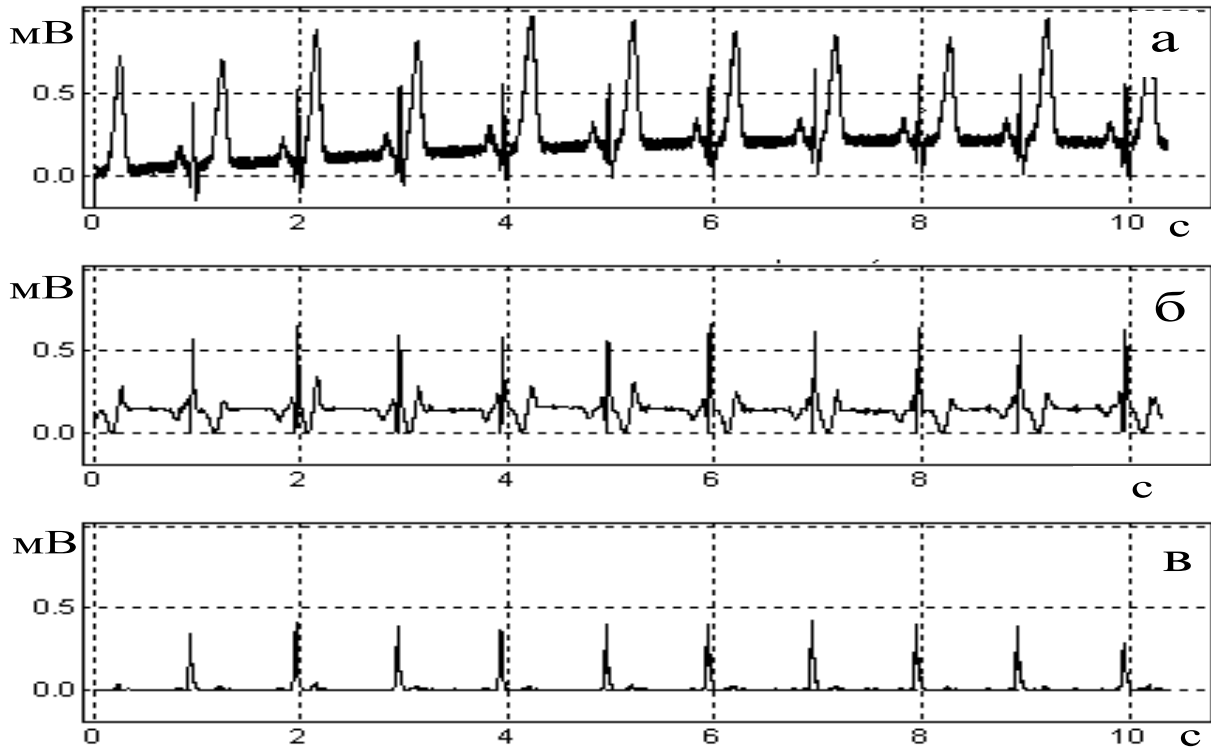


Рис. 4.8. Розділені ЕКГ (а) на окремі цикли за профілями $r_0^2[k]$ (б) та $\tilde{r}_0^2[k]$ (в)

Однак для класифікації циклів ЕКГ такий підхід виявився незастосовним. Складність задачі класифікації циклів обумовлена тим, що форма нетипового циклу ЕКГ одного пацієнта може збігатися з типовими циклами іншого та навпаки (рис. 4.9).



Рис. 4.9. ЕКГ з екстрасистою (вгорі), форма якої збігається з типовими циклами іншої ЕКГ (унизу)

Звідси випливає, що неможливо побудувати процес класифікації, який ґрунтується на аналітичному описі форми циклів, отже не варто шукати узагальнюючі ознаки методами машинного навчання.

Розглянемо інший підхід до розв'язування задачі, який передбачає єдине обмеження: кількість M_N нетипових циклів (екстрасистол і артефактів) значно менше від загальної кількості M циклів, тобто

$$M_N \ll M, \quad (4.25)$$

оскільки інакше означення «нетиповий» цикл втрачає сенс.

Підхід ґрунтується на визначенні близькості траєкторій на основі гаусдорфової метрики. Що визначити цю величину, спочатку нагадаємо основні положення поняття метричного простору.

У математиці часто розглядаються множини, між елементами яких визначено відстань (метрику). У цьому випадку вважають, що елементи задані в метричному просторі.

Відстань $\rho(A, B)$ дає змогу оцінювати близькість (схожість) об'єктів A і B та задовольняє таким умовам:

- $\rho(A, B) \geq 0$, зокрема $\rho(A, B) = 0$ в тому і тільки тому випадку, коли $A = B$;
- $\rho(A, B) = \rho(B, A)$ – властивість симетрії;
- $\rho(A, B) \leq \rho(A, C) + \rho(C, B)$ – властивість трикутника.

Наведемо лише декілька найбільш поширених відстаней.

Евклідова відстань оцінює близькість двох точок в евклідовому просторі. Зокрема, якщо цей простір двовимірний (площина (x, y)), то евклідова відстань між двома точками $A(x_1, y_1)$ і $B(x_2, y_2)$ обчислюється за формулою

$$\rho_1(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (4.26)$$

Манхеттенська відстань визначає близькість двох точок у міських кварталах (рис. 4.10) і обчислюється за формулою

$$\rho_2(A, B) = |x_1 - x_2| + |y_1 - y_2|. \quad (4.27)$$

Відстань Махаланобіса визначає близькість двох випадкових векторів \vec{A} і \vec{B} з одного розподілу ймовірностей з коваріаційною матрицею S та обчислюються за формулою

$$\rho_3(\vec{A}, \vec{B}) = \sqrt{(\vec{A} - \vec{B})^T S^{-1} (\vec{A} - \vec{B})}. \quad (4.28)$$

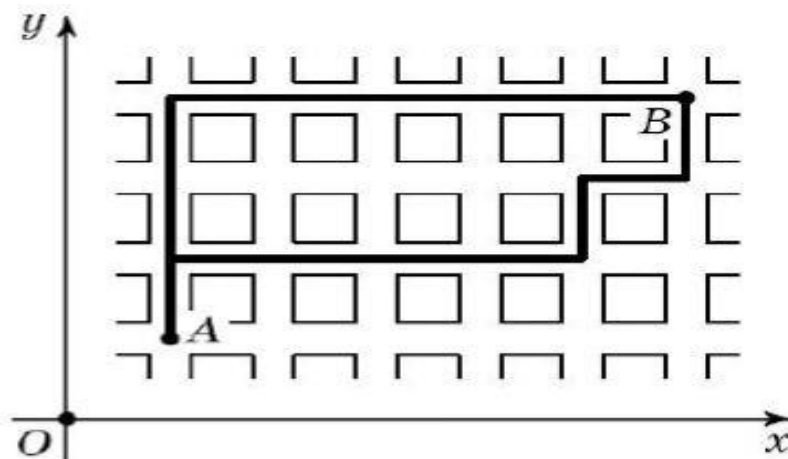


Рис. 4.10. Манхеттенська відстань

Відстань Хемінга визначає близькість двох N розрядних кодів $A = a_1 \dots a_N$ і $B = b_1 \dots b_N$, де $a_i \in \{0,1\}$, $b_i \in \{0,1\}$, $i = 1, \dots, N$ та обчислюється за формулою

$$\rho_4(A, B) = \sum_{i=1}^N a_i \oplus b_i, \quad (4.29)$$

де $a_i \in \{0,1\}$, $b_i \in \{0,1\}$, а \oplus – операція XOR (виключне АБО).

Відстань Левенштейна визначає відстань між словами (послідовностями символів) A і B можливо різної довжини, відповідає мінімальній кількості елементарних операцій редагування («ВСТАВКА», «ВИДАЛЕННЯ», «ЗАМІНА» одного символу), що забезпечує перехід від слова A до слова B . Розраховується за методом динамічного програмування, що буде показано далі.

Відстань Гаусдорфа використовують для визначення відстані між множинами точок $A = \{a_1, \dots, a_N\}$ і $B = \{b_1, \dots, b_M\}$ у припущенні, що в загальному випадку може виконуватися умова $N \neq M$, та визначається за формулою

$$\rho_4(A, B) = \max\left\{ \max_{b_j \in B} \min_{a_i \in A} \rho_1(a_i, b_j), \max_{a_i \in A} \min_{b_j \in B} \rho_1(a_i, b_j) \right\}, \quad (4.30)$$

де $\rho_1(a_i, b_j)$ – евклідова відстань між елементами $a_i \in A$ і $b_j \in B$.

Будемо використовувати відстань Гаусдорфа для оцінювання близькості фазових траєкторій (рис. 4.11).

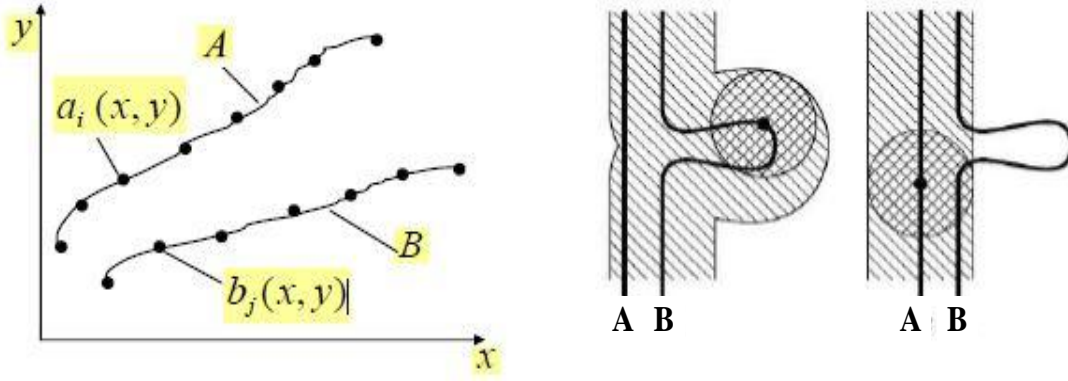


Рис. 4.11. Відстань Гаусдорфа між кривими на площині A і B

Обчислимо $M \times M$ матрицу $D = \|R_H(Q_i, Q_j)\|$ відстаней Гаусдорфа між всіма парами $Q_i \in Q$ і $Q_j \in Q$, $i, j = 1, \dots, M$ множини $Q = \{Q_1, \dots, Q_M\}$ фазових траєкторій, що відповідають окремим циклам:

$$R_H(Q_i, Q_j) = \max \left\{ \max_{q_j \in Q_j} \min_{q_i \in Q_i} \rho(q_i, q_j), \max_{q_i \in Q_i} \min_{q_j \in Q_j} \rho(q_i, q_j) \right\}, \quad (4.31)$$

де $q \triangleq (z^*, \dot{z}^*)$, а $\rho(q_i, q_j) = \|q_i - q_j\|$ – евклідова відстань між нормованими векторами $q_i \in Q_i$ і $q_j \in Q_j$.

Номер рядка матриці $D = \|R_H(Q_i, Q_j)\|$, сума елементів якого мінімальна, визначає підмножину $Q_0 = \{q[k]\}$ нормованих векторів $q[k] = (z^*[k], \dot{z}^*[k])$, що належать фазовій траєкторії одного з циклів. Будемо називати таку підмножину

$$Q_0 = \arg \min_{1 \leq j \leq M} \sum_{i=1}^M R_H(Q_i, Q_j) \quad (4.32)$$

опорною траєкторією в нормованому фазовому просторі.

Якщо виконується умова (4.25), то опорний цикл, знайдений згідно з (4.32), можна вважати «типовим», що дозволяє знайти «нетипові» цикли за відстанню до Q_0 .

З цією метою впорядкуємо за зростанням елементи рядка матриці $D = \|R_H(Q_i, Q_j)\|$ гаусдорфових відстаней (4.31), що відповідає опорному циклу.

Отримана послідовність відстаней

$$\mathfrak{R} = R_H(Q_0, Q_\mu), \quad \mu \in [1, M]$$

дає уявлення про варіабельність форми траєкторій спостережуваних сигналів відносно Q_0 .

Якщо сигнал, що спостерігається, містить лише типові цикли, тобто породжений одним еталоном $z_0(\cdot)$, то впорядковані значення $R_H(Q_0, Q_\mu)$ зростають певною мірою поступово зі зростанням μ (рис. 4.12).

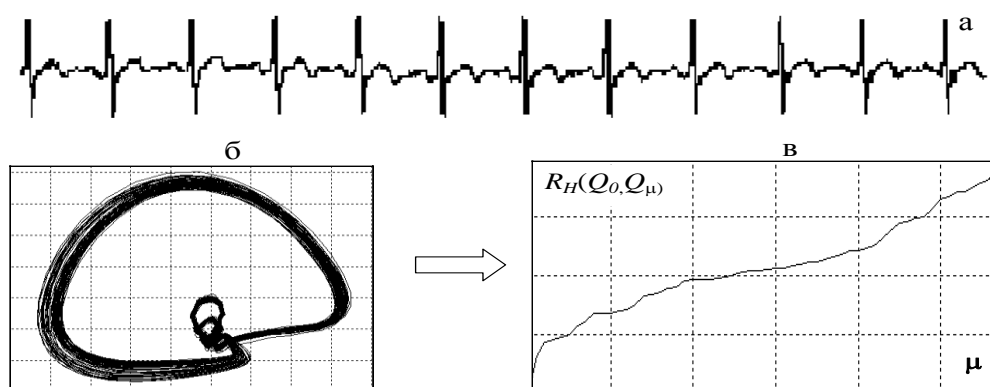


Рис. 4.12. Аналіз варіабельності форми ЕКГ (а) з типовими циклами по фазовому портрету (б) та кривою відстаней \mathfrak{R} (в)

Якщо ж на ЕКГ з'являються нетипові цикли, форма яких суттєво відрізняється від еталона (артефакти або екстрасистоли), інакше кажучи, якщо сигнал породжений кількома еталонами, то послідовність впорядкованих відстаней Гаусдорфа \mathfrak{R} містить один або кілька виражених стрибків (рис. 4.13).

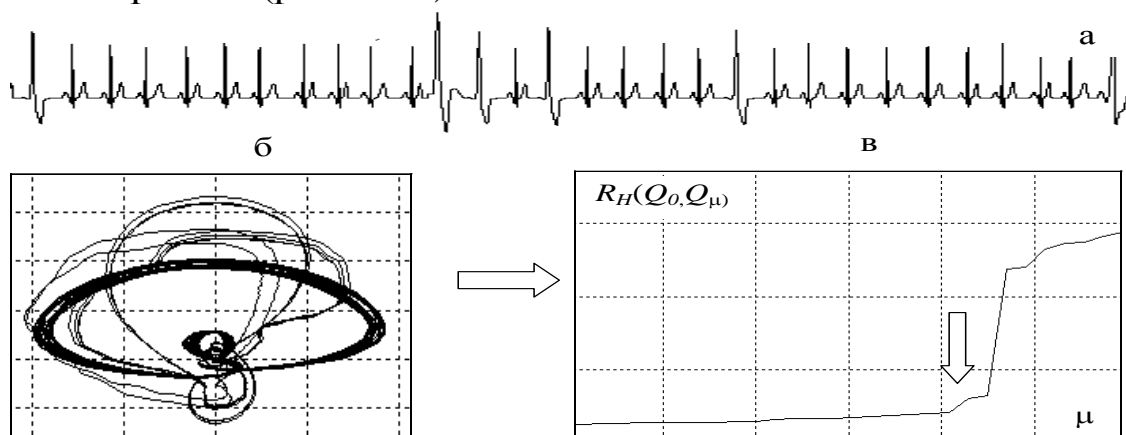


Рис. 4.13. Аналіз варіабельності форми ЕКГ (а) з нетиповими циклами по фазовому портрету (б) та кривій відстаней (в)

Визначивши положення першого з таких стрибків, можна знайти граничне значення \mathfrak{R}_0 , що дозволить виділити підмножину типових траєкторій

$$\Omega_N = \{Q_\mu : R_H(Q_0, Q_\mu) < \mathfrak{R}_0\},$$

які підлягають усередненню.

Експерименти, проведені на реальних і модельних ЕКГ (понад 1000 спостережень), показали ефективність запропонованого підходу: майже на всіх експериментах процедура безпомилково виявляла нетипові цикли (рис. 4.14).

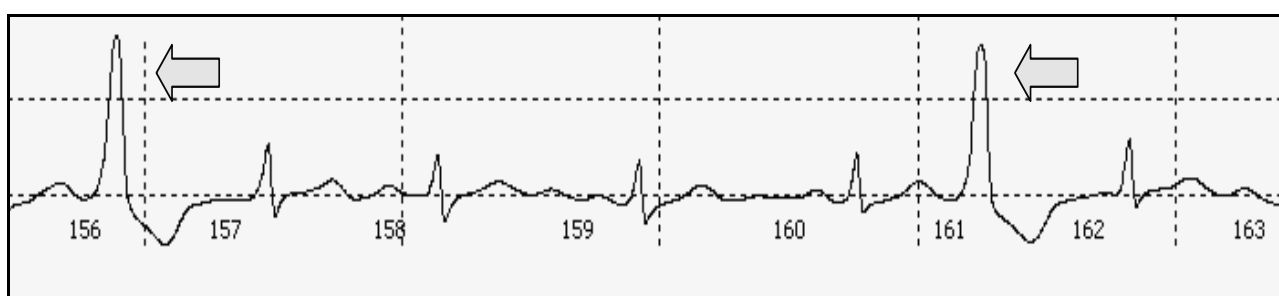


Рис. 4.14. Приклад виявлення нетипових циклів реальної ЕКГ (позначені стрілками)

Оскільки відповідно до (4.32) середня відстань від опорної траєкторії до інших траєкторій мінімальна, то траєкторія Q_0 розміщена «всередині» множини $Q = \{Q_1, \dots, Q_M\}$ (рис. 4.15).

Згідно з моделлю (2.29), за рахунок дії випадкових α_{im} , $\varepsilon_{im}^{(1)}$, $\varepsilon_{im}^{(2)}$, які мають нульове математичне сподівання та обмежену дисперсію, в околі кожної точки $q_0[k] \in Q_0$, що належить опорній траєкторії, розміщені відповідні точки інших траєкторій (рис. 4.15).

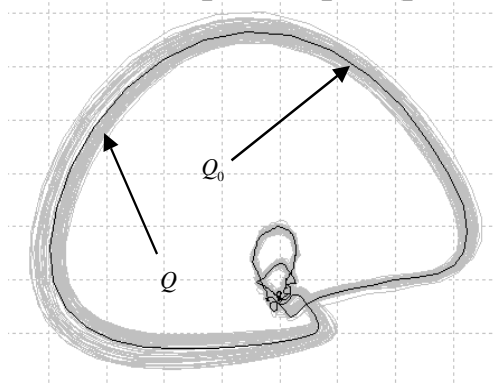


Рис. 4.15. Опорна траєкторія Q_0 у фазовому просторі

Будемо використовувати траєкторію Q_0 як перше наближення траєкторії еталонного циклу $z_0(\cdot)$, що не спостерігається, а для отримання більш точного оцінювання усереднимо точки траєкторій в околі відповідних точок опорної траєкторії (рис. 4.16).

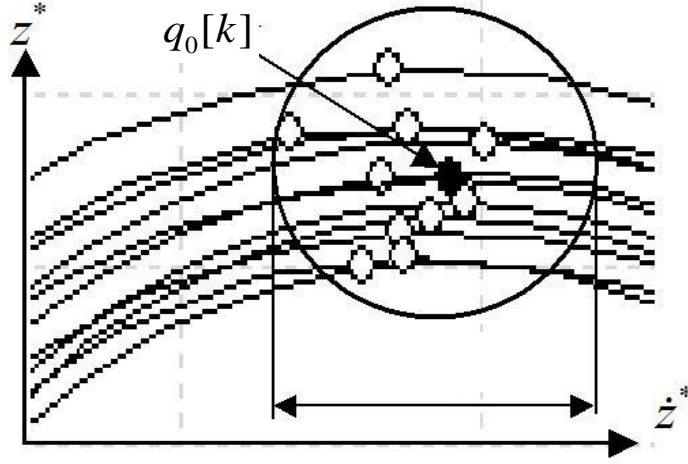


Рис. 4.16. Розміщення точок фазових траєкторій в околі точки $q_0[k] \in Q_0$ опорного циклу

Обчислимо послідовність векторів $\hat{q}[k]$, $k = 1, \dots, K_0$ за формулою

$$\hat{q}[k] = \frac{q_0[k] + \sum_{m=1}^{M-1} q_{mk}^{\min}}{M}, \quad (4.33)$$

де q_{mk}^{\min} – вектор, що належить m -й траєкторії (яка не є опорною), та розміщений на мінімальній евклідовій відстані

$$\rho = \sqrt{(z_0^*[\cdot] - z_m^*[\cdot])^2 + (\dot{z}_0^*[\cdot] - \dot{z}_m^*[\cdot])^2} \quad (4.34)$$

від вектора $q_0[k] \in Q_0$, а K_0 – кількість точок опорної траєкторії.

Послідовність векторів

$$\hat{q}[k] = (\hat{z}^*[k], \hat{\dot{z}}^*[k]), \quad k = 1, \dots, K_0,$$

обчислених згідно з (4.33), належить середній траєкторії в нормованому фазовому просторі, а відповідна послідовність значень $\hat{z}[1], \dots, \hat{z}[K_0]$ дає оцінювання еталону $z_0(\cdot)$ у часовій області (рис. 4.17).

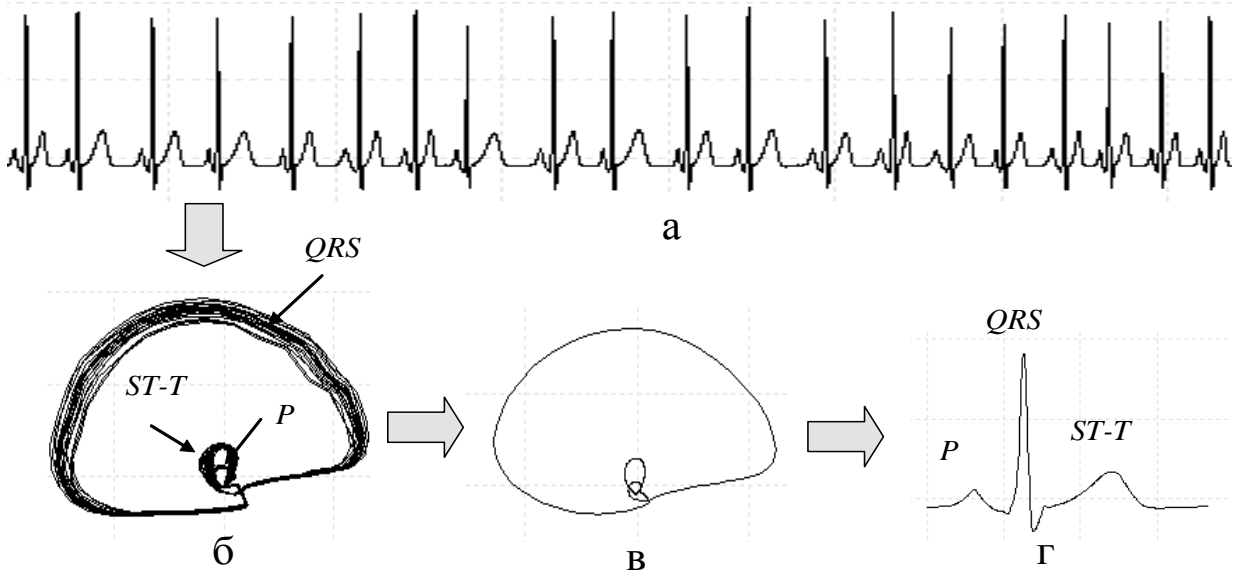


Рис. 4.17. Послідовність етапів оброблення ЕКГ: ЕКГ(а); її фазова траєкторія (б); усереднена фазова траєкторія (в); еталонний цикл ЕКГ у часовій області (г)

Для побудови опорного циклу (4.32) було використано класичну метрику Гаусдорфа (4.31), яка, загалом, призначена для визначення відстані між двома множинами точок. Водночас нас цікавлять відстані між двома послідовностями точок.

До того ж у загальному випадку фазові координати будь-якої з траєкторій, що спостерігаються, можуть перетинатися в різні моменти часу, наприклад, якщо зубець T має велику амплітуду (рис. 4.18). У такому разі найближчі за відстанню (4.34) точки можуть належати різним фрагментам серцевого циклу.

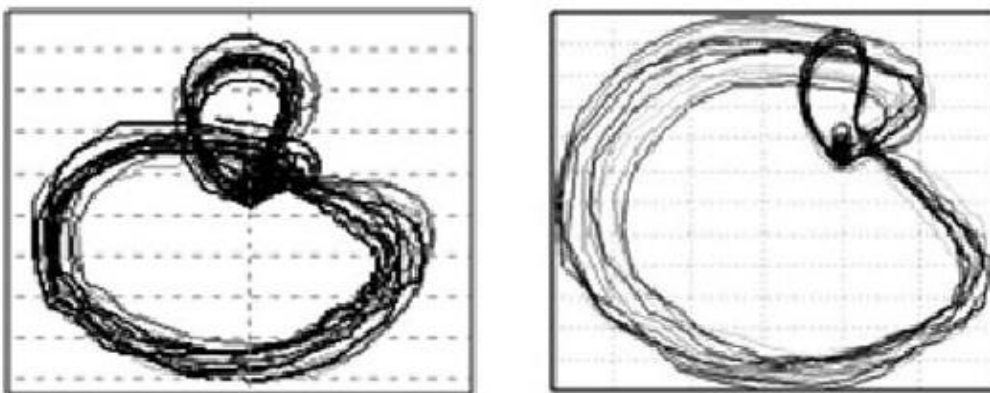


Рис. 4.18. Приклади самоперетину фазових траєкторій

Для поліпшення оцінювання еталона треба вдосконалити запропонований алгоритм. Для цього подамо опорну траєкторію Q_0 і кожну з інших $M-1$ траєкторій послідовністю розширених векторів $u[k] = (q[k], \tau[k])$, які, крім нормованих фазових координат $q[k] = (z^*[k], \dot{z}^*[k])$, містять додаткову компоненту $\tau[k]$, що має сенс відносного часу появи k -ї точки на m -й траєкторії.

Величину $\tau[k]$ визначимо за формулою

$$\tau[k] = (k-1)/(K_m-1), \quad (4.35)$$

де K_m – кількість точок m -ї траєкторії. Зауважимо, що величина $\tau[k]$ нормована у тих самих межах, як і компоненти $z^*[k], \dot{z}^*[k]$ векторів $q[k]$, тобто $\tau[k] \in [0, 1]$.

Введення додаткової компоненти $\tau[k] \in [0, 1]$ дає змогу оцінювати близькість траєкторій, що усереднюються, не тільки з погляду значень фазових координат $z^*[k], \dot{z}^*[k]$, які можуть збігатися для різних моментів часу, але й їх синхронності в часі.

Для визначення послідовності $\hat{q}[k], k = 1, \dots, K_0$ будемо, як і раніше, використовувати співвідношення (4.33), але замість (4.34) для пошуку «найближчих» векторів q_{mk}^{\min} використаємо евклідову відстань:

$$\rho_\tau(u_0[\cdot], u_m[\cdot]) = \sqrt{(z_0^*[\cdot] - z_m^*[\cdot])^2 + (\dot{z}_0^*[\cdot] - \dot{z}_m^*[\cdot])^2 + (\tau_0[\cdot] - \tau_m[\cdot])^2}. \quad (4.36)$$

Результати доказових експериментів з використанням запропонованих моделей генерації штучних ЕКГ реалістичної форми показали, що оброблення сигналу у фазовому просторі забезпечує високу точність оцінювання еталону. Навіть в умовах високих внутрішніх спотворень середня похибка в оцінюванні значень діагностичних ознак (амплітуди та тривалості зубця T , і зсуву сегмента ST), не перевищувала 1,5–2,5 %. Водночас під час використання традиційного алгоритму усереднення у часовій області межі помилок досягали 15–20 %. І це не випадково, оскільки майже на всіх експериментах усереднення в часовій області призводило до «розмивання» форми інформативних фрагментів.

4.2. Оцінювання ступеня хаотичності сигналів

Під час вивчення динаміки поведінки складних медико-біологічних систем все більшу увагу привертають методи теорії хаосу і синергетики, які дають змогу повніше розкрити і проаналізувати механізми функціонування живої складноорганізованої системи. Застосовують ці методи і в кардіології для оцінювання хаотичності серцевого ритму, який має інформацію про функціональний стан усіх ланок регулювання життєдіяльності людини, як у нормі, так і за різних патологій. Таке оцінювання зазвичай проводиться за динамічним рядом тривалостей серцевих циклів (RR -інтервалів), які визначаються в процесі реєстрації ЕКГ.

Водночас лише з аналізу хаотичності ритму серця не можна повноцінно зробити висновок про функціональний стан самого серця як основного системоутворюючого органу. Тому разом з оцінюванням хаотичності тривалостей RR -інтервалів доцільно проаналізувати хаотичності інших інформативних показників ЕКГ, що мають інформацію про форму фрагментів ЕКГ від циклу до циклу – зубців P, Q, R, S, T , інтервалів PQ, QT і сегмента ST .

Розглянемо спочатку математичні основи теорії інформації та методи оцінювання хаотичності динамічних рядів.

В основі методів є відома формула інформаційної ентропії

$$H = -\sum_{i=1}^N p_i \log_2 p_i, \quad (4.37)$$

запропонована К. Шенноном для оцінювання невизначеності системи, яка перебуває в одному зі N станів з ймовірностями p_i , $\sum_{i=1}^N p_i$. Чим більша величина H , тим більше система відхиляється від упорядкованого стану, причому максимальне значення ентропії досягається, коли стани системи рівноможливі, тобто $p_1 = p_2 = \dots = p_N$.

На перший погляд здається, що для аналізу хаотичності кінцевого часового ряду

$$A = a_1, a_2, \dots, a_N, \quad (4.38)$$

елементи якого є скалярні величини $a_i \in [a_i^-, a_i^+]$, $i = 1, \dots, N$, що належать заданим інтервалам, можна скористатися формулою (4.5) на ос-

нові оцінювання ймовірностей (частот) потрапляння значень (4.38) у зазначені інтервали.

Однак це буде неправильно, оскільки величина (4.37) інваріантна щодо перестановок елементів ряду (4.38). Звідси випливає, що безпосереднє використання формули (4.37) оцінює не хаотичність послідовності

(4.38), а невизначеність випадкової величини, яка породжує цю послідовність, що не те саме.

Наприклад, дві послідовності – регулярна

$$A_1 = 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0$$

і хаотична

$$A_2 = 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1$$

матимуть однакові значення ентропії $H = 1$.

Оцінити хаотичність часового ряду під час використання формули (4.41) можна лише в тому разі, якщо послідовність (4.38) обробляти частинами, наприклад, за допомогою умовної ентропії (**Conditional Entropy**). Для цього послідовність (4.38) розділяється на підпослідовності (патерни)

$$x(i) = [a(i), a(i+1), \dots, a(i+m-1)], \quad i = 1, \dots, N-m+1 \quad (4.39)$$

з розмірністю вкладення m , оцінюються ймовірності (частоти) виникнення конкретних патернів і обчислюється умовна ентропія $E(m|m-1)$ як збільшення шенонівської ентропії під час переходу від патернів з розмірністю $m-1$ до m , тобто

$$E(m|m-1) = E(m) - E(m-1) = - \sum_{i=1}^{N-m+1} p_m \ln p_m + \sum_{i=1}^{N-m+2} p_{m-1} \ln p_{m-1}. \quad (4.40)$$

Зауважимо, що перехід від послідовності скалярних величин (4.38) до векторів (4.39) можна інтерпретувати як перехід у псевдофазовий простір, в якому проводиться подальший аналіз.

Для повнішого аналізу складності медико-біологічних систем стали популярними інші ентропійні оцінювання, зокрема, апроксимаційна ентропія (**Approximation Entropy**), яку називають також ентропією подібності.

Під час її обчислення вихідна послідовність також розділяється на патерни (4.39), близькість яких у фазовому просторі оцінюють відстанню

$$d[x(i), x(j)] = \max_{k=1, \dots, m} \{a(i+k-1) - a(j+k-1)\} \quad (4.41)$$

між усіма парами $x(i)$ та $x(j)$, $i = 1, \dots, N - m + 1$, $j = i, \dots, N - m + 1$. Далі оцінюються ймовірності (частоти) виникнення в послідовності (4.6) таких пар патернів, відстань між якими не перевищує заданий поріг d_0 .

Для цього визначають величини

$$C_r^{(m)}(i) = \frac{U^{(m)}(i)}{N - m + 1}, \quad (4.42)$$

в яких $U^m(i)$ – кількість значень $d[x(i), x(j)]$, що задовольняють умову

$$d[x(i), x(j)] \leq d_0, \quad j = 1, \dots, N - m + 1, \quad (4.43)$$

або, в еквівалентній формі запису,

$$C_r^{(m)}(i) = \frac{1}{N - m + 1} \sum_{j=1}^{N-m+1} \Theta(d_0 - d[x(i), x(j)]), \quad (4.44)$$

де

$$\Theta(\eta) = \begin{cases} 1, & \text{коли } \eta \geq 0, \\ 0, & \text{коли } \eta < 0 \end{cases} \quad (4.45)$$

є функцією Хевісайда.

Використовуючи вирази (4.43) або (4.44), можна обчислити величину

$$\theta^{(m)}(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_r^{(m)}(i), \quad (4.46)$$

яку називають безумовною ентропією.

Апроксимаційна ентропія (**Approximate entropy**) визначається як збільшення безумовної ентропії (4.46) під час переходу від послідов-

ності патернів завдовжки m до послідовності патернів завдовжки $m + 1$ за формулою

$$ApEn = \theta^{(m)}(r) - \theta^{(m+1)}(r). \quad (4.47)$$

Зазвичай для розрахунку використовують наближену формулу

$$ApEn = \frac{1}{N - m} \sum_{i=1}^{N-m} \ln \frac{C_r^m}{C_r^{m+1}}. \quad (4.48)$$

Значення апроксимаційної ентропії для наведених вище регулярної послідовності $A_1 = 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0$ та хаотичної послідовності $A_2 = 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1$ будуть відповідно становити

$$ApEn(A_1) = 0,08 \text{ і } ApEn(A_2) = 0,4.$$

Слід зауважити, що оцінювання (4.48) є зміщеним оцінюванням апроксимаційної ентропії, значення якої асимптотично зростають із збільшенням кількості елементів часового ряду. Для усунення цього запропоновано невелику модифікацію методу обчислення, що ґрунтується на величині

$$C_r^{(m)}(i) = \frac{\sigma + U^{(m)}(i)}{N - m + 1}, \quad (4.49)$$

чисельник якої містить фіксовану домішку $\sigma > 0$, а в умову (4.44) вводиться додаткове обмеження $i \neq j$.

Подальша модифікація апроксимаційної ентропії – ентропія шаблонів (**Sample entropy**), обчислення якої також виконується за обмеженням $i \neq j$, а саму ентропію визначає вираз

$$SampEn = -\ln \frac{U^{m+1}(r)}{U^m(r)} = \ln U^m(r) - \ln U^{m+1}(r). \quad (4.50)$$

Таке оцінювання, на відміну від (4.48), є незміщеним та його значення майже не залежить від кількості елементів часового ряду.

Ще одним популярним засобом оцінювання хаотичності часових рядів є переставна ентропія PE (**Permutation Entropy**), яка ґрунтується на аналізі патернів не з погляду відстані між ними у фазовому просторі, а з погляду їх форми (класу).

Для класифікації патернів кожному елементу послідовності $a(i)$, $i = 1, 2, \dots, N$ присвоюється позначка, що характеризує рівень елемента відносно сусіднього, наприклад, позначка 0, 1, 2. Тоді можна розрізняти $3!$ (шість) класів патернів $\pi_1, \pi_2, \dots, \pi_6$, утворених трійкою послідовних значень відповідного рівня (рис. 4.19).

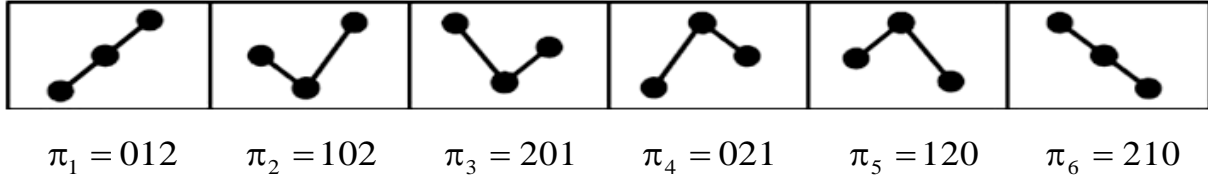


Рис. 4.19. Класи патернів традиційної переставної ентропії ($m = 3$).

Переставну ентропію обчислюють за формулою

$$PE(m) = - \sum_{j=1}^{N-m+1} p(\pi_i) \log_2 p(\pi_i), \quad i = 1, \dots, m!, \quad (4.51)$$

в якій $p(\pi_i)$ частота виникнення патерна i -го класу під час оброблення послідовності (4.38) ковзним вікном з m точок.

Значення переставної ентропії (4.51) для регулярної послідовності $A_1 = 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0$ та хаотичної послідовності $A_2 = 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1$ становлять

$$PE(A_1) = 1 \text{ та } PE(A_2) = 1,96.$$

Під час розв'язування прикладних задач, зокрема, задачі оброблення ЕКГ, виявилось доцільним дещо модернізувати відомий підхід в обчисленні переставної ентропії, аналізуючи у ковзному вікні трійки значень a_{i-1}, a_i, a_{i+1} , $i = 2, \dots, N - 1$ послідовності (4.38). В цьому випадку переставна ентропія враховує частоти виникнення вже таких п'яти класів патернів (рис. 4.20).



Рис. 4.20. Класи патернів модернізованої переставної ентропії ($m = 3$).

Класи патернів визначають послідовною перевіркою таких умов:

клас 1 «**Максимум**»,
 якщо $(a_i - a_{i-1}) > h \wedge (a_i - a_{i+1}) > h$,
 Клас 2 «**Мінімум**»,
 якщо $(a_{i-1} - a_i) > h \wedge (a_{i+1} - a_i) > h$,
 Клас 3 «**Зростання**»,
 якщо $(a_i - a_{i-1}) > h \vee (a_{i+1} - a_i) > h \vee (a_{i+1} - a_{i-1}) > h$,
 клас 4 «**Спадання**»,
 якщо $(a_{i-1} - a_i) > h \vee (a_i - a_{i+1}) > h \vee (a_{i-1} - a_{i+1}) > h$,
 клас 5 «**Постійний**»,
 якщо не виконується жодне з наведених вище умов,

де h – поріг нечутливості до локальних змін сигналу.

Якщо елементи часового ряду (4.38) утворюють не скалярні величини a_i , а R -вимірні вектори $\vec{a}_i = (a_{i1}, \dots, a_{iR})$, для інтегрального оцінювання хаотичності ряду використовують кореляційний інтеграл. Для його обчислення компоненти векторів $\vec{a}_i = (a_{i1}, \dots, a_{iR})$ попередньо нормують і подальше оброблення здійснюють у фазовому просторі з нормованими координатами $a_{i1}^*, \dots, a_{iR}^*$.

Для скінченної послідовності спостережень оцінювання кореляційного інтегралу має вигляд

$$C(d_0, N) = \frac{1}{N^2} \sum_{m,l=1}^N \Theta(d_0 - d(\vec{a}_l, \vec{a}_m)), \quad (4.52)$$

де $d(\vec{a}_l, \vec{a}_m)$ – евклідова відстань між l -м і m -м векторами в R -вимірному фазовому просторі; d_0 – заданий поріг, а $\Theta(\cdot)$ – функція Хевісайда, що визначається за співвідношенням (4.45).

З виразу (4.52) випливає, що кореляційний інтеграл оцінює середню ймовірність того, що відстань між нормованими векторами $\vec{a}_i = (a_{i1}, \dots, a_{iR})$ не перевищує заданий поріг d_0 .

Відомо, що для фрактальних (самоподібних) множин характерна така залежність:

$$\lim_{d_0 \rightarrow 0} C(d_0, N) = d_0^D, \quad (4.53)$$

де D – розмірність атрактора (**Fractal dimension**), яка також оцінює хаотичність часових сигналів у R -вимірному фазовому просторі.

З (4.53) випливає, що за обчисленим кореляційним інтегралом (4.52), можна визначити розмірність атрактора за формулою

$$D = \lim_{d_0 \rightarrow 0} \frac{\log(C(d_0))}{\log d_0}. \quad (4.54)$$

На підставі (4.54) можна зробити висновок, що фрактальну розмірність можна оцінити за нахилом прямої лінії регресії, побудованої за експериментальними даними в координатах $\ln C(d_0)$ і $\ln d_0$.

Для зручності запишемо цю залежність у явному вигляді

$$D = \frac{N \sum_{i=1}^N \log d_0(i) \cdot \log(C(d_0(i))) - \sum_{i=1}^N \log d_0(i) \sum_{i=1}^N \log(C(d_0(i)))}{N \sum_{i=1}^N (\log d_0(i))^2 - (\sum_{i=1}^N \log d_0(i))^2}. \quad (4.55)$$

Формулу (4.55) доцільно використовувати під час розв'язування прикладних задач.

Нагадаємо ще один відомий підхід до обчислення фрактальної розмірності, який ґрунтується на розмірності Мінковського D_{bc} (**Box-counting dimension**). Метод полягає у визначенні мінімальної кількості «осередків» фазового простору діаметром $d \rightarrow 0$, якими можна «накрити» вихідну множину точок.

Після низки перетворень отримаємо в явному вигляді остаточну формулу для обчислення розмірності Мінковського як кутовий коефіцієнт прямої лінії регресії, побудованої за експериментальними точками в координатах $\ln M(d), \ln d_0(i)$:

$$D_{bc} = \frac{N \sum_{i=1}^N \ln d_0(i) \cdot \ln M(d(i)) - \sum_{i=1}^N \ln d_0(i) \sum_{i=1}^N \ln M(d(i))}{N \sum_{i=1}^N (\ln d_0(i))^2 - (\sum_{i=1}^N \ln d_0(i))^2}. \quad (4.56)$$

Ще один досить простий і ефективний метод аналізу хаотичності нестационарних часових рядів – R/S -аналіз, який ґрунтується на обчисленні індексу Херста (**Hurst exponent**) за модифікованою часовою послідовністю:

$$z_i = \sum_{i=1}^k (a_i - a_{cp}), \quad i = 1, \dots, k, \quad (4.57)$$

де $k \leq N$, a_{cp} – середнє значення елементів у послідовності (4.42), що містить N елементів, або ж за іншою модифікованою послідовністю у вигляді логарифмів відносин

$$z_i = \log\left(\frac{a_{i+1}}{a_i}\right), \quad i = 1, 2, \dots, N-1 \quad (4.58)$$

Для обчислення індексу Херста достатньо визначити розмах

$$R^{(k)} = \max_{i=1, \dots, k} (z_i) - \min_{i=1, \dots, k} (z_i) \quad (4.59)$$

і стандартне відхилення

$$S^{(k)} = \sqrt{\frac{1}{k-1} \sum_{i=1}^k (a_i - a_{cp})^2}, \quad (4.60)$$

для певних фіксованих значень $k \leq N$.

Встановлено, що для більшості стохастичних рядів виконується умова

$$R^{(k)} / S^{(k)} = Lk^{H_0}, \quad (4.61)$$

тобто нормований розмах пропорційний ступеню k , де H_0 – показник Херста, а L – деяка константа.

Для визначення показника Херста у зручній формі прологарифмуємо (4.61):

$$\log(R^{(k)} / S^{(k)}) = \log L + H_0 \log k \quad (4.62)$$

та знайдемо значення індексу Херста у вигляді кутового коефіцієнта прямої лінії регресії (4.62), яка найкраще апроксимує експериментальні точки $\log(R^{(k)} / S^{(k)})$ і $\log k$, отримані для різних фіксованих значень $k \leq N$.

У результаті, після низки очевидних перетворень, які ґрунтуються на обробленні значень часового ряду для всіх $k = 1, \dots, N$ методом найменших квадратів, отримаємо

$$H_0 = \frac{\log k \sum_{k=1}^N \log k \log(R^{(k)} / S^{(k)}) - \sum_{k=1}^N \log k \sum_{k=1}^N \log(R^{(k)} / S^{(k)})}{\log k \sum_{k=1}^N (\log k)^2 - (\sum_{k=1}^N \log k)^2}. \quad (4.63)$$

З першого погляду може виникнути враження, що ентропійне оцінювання хаотичності динамічного ряду (4.38) має таку саму діагностичну цінність, як і середньоквадратичне відхилення. Однак у загальному випадку це не правильно.

Відомо, що для деяких розподілів випадкової величини X , яка породжує ряд (4.38), шененонівська ентропія пов'язана з середньоквадратичним відхиленням певної залежності. Наприклад, якщо величина має безперервний нормальний розподіл, то

$$\begin{aligned} H &= - \int_{-\infty}^{+\infty} p(X) \log p(X) dX = - \int_{-\infty}^{+\infty} p(X) \left[\log \frac{1}{\sqrt{2\pi}\sigma} - \frac{X^2}{2\sigma^2} \log e \right] dX = \\ &= - \log(\sqrt{2\pi}\sigma)^{-1} \int_{-\infty}^{+\infty} p(X) dX + \frac{\log e}{2\sigma^2} \int_{-\infty}^{+\infty} p(X) X^2 dX = \log \sigma \sqrt{2\pi} + \frac{1}{2} \log e = \\ &= \log \sigma \sqrt{2\pi e} = 0,5 \log_2 2\pi e + \log_2 \sigma. \end{aligned}$$

Звідси випливає, що для неперервного нормального розподілу є залежність між ентропією H і середньоквадратичним відхиленням σ :

$$H \cong 2,05 + \log_2 \sigma. \quad (4.64)$$

Для експериментальної перевірки залежності H і σ проводились модельні експерименти за допомогою генерації послідовностей незалежних нормально розподілених величин a_i , $i = 1, \dots, N$ з нульовим середнім і $\sigma = 0,1$ за різних N . Модельні експерименти показали, що коли $N \geq 400$ теоретична та експериментальна залежності мають майже однакову форму (рис. 4.21).

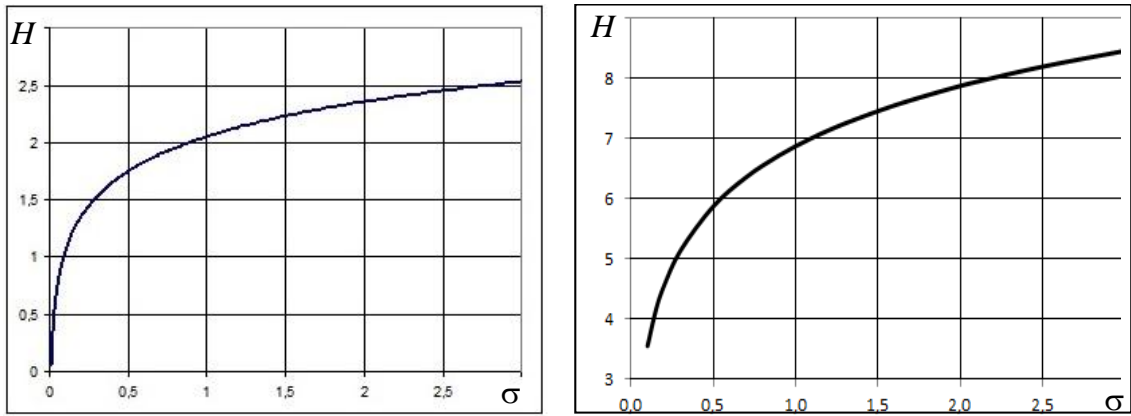


Рис. 4.21. Теоретична (ліворуч) та експериментальна (праворуч) залежності ентропії від середньоквадратичного відхилення за нормальним розподілом ($N = 400$ точок)

Нелінійний характер залежностей (4.64) дозволяє зробити важливий висновок. За нормальним розподілом випадкової величини, що породжує послідовність (4.38), навіть незначні зміни середньоквадратичного відхилення σ в області малих значень призводять до великих змін ентропії H , тоді як зміна σ на той самий відсоток в області великих значень майже не призводить до будь-яких змін H . Межу зазначених областей визначає поріг

$$\sigma_0 \approx 1,443, \quad (4.65)$$

який задовольняє умову $\partial H / \partial \sigma = 1$.

Для ілюстрації розглянемо результати оцінювання хаотичності двох модельних сигналів, що є послідовністю незалежних нормально розподілених величин з нульовими середнім і різними значеннями σ на першій і другій половині спостережень (рис. 4.22).

Перший сигнал складається з двох фрагментів по $N = 400$ точок з параметрами $\sigma_{11} = 0,10$ і $\sigma_{12} = 0,13$ відповідно (рис. 4.22, а), а другий сигнал складається з двох фрагментів по $N = 400$ точок з параметрами $\sigma_{21} = 2,70$ і $\sigma_{22} = 3,51$ відповідно (рис. 4.22, б). Таким чином, на другій половині обох сигналів спостерігається однаковий приріст $\Delta\sigma$ середньоквадратичного відхилення σ порівняно з першою половиною, що дорівнює $\Delta\sigma = 30\%$. Однак самі значення σ першого сигналу належать області $\sigma < \sigma_0$, але другого сигналу – області $\sigma > \sigma_0$, де $\sigma_0 \approx 1,443$ – граничне значення, що визначається співвідношенням (4.65).

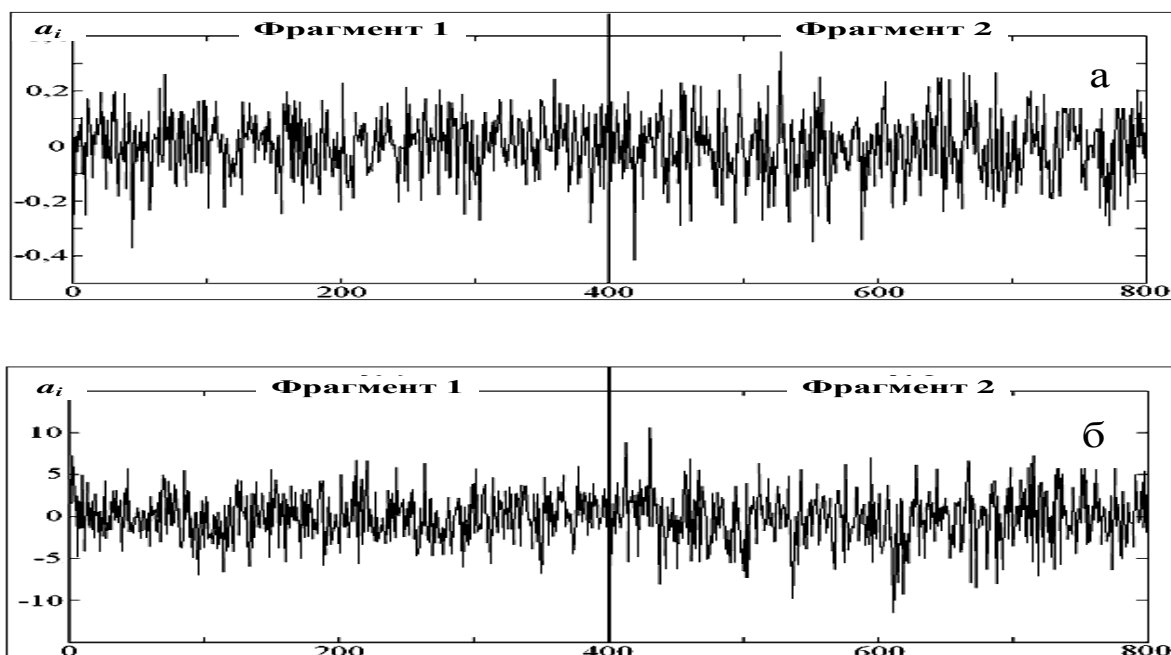


Рис. 4.22. Модельні сигнали з низьким (а) та високим (б) значеннями середньоквадратичного відхилення σ

У табл. 4.1 наведено результати обчислення середньоквадратичних відхилень σ та ентропій H на фрагментах зазначених рядів.

Таблиця 4.1. Результати моделювання

Тестовий сигнал	Середнє квадратичне відхилення		Ентропія	
	σ , од.	$\Delta\sigma$	H , од.	ΔH
Часовий ряд 1	0,1	30 %	1,90	20 %
	0,13		2,29	
Часовий ряд 2	2,7	30 %	4,93	2 %
	3,51		5,05	

Проведемо тепер порівняльний аналіз результатів оцінювання мінливості динамічного ряду з використанням ковзних вікон. Модельний сигнал було згенеровано як послідовність з $M = 10$ фрагментів, кожен із яких складався з точок $K_0 = 50$ і породжувався моделлю авторегресії вигляду

$$a_{kl} = \lambda_l \cdot a_{k-1} + (1 - \lambda_l) \cdot \xi_{k-1}, \quad k = 1, \dots, K_0, \quad l = 1, \dots, M, \quad (4.66)$$

де λ_l ($0 < \lambda_l < 1$) – параметр, що визначає мінливість сигналу на l -му фрагменті; $a_{0l} = 0,1$ – початкове значення; $\xi \in N(0, \sigma^2)$ – послідов-

ність незалежних нормально розподілених випадкових чисел з нульовим математичним сподіванням і дисперсією $\sigma^2 = 3$.

Таким чином, увесь сигнал містив $N = 500$ точок. Значення параметрів $\lambda_1, \lambda_2, \dots, \lambda_{10}$ на фрагментах тестового сигналу наведено у табл. 4.2, а графік цього сигналу показано на рис. 4.23.

Таблиця 4.2. Значення параметрів λ_l на фрагментах сигналу

l	1	2	3	4	5	6	7	8	9	10
λ_l	0,93	0,86	0,75	0,1	0,2	0,3	0,6	0,7	0,8	0,9

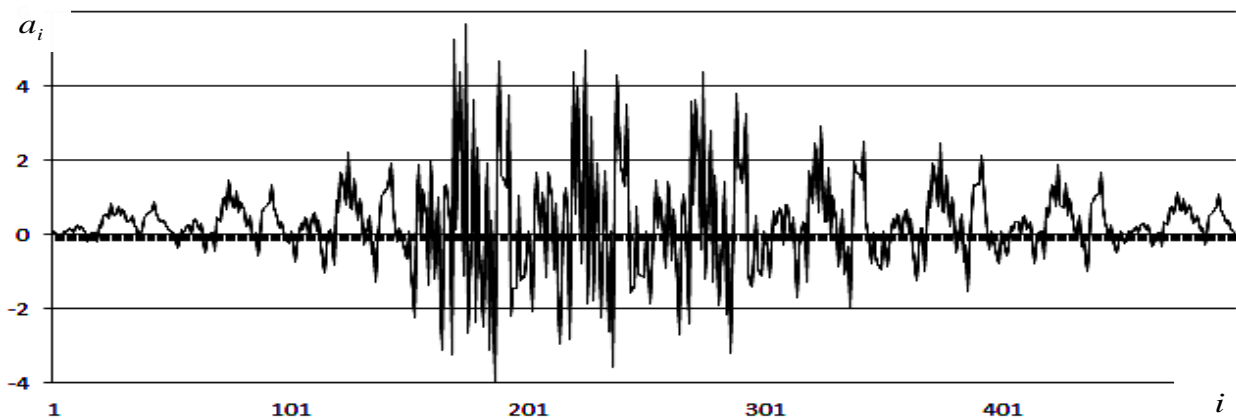


Рис. 4.23. Тестовий сигнал з 10 фрагментів

Результати оцінювання мінливості тестового сигналу за допомогою ковзних вікон, в яких обчислювалися ентропія H та середнє квадратичне відхилення (СКВ) σ наведено на рис. 4.24. Можна побачити, що обидва показники показали схожі результати – збільшення σ супроводжувалося збільшенням H , і навпаки. Однак в області значень $\sigma > \sigma_0$ чутливість ентропії до мінливості сигналу майже в 2 рази менша, ніж чутливість СКВ, що узгоджується з результатами дослідження теоретичної залежності між цими величинами.

Розглянемо результати аналізу хаотичності одного й того ж динамічного ряду з використанням різних ентропійних оцінок. Тестовий сигнал (рис. 4.25) генерувався за допомогою моделі авторегресії

$$X_k = bX_{k-1} + (1-b)\xi_{k-1}, \quad k \in [0; N-1] \quad (4.67)$$

з початковим значенням $X_0 = 0,1$, в якому b – параметр налаштування ($0 \leq b \leq 1$), $N = 1000$ – кількість дискретних відліків, $\xi \in N(0, \sigma^2)$ – вектор незалежних нормально розподілених випадкових величин з

нульовим математичним сподіванням і обмеженою дисперсією $\sigma^2 = 10^2$.

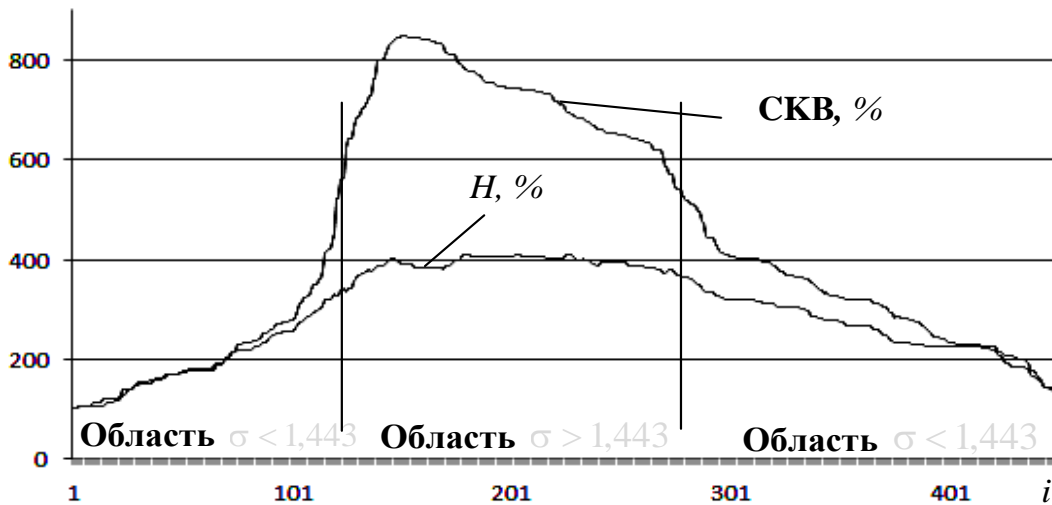


Рис. 4.24. Динаміка змін СКВ та ентропії H під час оброблення тестового сигналу (4.66)

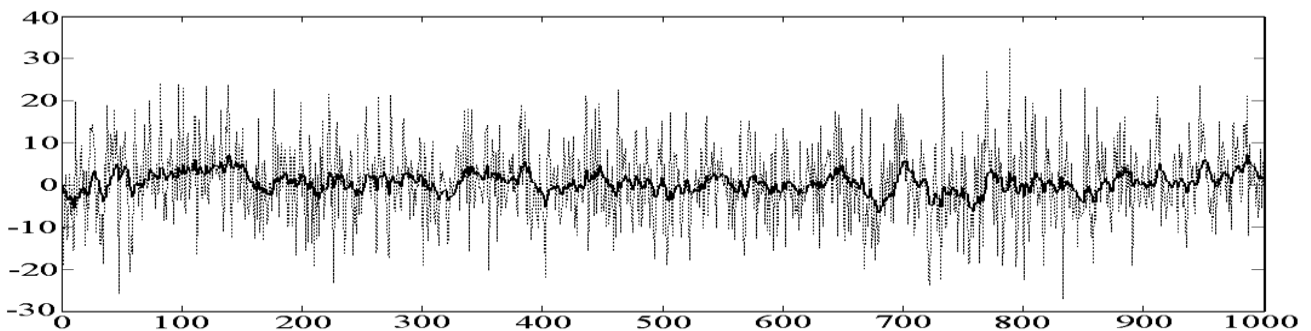


Рис. 4.25. Графіки модельного сигналу (4.67) з параметрами $b = 0,9$ (чорний) і $b = 0,1$ (сірий)

Усереднені результати оцінювання ступеня хаотичності тестового сигналу (4.67) у серії експериментів за різних значень параметра b подано в табл. 4.3. Ступінь хаотичності, обчислений різними методами, наведено в абсолютних значеннях (x , од.) і відносних величинах (δ , %) відносно її значення, отриманого на тестовому сигналі (4.67) з показником $b = 0,9$.

Як видно з таблиці, значення всіх ентропійних оцінок збільшувалися зі збільшенням хаотичності сигналу (зменшення значення параметра b), причому найчутливішою виявилася умовна ентропія $E(m | m - 1)$. За значення параметра $b = 0,9$ показник Херста набував значення $H_0 = 0,73$, тобто належав області $0,5 < H_0 \leq 1$ стійких (персистентних) сигналів з ознаками самоподібності. Тому зі збільшен-

ням ступеня хаотичності індекс Херста, на відміну від ентропійного оцінювання, зменшувався.

Таблиця 4.3. Усереднені результати оцінювання хаотичності тестового сигналу (4.67) у серії модельних експериментів

Метод оцінювання	Параметр хаотичності b тестового сигналу									
	0,9		0,7		0,5		0,3		0,1	
	x	x	$\delta, \%$	x	$\delta, \%$	x	$\delta, \%$	x	$\delta, \%$	
Умовна ентропія $E(m m - 1)$	0,73	0,99	35,8	1,13	55,5	1,19	62,8	1,21	65,3	
Апроксимаційна ентропія $ApEn$	1,06	1,26	19,1	1,3	23	1,31	24,6	1,32	25,2	
Ентропія шаблонів $SampEn$	1,40	1,86	33,8	2,05	47,2	2,16	55,2	2,18	56,3	
Переставна ентропія PE	1,29	1,46	14	1,49	16	1,5	16,7	1,51	17,7	
Індекс Херста H_0	0,73	0,65	-10,4	0,61	-16,2	0,58	-20,8	0,55	-24,7	

Зрозуміло, наведений аналіз наявних методів оцінювання хаотичності динамічних послідовностей не може бути повним, тим більше що методи оцінювання хаотичності часових рядів постійно розвиваються. Ми розглянули лише найпопулярніші підходи та отримали остаточні формули, зручні для практичного використання.

Можна зробити висновок, що сучасна наука має ефективні методи й обчислювальні алгоритми, які дають змогу з різного погляду оцінити ступінь хаотичності часових рядів, що породжуються.

Важливу інформацію про властивості динамічної системи має як сама ентропія, так й її зміни у часі. На основі аналізу таких змін В. С. Аніщенко виявив гендерні відмінності реакції організму на стресові дії навколишнього середовища.

Розглянемо ефективний метод оцінювання змін у часі ентропії показників, які характеризують динаміку форми циклів сигналу. Одним з таких показників є показник β , який характеризує симетрію зубця T ЕКГ.

Визначимо послідовність,

$$\beta_1, \beta_2, \dots, \beta_M, \quad (4.68)$$

що відображає динаміку симетрій зубців T на послідовних циклах ЕКГ. Будемо оброблювати послідовність (4.68) ковзним вікном, що містить W точок і в кожному l -му вікні, оцінимо процентне прирощенні ентропії відносно ентропії в першому вікні, тобто обчислимо співвідношення

$$h_l = \frac{H_l}{H_1} \cdot 100 \% , \quad l=1, \dots, L, \quad (4.69)$$

вважаючи, що $H_1 \neq 0$.

Принципово для обчислення H_l можна використати будь-яке ентропійне оцінювання, аналіз яких наведено вище, наприклад, модернізовану переставну ентропію. У цьому випадку відношення (4.73) набуває вигляду

$$h_l = \frac{-\sum_{j=1}^5 p(\pi_{jl}) \log p(\pi_{jl})}{H_1} \cdot 100 \% , \quad l=1, \dots, L , \quad (4.70)$$

де $p(\pi_j)$ – частота виникнення одного з п'яти характерних патернів π_1, \dots, π_5 (рис. 4.20) в l -му вікні.

Процедуру (4.70) можна реалізувати під час зсуву вікна $l+1$ відносно l -ї ширини вікна W або під час зміщення вікон на одну точку (режим ковзного вікна). Зрозуміло, що в останньому випадку обсяг необхідних обчислень буде більшим, але графік зміни ентропії h_l буде більш плавним, причому вигляд графіка залежить від ширини вікна W і порога δ нечутливості до зміни значень динамічного ряду, що обробляється (рис. 4.26).

Оскільки немає формальних методів визначення оптимальних значень W і δ , під час розв'язування прикладних задач значення цих параметрів обирають експериментально з урахуванням прагматичного критерію – максимуму середнього квадрата відхилення змін ентропії представників досліджуваних груп. Наприклад, груп із різним рівнем адаптаційних резервів.

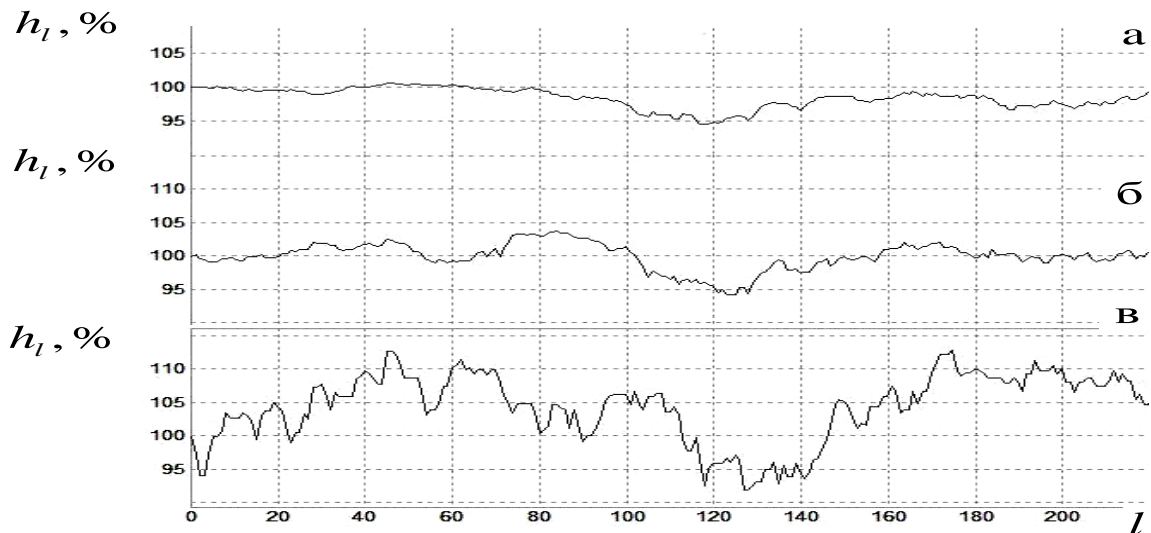


Рис. 4.26. Графіки ковзної ентропії $h(l)$:

$W = 100$ точок, $\delta = 0,1$ од. (а); $W = 60$ точок, $\delta = 0,04$ од. (б);
 $W = 30$ точок, $\delta = 0,02$ од. (в)

З цією метою проводять серію обчислювальних експериментів з різними значеннями W та δ та оцінюють значення функції $\Psi(W, \delta)$ у вигляді середнього квадрата відмінностей значень ентропії $h_1(i)$ та $h_2(i)$, обчислених для пар представників першої та другої груп:

$$\Psi(W, \delta) = \frac{1}{M-1} \sum_{i=1}^M [h_1(i) - h_2(i)]^2, \quad (4.71)$$

де M – число представників груп в експериментальній вибірці.

На рис. 4.27 наведено графік функції $\Psi(W, \delta)$, побудований для спортсменів та не спортсменів. Як видно, взяті значення параметрів налаштування, які забезпечують найбільш виражені відхилення ентропій $h_1(i)$ та $h_2(i)$ у групах, перебувають у досить широкому діапазоні.

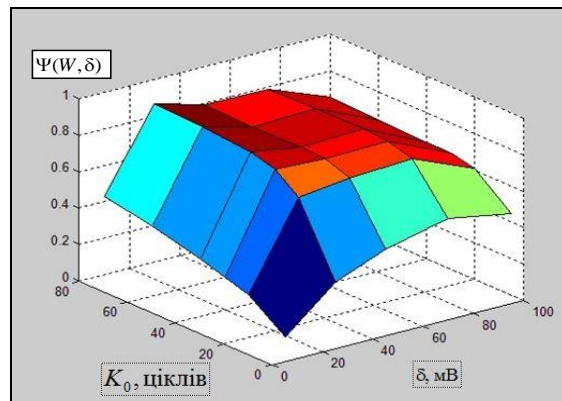


Рис. 4.27. Графік функції $\Psi(W, \delta)$

Слід зазначити, що швидкість зміни ентропії активно вивчається в термодинамічних системах. Незважаючи на те, що питання взаємозв'язку термодинамічної ентропії Больцмана та інформаційної ентропії Шеннона досі залишається предметом наукових дискусій, такі аналогії все ж таки корисні.

Застосуємо процедуру чисельного диференціювання (рис. 4.4) і оцінимо похідну $\dot{h}(l)$ в l -й точці. Це дозволяє перейти від дискретних значень ковзної ентропії $h(l)$ до її фазового портрета на площині $h(l), \dot{h}(l)$ і побудувати його опуклу оболонку (рис. 4.28).

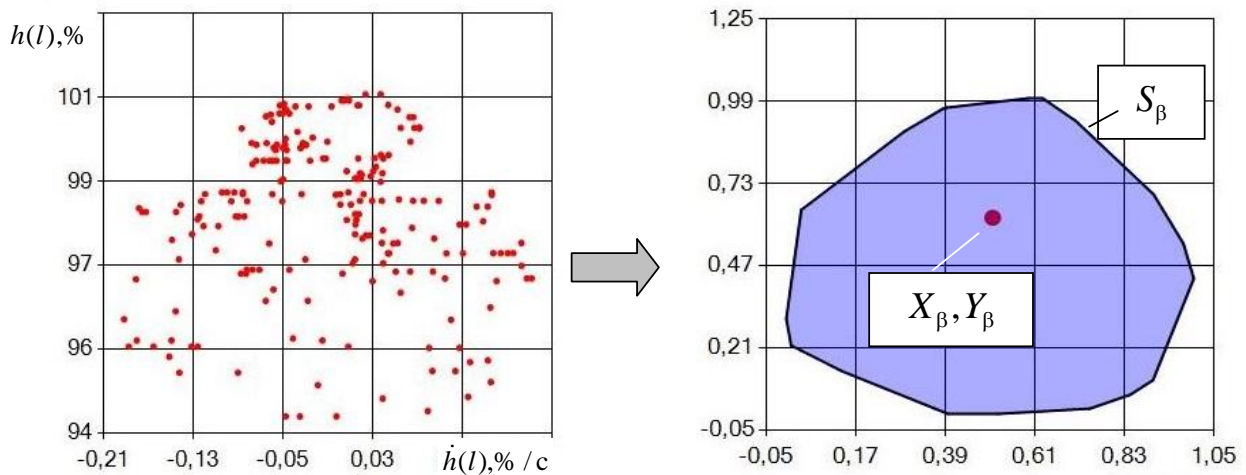


Рис. 4.28. Фазовий портрет ковзної ентропії (ліворуч) та його опукла оболонка (праворуч)

Оскільки, як вже зазначалось, інформацію про властивості динамічної системи має не тільки ентропія, а й її зміни у часі, можна припустити, що форма фазового портрета ентропії та його «розмір» на фазовій площині має додаткову інформацію про поточний стан об'єкта, зокрема, про стан серцево-судинної системи.

Тому для інтегрального оцінювання хаотичності показника β за послідовністю (4.68) будемо використовувати площу опуклої оболонки S_β фазового портрета ентропії, координати центру тяжкості та низку інших показників. Аналогічно можна оцінювати інтегральні показники, що характеризують хаотичність інших параметрів форми циклічного сигналу, зокрема амплітуд та тривалості зубців ЕКГ.

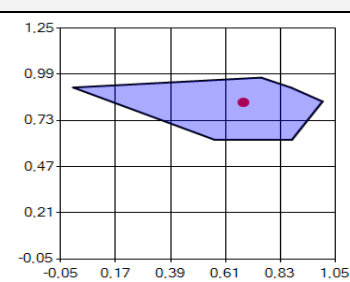
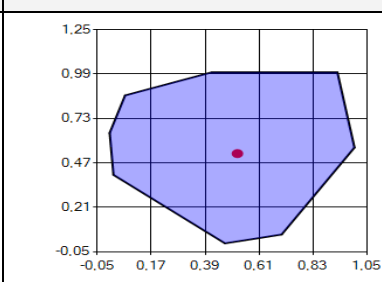
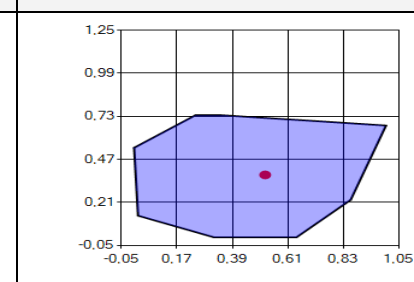
Розглянемо результати, які демонструють діагностичну ефективність аналізу фазового портрета ентропії (ФПЕ).

У табл. 4.4 подано результати тестування на тредмілі умовно здорового добровольця 44 років. Швидкість стрічки тредміли досяга-

ла 5,5 км/год, а кут нахилу поступово збільшувався до 14 %, після чого був відпочинок протягом трьох хвилин. У таблиці наведено динаміку зміни інтегрального показника S_β – площі опуклої оболонки фазового портрета перестановочної ентропії, побудованої для показника β_T .

Можна побачити, що на висоті навантаження показник збільшився майже на 250 % відносно початкового стану, а в період відпочинку поступово повертався до вихідних значень.

Таблиця 4.4. Динаміка показника S_β під час тредміл тесту

Спокій	Навантаження 15 хв	Відпочинок 3 хв
		
$S_\beta = 0,217$ од.	$S_\beta = 0,749$ од.	$S_\beta = 0,58$ од.

З таблиці видно, що динаміка інтегрального показника S_β під час навантаження мала протилежну тенденцію порівняно з динамікою стандартного показника $SDNN$, що характеризує середньоквадратичне відхилення RR -інтервалів (рис. 4.29), який під час навантаження зменшувався швидше за все через зростаючий вплив симпатичної частини вегетативної нервової системи.

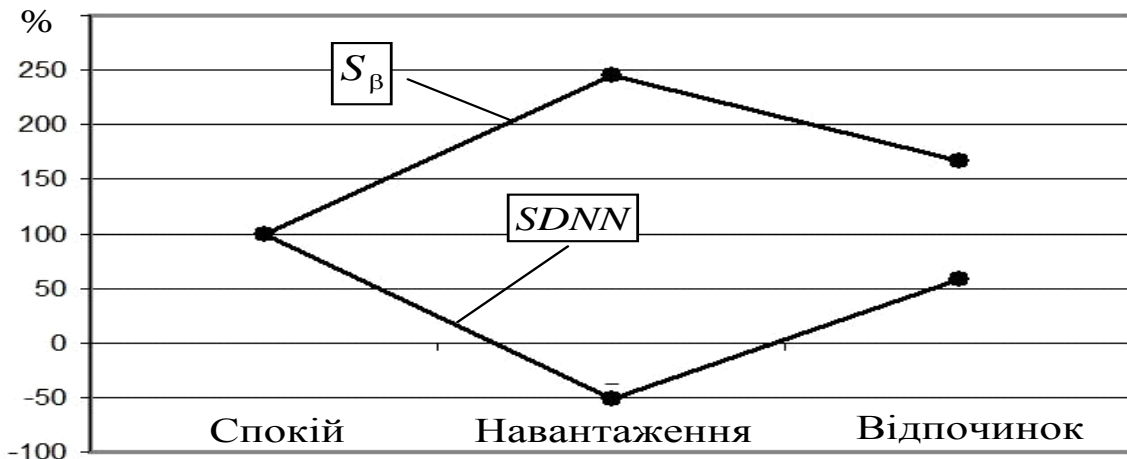


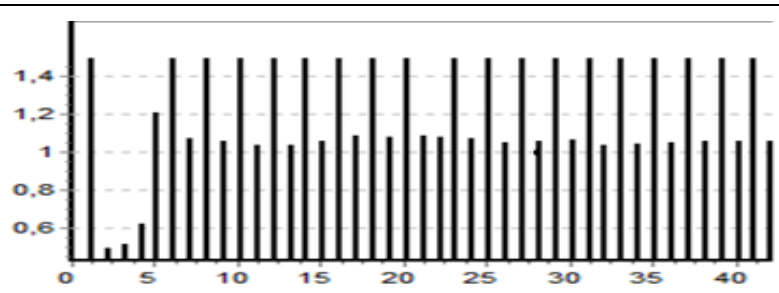
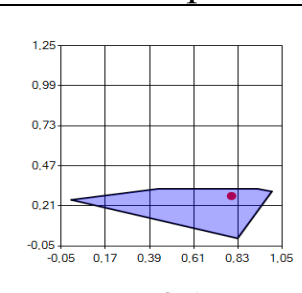
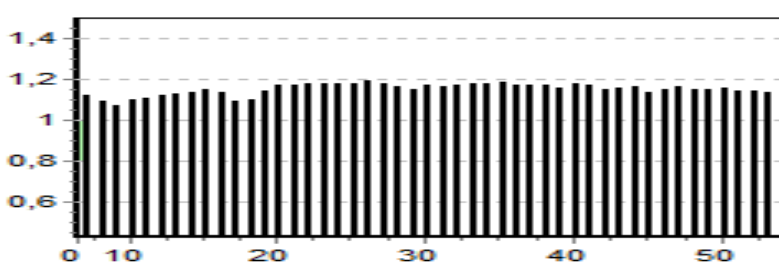
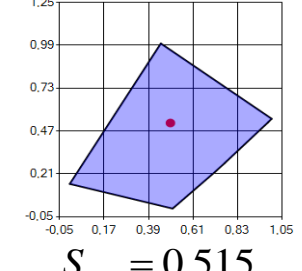
Рис. 4.29. Динаміка показників S_β і $SDNN$ під час навантаження

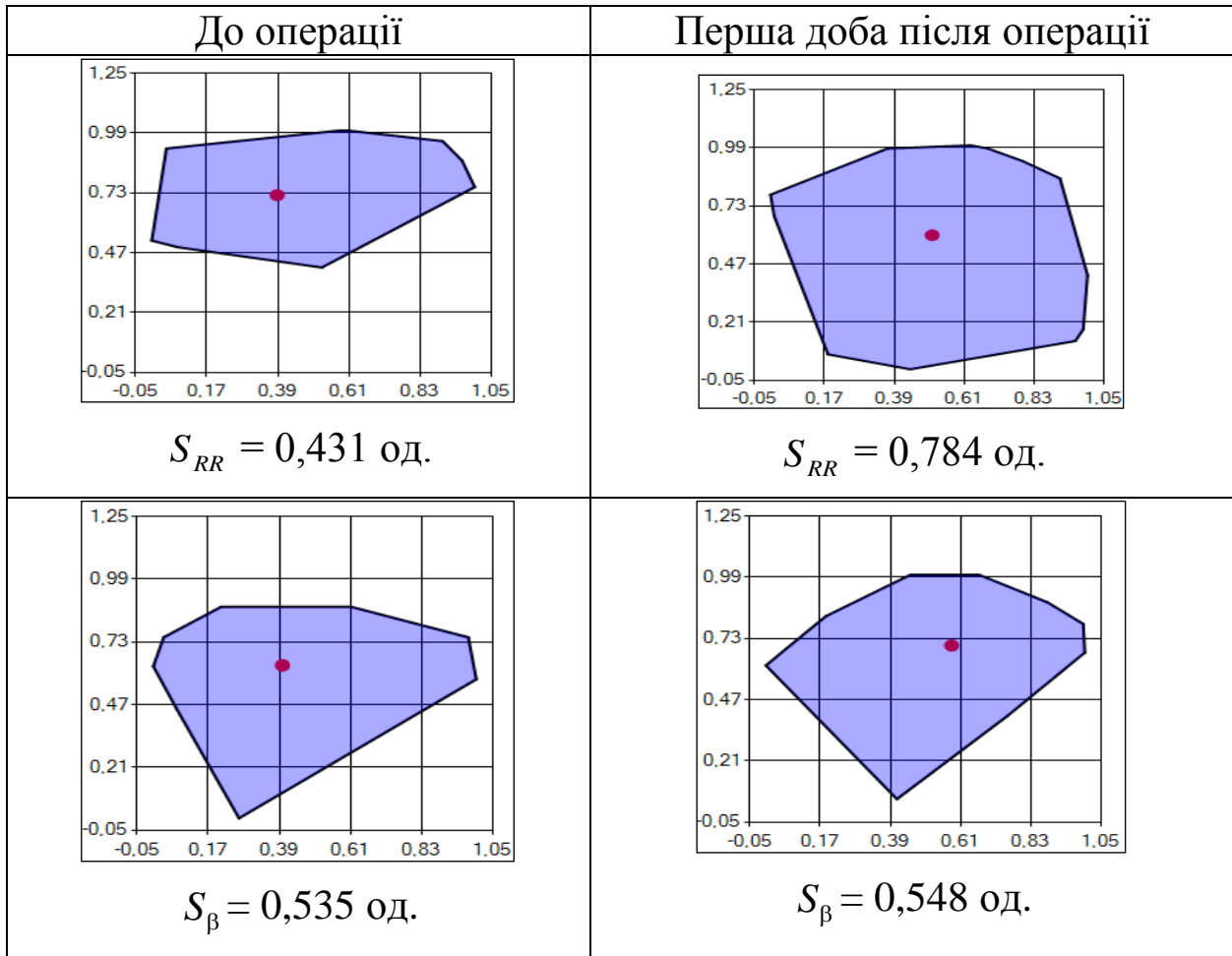
Розглянемо ще один приклад, який демонструє динаміку інтегральних показників ФПЕ за результатами оброблення ЕКГ тривалістю 150 с під час краплинного введення лікарських препаратів. У табл. 4.5 наведено результати обчислення площі S_{RR} опуклої оболонки фазового портрета ентропії (ФПЕ) послідовності RR -інтервалів, отримані під час краплинного введення хворому 72 років кардіологічних препаратів Тівомакс, Армадін і T -тріомакс.

До введення препаратів у хворого спостерігалася виражена бігемінія – регулярне чергування нормальних та екстрасистолічних серцевих циклів, що можна побачити на верхній ритмограмі. На момент часу 13:02 (через годину після початку введення препаратів) серцевий ритм нормалізувався, що відбивалося збільшенням майже втричі показника S_{RR} порівняно з початковим значенням.

Цікаві результати отримані також під час аналізу динаміки зміни показників S_{RR} до та після операції на відкритому серці – аортокоронарному шунтуванні (АКШ), які подано у табл. 4.6. Як видно з таблиці, показник S_{RR} збільшився на 82 %, а показник S_{β} – на 2,4 %.

Таблиця 4.5. Динаміка показника S_{RR} під час введення ліків

Час	Фрагменти ритмограми	ФПЕ RR -інтервалів
11:35		 $S_{RR} = 0,176$
13:02		 $S_{RR} = 0,515$

Таблиця 4.6. Динаміка показника S_{RR} і S_{β} під час АКШ

Наочність інтегральних показників S_{RR} та S_{β} їх висока чутливість до зовнішніх впливів дає підстави їх використовувати як додаткові діагностичні ознаки під час вирішення актуальних наукових та практичних задач профілактичної та клінічної медицини.

4.3. Лінгвістичний підхід до оброблення циклічних сигналів

Аналіз динаміки зміни форми циклічних сигналів можна проводити в межах лінгвістичного підходу, який ґрунтується на переході від k -ї реалізації сигналу $z_k(t)$, $k = 1, 2, \dots$, до слова $S_k = \alpha_1 \alpha_2 \dots \alpha_K$, що являє собою кінцевий ланцюжок символів $\alpha_j \in \Psi$, $j = 1, \dots, K$ деякого алфавіту Ψ «назв» фрагментів. У загальному випадку лінгвістичний підхід передбачає, що множина $\{S_k\}$ різноманітних слів (не обов'язково

кінцева) утворює формальну мову, для якої будується граматика у вигляді четвірки

$$G = \langle \Omega_0, \Omega_T, P_G, \omega_0 \rangle, \quad (4.72)$$

де Ω_0 – множина нетермінальних символів (змінних); Ω_T – множина термінальних символів (констант); $\Omega_T \cup \Omega_0 = \Psi$, $\Omega_T \cap \Omega_0 = \emptyset$, P_G – множина правил підстановки, а $\omega_0 \in \Omega_0$ – початковий (кореневий) нетермінальний символ.

Під час розв’язування практичних задач найчастіше передбачається, що алфавіт Ψ заздалегідь відомий, а конструювання граматики (4.72), адекватних множині сигналів, що спостерігаються, здійснюється людиною на підставі неформальних знань експерта в предметній області.

Будемо аналізувати параметри x_1, \dots, x_M , що характеризують форму окремих циклів $z_k(t)$, і за різницею значень m -го параметра ($m = 1, \dots, M$) на послідовних циклах обчислимо значення двозначних індикаторних функцій

$$V_n^{(m)} = \begin{cases} +1, & \text{if } x_n^{(m)} - x_{n-1}^{(m)} \geq 0, \\ -1, & \text{if } x_n^{(m)} - x_{n-1}^{(m)} < 0, \end{cases} \quad m = 1, \dots, M, \quad n = 2, \dots, N. \quad (4.73)$$

Можливі комбінації значень $V_n^{(1)}, V_n^{(2)}, \dots, V_n^{(M)}$ визначають 2^M різних символів $\alpha_n \in A$, а ланцюжок символів $S_k = \alpha_1 \alpha_2 \dots \alpha_{N-1}$ формує $N-1$ розрядне слово S_k , яке однозначно кодує оброблюваний сигнал $z_k(t)$.

Розглянемо правило кодування електрокардіограми (ЕКГ), цикли якої описують два відомі параметри – тривалість RR -інтервалу і показник β симетрії зубця T . У цьому випадку цикли кодує один з чотирьох символів алфавіту $A = \{A, B, C, D\}$ за правилом, поданим у табл. 4.7.

Перехід від спостережуваного сигналу $z_k(t)$ до кодового слова (кодограми S_k) дає змогу використовувати методи математичної лінгвістики для розв’язування задачі аналізу та інтерпретації $z_k(t)$. Зокрема метод передбачає оцінювання близькості між будь-якою парою кодограм на основі відстані Левенштейна $L(S_\mu, S_\nu)$, яка визначає мі-

німальну кількість операцій редагування (вставки, видалення та заміни символу), що забезпечує перехід від S_μ до S_ν .

Таблиця 4.7. Правило кодування ЕКГ на основі функції (4.73)

№ з/п	Значення індикаторних функцій		Символ $\alpha_n \in A$
	$V_n^{(1)}$	$V_n^{(2)}$	
	RR-інтервал	Симетрія зубця T	
1	1	1	A
2	1	-1	B
3	-1	1	C
4	-1	-1	D

Для обчислення $L(S_\mu, S_\nu)$ використаємо алгоритм Вагнера-Фішера (Wagner-Fischer), який ґрунтується на методі динамічного програмування. Сформуємо матрицю U розмірністю $N_\mu \times N_\nu$, де N_μ і N_ν відповідно кількості символів у словах S_μ і S_ν .

Заповнимо перший рядок і перший стовпець матриці U таким чином:

$$\begin{aligned} U(i, 0) &= i, \quad \forall i = 1 \dots N_\mu, \\ U(0, j) &= j, \quad \forall j = 1 \dots N_\nu. \end{aligned} \quad (4.74)$$

а інші елементи матриці U ($i > 1, j > 1$) заповнюємо відповідно до правила:

$$U(i, j) = \min \begin{cases} U(i, j-1) + 1, & U(i-1, j) + 1, \\ U(i-1, j-1) + m(S_\mu(i), S_\nu(j)), \end{cases} \quad (4.75)$$

де

$$m(S_\mu(i), S_\nu(j)) = \begin{cases} 0, & \text{если } S_\mu(i) = S_\nu(j), \\ 1, & \text{если } S_\mu(i) \neq S_\nu(j). \end{cases} \quad (4.76)$$

У результаті відстань Левенштейна $L(S_\mu, S_\nu)$ між словами S_μ і S_ν визначає елемент матриці $U(N_\mu, N_\nu)$.

У загальному випадку можливо декілька оптимальних способів переходу від комірки $U(1, 1)$ до комірки $U(N_\mu, N_\nu)$, які забезпечують мінімальну кількість операцій редагування. Це зображено на рис. 4.30, на якому показано «оптимальну» послідовність переходу від слова

$$S_\mu = DDABDCBADCBADCA \quad (4.77)$$

до слова

$$S_\nu = BACDAAACDADCCBB. \quad (4.78)$$

		A	D	A	B	D	C	B	A	D	C	B	A	D	C	A
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
B	1	①	②	3	3	4	5	6	7	8	9	10	11	12	13	14
A	2	2	2	②	3	4	5	6	6	7	8	9	10	11	12	13
C	3	3	3	3	③	4	4	5	6	7	7	8	9	10	11	12
D	4	3	3	4	4	③	④	5	6	6	7	8	9	9	10	11
A	5	4	4	3	4	4	④	⑤	5	6	7	8	8	9	10	10
A	6	5	5	4	4	5	5	⑤	⑤	6	7	8	8	9	10	10
A	7	6	6	5	5	5	6	6	⑤	⑥	7	8	8	9	10	10
C	8	7	7	6	6	6	5	6	6	6	⑥	7	8	9	9	10
D	9	8	7	7	7	6	6	6	7	6	7	⑦	8	8	9	10
A	10	9	8	7	8	7	7	7	6	7	7	8	⑦	8	9	9
D	11	10	9	8	8	8	8	8	7	6	7	8	8	⑦	8	9
C	12	11	10	9	9	9	8	9	8	7	6	7	8	⑧	⑦	8
C	13	12	11	10	10	10	9	9	9	8	7	7	8	9	⑧	⑧
B	14	13	12	11	10	11	10	9	10	9	8	7	8	9	⑨	⑨
B	15	14	13	12	11	11	11	10	10	10	9	8	8	9	10	⑩

Рис. 4.30. Можливі способи переходу від слова S_μ до слова S_ν .

Стали популярними інші варіанти редакторської відстані, зокрема, відстань Дамерау–Левенштейна (**Damerau–Levenshtein distance**), в яку введено додаткову операцію редагування у вигляді транспозиції (перестановки) двох сусідніх символів. Відомо також модифікації відстані Левенштейна за рахунок введення різних цін елементарних операцій.

На основі обчислення відстані Левенштейна $L(S_\mu, S_\nu)$ між парами кодограм S_μ, S_ν можна побудувати алгоритми, що забезпечують класифікацію циклічних сигналів. Для цього формують навчальну вибірку сигналів з відомою приналежністю до класів Ψ_1, \dots, Ψ_G .

Нехай у результаті експериментів зареєстровано Q_g спостережень класу $V_g \in \{V_1, \dots, V_G\}$, які закодовані словами $S_1^{(g)}, S_2^{(g)}, \dots, S_{Q_g}^{(g)}$ відповідно до табл. 4.7. Обчислимо за формулами (4.74)–(4.76) відстані Левенштейна $L(S_\mu^{(g)}, S_\nu^{(g)})$ між кожною парою $S_\mu^{(g)}, S_\nu^{(g)}$, $\mu = 1, \dots, Q_g$, $\nu = 1, \dots, Q_g$ кодограм та сформуємо квадратну матрицю:

$$\Lambda^{(g)} = \begin{pmatrix} L(S_1^{(g)}, S_1^{(g)}) & L(S_1^{(g)}, S_2^{(g)}) & \dots & L(S_1^{(g)}, S_{Q_g}^{(g)}) \\ L(S_2^{(g)}, S_1^{(g)}) & L(S_2^{(g)}, S_2^{(g)}) & \dots & L(S_2^{(g)}, S_{Q_g}^{(g)}) \\ \dots & \dots & \dots & \dots \\ L(S_{Q_g}^{(g)}, S_1^{(g)}) & L(S_{Q_g}^{(g)}, S_2^{(g)}) & \dots & L(S_{Q_g}^{(g)}, S_{Q_g}^{(g)}) \end{pmatrix}. \quad (4.79)$$

Еталон класу визначить рядок матриці (4.79), сума елементів якої є мінімальною, тобто

$$S_0^{(g)} = \arg \min_{1 \leq \nu \leq Q_g} \sum_{\mu=1}^{Q_g} L(S_\mu^{(g)}, S_\nu^{(g)}). \quad (4.80)$$

Аналогічно визначимо еталони інших класів.

Побудовані еталони дають змогу приймати рішення про поточну кодограму S_t сигналу, що аналізується, за правилом:

$$\text{Клас } \Psi_n, \text{ якщо } L(S_t, S_0^{(n)}) = \min_{1 \leq g \leq G} L(S_t, S_0^{(g)}), \quad n = 1, \dots, G. \quad (4.81)$$

Для оцінювання ефективності правила (4.81) проведено дослідження з використанням бази даних, яка включала 100 записів ЕКГ верифікованих хворих на хронічну форму ішемічної хвороби серця (CAD) та 100 записів ЕКГ здорових добровольців (контрольна група HEALTHY). Діагноз було попередньо встановлено за результатами коронарографії.

Важливо зазначити, що на ЕКГ хворих були відсутні традиційні ознаки ішемії міокарда. Однак, навіть на такому складному клінічному матеріалі запропонований підхід дав змогу класифікувати наявні дані з чутливістю $S_E = 72\%$ та специфічністю $C_p = 79\%$ за правилом

$$\text{CAD, якщо } L(S_t, S_0^{(1)}) \leq L(S_t, S_0^{(2)}), \quad (4.82)$$

$$\text{HEALTHY, якщо } L(S_t, S_0^{(1)}) > L(S_t, S_0^{(2)}), \quad (4.83)$$

де S_t – кодограма аналізованої ЕКГ, а $S_0^{(1)}$ та $S_0^{(2)}$ відповідно еталони кодограм хворих на ІХС та здорових пацієнтів.

На рис. 4.31 зображено оцінювання умовних розподілів $P(L(S_t, S_0^{(1)}))$ та $P(L(S_t, S_0^{(2)}))$ відстаней Левенштейна між кодовими словами навчальної вибірки відносно еталонів $S_0^{(1)}$ та $S_0^{(2)}$. Перевірка гіпотези про однорідність умовних розподілів $P(L(S_t, S_0^{(1)}))$ і $P(L(S_t, S_0^{(2)}))$ за критерієм Колмогорова–Смирнова показала, що з високою статистичною значимістю ($p < 0,001$) гіпотеза про рівність розподілів $P(L(S_t, S_0^{(1)}))$ і $P(L(S_t, S_0^{(2)}))$ має бути відкинута. Аналогічно це було підтверджено критерієм Мана–Уїтні для незалежних вибірок.

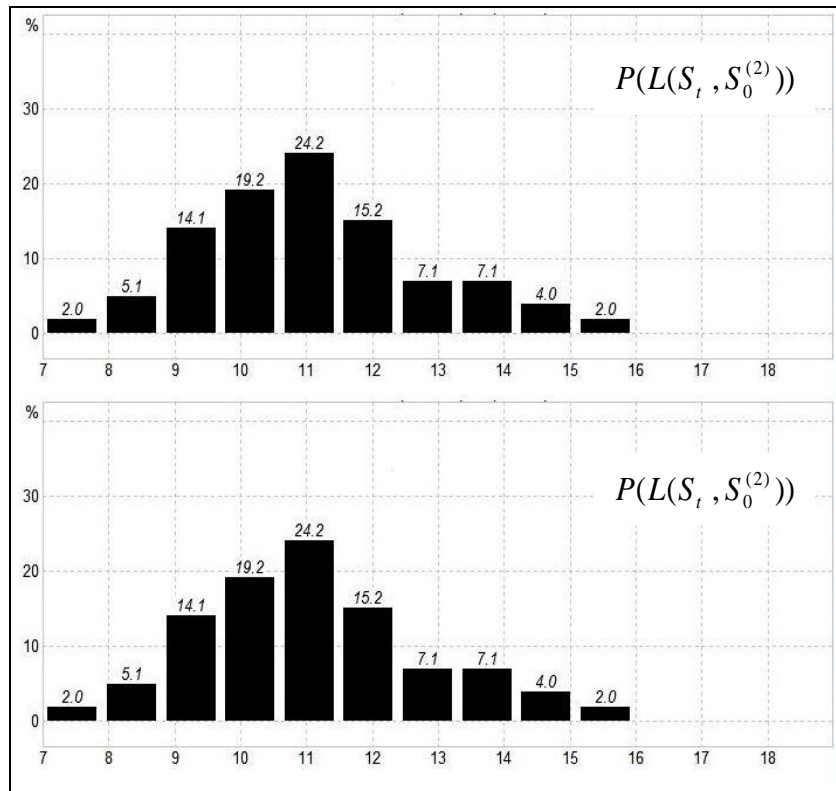


Рис. 4.31. Оцінювання умовних розподілів $P(L(S_t, S_0^{(1)}))$ та $P(L(S_t, S_0^{(2)}))$

Процедури (4.74)–(4.76) дають змогу обчислити відстань Левенштейна $L(S_\mu, S_\nu)$ у загальному випадку, коли сигнали $z_\mu(t)$, $z_\nu(t)$ мають різну кількість циклів, тобто коли числа символів N_μ та N_ν у словах S_μ і S_ν неоднакові. Вочевидь, якщо $N_\mu \neq N_\nu$, то з відстань

Левенштейна задовольняє умову $L(S_\mu, S_\nu) \geq |N_\mu - N_\nu|$, причому $L(S_\mu, S_\nu) = |N_\mu - N_\nu|$, навіть якщо сигнали тотожні $z_\mu(t) \equiv z_\nu(t)$.

Звідси випливає, що коли $N_\mu \neq N_\nu$, відстань Левенштейна $L(S_\mu, S_\nu)$ характеризує не тільки відмінність форми $z_\mu(t)$ і $z_\nu(t)$, а й відмінність тривалості сигналів, що додає неоднозначність в інтерпретацію $L(S_\mu, S_\nu)$. Тому перед формуванням матриці (4.79) треба вирівняти тривалості спостережень, забезпечивши однакову кількість $N_0 = \min_k N_k$ циклів усіх спостережень навчальної вибірки.

Розглянемо можливість покращення оцінювання динаміки форми циклів за рахунок переходу від двозначних індикаторних функцій (4.73) до тризначних:

$$V_n^{(m)} = \begin{cases} +1, & \text{if } x_n^{(m)} - x_{n-1}^{(m)} > \varepsilon, \\ 0, & \text{if } |x_n^{(m)} - x_{n-1}^{(m)}| \leq \varepsilon, \\ -1, & \text{if } x_n^{(m)} - x_{n-1}^{(m)} < -\varepsilon, \end{cases} \quad m = 1, \dots, M, \quad n = 2, \dots, N_0, \quad (4.84)$$

де ε – поріг нечутливості до зміни m -го параметра ($m = 1, \dots, M$) на послідовних циклах.

На відміну від (4.73) змінні (4.84) оцінюють не тільки напрямок, а й ступінь зміни значень m -го параметра, що розширює можливості методу. В такому випадку алфавіт A містить вже 3^M різних символів, за допомогою яких породжуються $N_0 - 1$ розрядні кодограми $S_k = \alpha_1 \alpha_2 \cdots \alpha_{N_0-1}$ сигналу $z_k(t)$.

У табл. 4.8. наведено правило кодування циклів ЕКГ на основі значень індикаторних функцій $V_n^{(1)}$, $V_n^{(2)}$, $V_n^{(3)}$, які характеризують динаміку зміни трьох діагностичних показників – тривалостей RR -інтервалів, симетрій зубців T і амплітуд зубців R . У цьому випадку алфавіт містить 27 символів.

На рис. 4.32 показано фрагмент реальної ЕКГ та графіки змін тривалостей RR -інтервалів, симетрій зубців T та амплітуд зубців R .

Таблиця 4.8. Правило кодування ЕКГ на основі функції (4.84)

№ з/п	Значення індикаторних функцій			Символ $\alpha_n \in A$
	$V_n^{(1)}$	$V_n^{(2)}$	$V_n^{(3)}$	
	<i>RR</i> -інтервали	Симетрія зубця <i>T</i>	Амплітуда зубця <i>R</i>	
0	0	0	0	=
1	0	-1	-1	A
2	0	+1	+1	B
3	-1	0	0	C
4	-1	-1	-1	D
5	-1	+1	+1	E
6	+1	0	0	F
7	+1	-1	-1	G
8	+1	+1	+1	H
9	0	0	0	I
10	0	-1	-1	J
11	0	+1	+1	K
12	-1	0	0	L
13	-1	-1	-1	M
14	-1	+1	+1	N
15	+1	0	0	O
16	+1	-1	-1	P
17	+1	+1	+1	Q
18	0	0	0	R
19	0	-1	-1	S
20	0	+1	+1	T
21	-1	0	0	U
22	-1	-1	-1	V
23	-1	+1	+1	W
24	+1	0	0	X
25	+1	-1	-1	Y
26	+1	+1	+1	Z

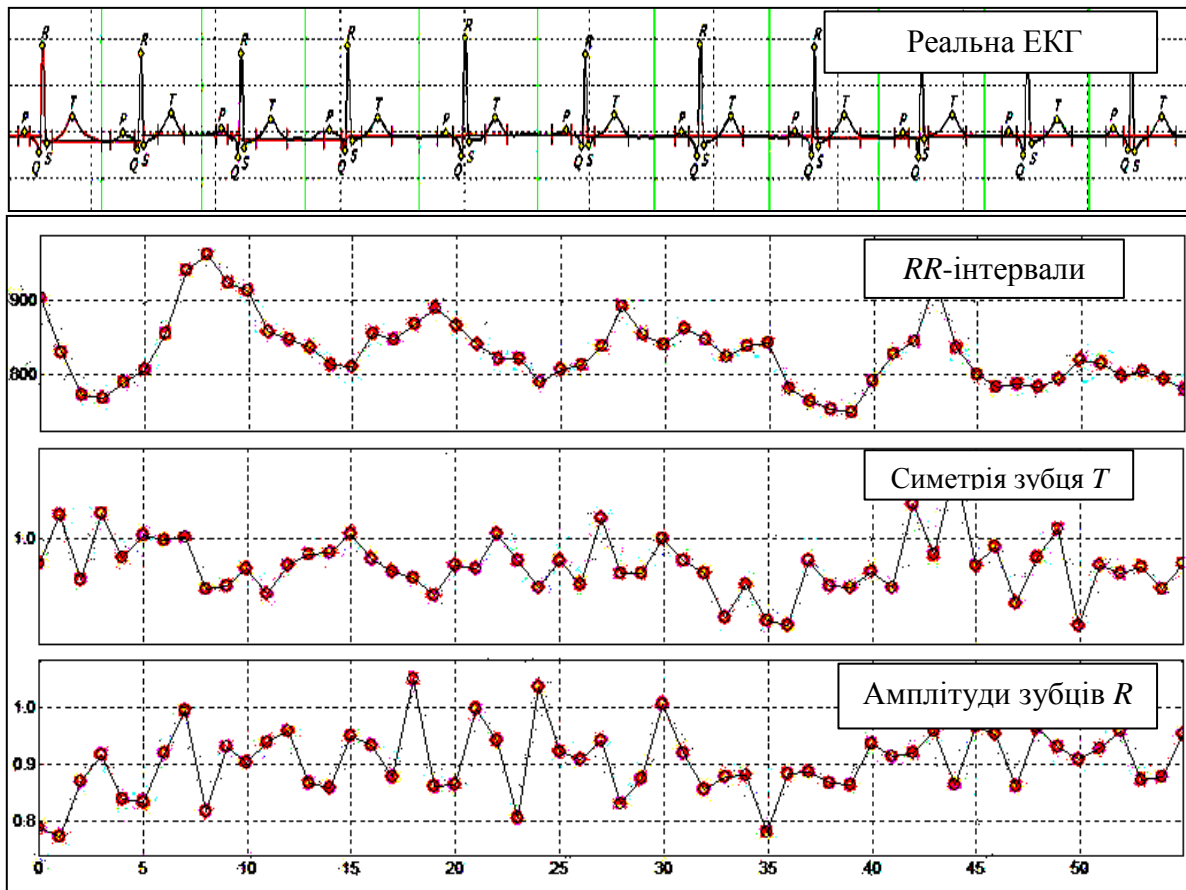


Рис. 4.32. Графіки змін тривалостей RR -інтервалів, симетрій зубців T та амплітуд зубців R на реальній ЕКГ

Як видно з табл. 4.9, на якій подано кодограми одного і того самого сигналу за різними значеннями порогу ε , зі збільшенням ε зменшується кількість різних символів $\alpha_n \in A$, що формують кодограми, і одночасно збільшується кількість символів «=», які кодують стабільні ділянки сигналу. Тому виникає завдання вибору прийняттого значення порогу ε , що забезпечує контроль лише таких змін сигналу, які мають діагностичну цінність.

Для побудови бінарного класифікатора, коли сигнали належать до одного з двох класів Ψ_1 або Ψ_2 , оптимальне значення порогу ε можна визначити таким чином.

Маємо навчальну вибірку Q_1 спостережень класу Ψ_1 і Q_2 спостережень класу Ψ_2 . Кодуватимемо спостереження класу Ψ_1 відповідно до (4.88) за різними фіксованими значеннями ε , заданих з певним кроком $\Delta\varepsilon$ в інтервалі $0 \leq \varepsilon \leq \varepsilon_{\max}$. В результаті сформуємо $\frac{\varepsilon_{\max}}{\Delta\varepsilon}$ мат-

риць внутрішньокласових відстаней Левенштейна $L_\varepsilon(S_\mu^{(1)}, S_\nu^{(1)})$ за різними дискретними значеннями ε для спостережень першого класу:

$$\Lambda_\varepsilon^{(1)} = \begin{pmatrix} L_\varepsilon(S_1^{(1)}, S_1^{(1)}) & L_\varepsilon(S_1^{(1)}, S_2^{(1)}) & \dots & L_\varepsilon(S_1^{(1)}, S_{Q_1}^{(1)}) \\ L_\varepsilon(S_2^{(1)}, S_1^{(1)}) & L_\varepsilon(S_2^{(1)}, S_2^{(1)}) & \dots & L_\varepsilon(S_2^{(1)}, S_{Q_1}^{(1)}) \\ \dots & \dots & \dots & \dots \\ L_\varepsilon(S_{Q_1}^{(1)}, S_1^{(1)}) & L_\varepsilon(S_{Q_1}^{(1)}, S_2^{(1)}) & \dots & L_\varepsilon(S_{Q_1}^{(1)}, S_{Q_1}^{(1)}) \end{pmatrix}, \varepsilon = 0, \dots, \varepsilon_{\max}. \quad (4.85)$$

Таблиця 4.9. Приклад кодування однієї і тієї самої ЕКГ за різними значеннями порога ε

Поріг ε	Кодограма
1	QSMWNVMZPQYOQ=MXWVZQYKZVKWMZGMZZVIZYPTDMWMZQYJTMNTPSQY
2	QSMWNVDZAQSIQ=MRWVZQGKZVKWMZAMTZSIZGITCMWMZQYITJNTPSKS
3	QSJRNDAZAQSIQ=MRTSTOAKZSKWLZAJTTRIZAITCMTLXQSITJNTJSKR
4	QSJRNDAZAOSIK=MRTATIAKTSKTLZAJTTRIZAITCMTLXQSITJKTJ=BR
5	QSJRNDATAORIK=MRTATIAKTSKRLT=JTTRIZAIT=JRLXKSITJITJ=BR
6	OSJRKD=TAORIB=JRTATIAKTSKRLT=JTTRIZAIT=JRLXISITJIRI=BR
7	ISJRKD=TAIRIB=JRBATIAITSKRIT=JTBRIZAIT=JRLXISITJIRI=BR
8	ISIRKD=TA=RIB=JR=ATIAITSKRIT=JT=RITAIR=A=LXISITJIRI=BR
9	IRIRID=TA=RIB=JR=ABIAITSKRIT=JT=RITAIR=A=LXIR=TJIRI=BR
10	IRIRID=TA=RI===R=ABIAITSKRIT=JR=RITAIR=A=IXIR=TJIRI==R

За елементами кожної з матриць (4.85) обчислимо середню внутрішньокласову відстань для першого класу, яка залежить від ε :

$$\bar{L}_\varepsilon^{(1)} = \frac{2}{Q_1(Q_1 - 1)} \sum_{v=1}^{Q_1} \sum_{\mu=1}^{Q_1} L_\varepsilon(S_\mu^{(1)}, S_\nu^{(1)}). \quad (4.86)$$

Аналогічно за елементами матриць $\Lambda_\varepsilon^{(2)}$ обчислимо середню внутрішньокласову відстань для другого класу, яка також залежить від ε :

$$\bar{L}_\varepsilon^{(2)} = \frac{2}{Q_2(Q_2 - 1)} \sum_{v=1}^{Q_2} \sum_{\mu=1}^{Q_2} L_\varepsilon(S_\mu^{(2)}, S_\nu^{(2)}). \quad (4.87)$$

Сформуємо тепер $Q_1 \times Q_2$ матриці міжкласових відстаней $L_\varepsilon(S_\mu^{(1)}, S_\nu^{(2)})$ між усіма парами кодограм класів Ψ_1 і Ψ_2 за фіксованими значеннями ε в інтервалі $0 \leq \varepsilon \leq \varepsilon_{\max}$:

$$\Lambda_\varepsilon^{(1,2)} = \begin{pmatrix} L_\varepsilon(S_1^{(1)}, S_1^{(2)}) & L_\varepsilon(S_1^{(1)}, S_2^{(2)}) & \dots & L_\varepsilon(S_1^{(1)}, S_{Q_2}^{(2)}) \\ L_\varepsilon(S_2^{(1)}, S_1^{(2)}) & L_\varepsilon(S_2^{(1)}, S_2^{(2)}) & \dots & L_\varepsilon(S_2^{(1)}, S_{Q_2}^{(2)}) \\ \dots & \dots & \dots & \dots \\ L_\varepsilon(S_{Q_1}^{(1)}, S_1^{(2)}) & L_\varepsilon(S_{Q_1}^{(1)}, S_2^{(2)}) & \dots & L_\varepsilon(S_{Q_1}^{(1)}, S_{Q_2}^{(2)}) \end{pmatrix}, \varepsilon = 0, \dots, \varepsilon_{\max}. \quad (4.88)$$

За елементами матриць (4.88) обчислимо середню міжкласову відстань, яка залежить від ε :

$$\bar{L}_\varepsilon^{(1,2)} = \frac{1}{Q_1 Q_2} \sum_{\rho=1}^{Q_2} \sum_{\mu=1}^{Q_1} L_\varepsilon(S_\mu^{(1)}, S_\rho^{(2)}), \quad (4.89)$$

Використовуючи величини (4.86), (4.87) і (4.89) сформуємо критерій оптимальності

$$\eta(\varepsilon) = \frac{\bar{L}_\varepsilon^{(1)} + \bar{L}_\varepsilon^{(2)}}{\bar{L}_\varepsilon^{(1,2)}}. \quad (4.90)$$

Критерій (4.90) дає змогу методом перебору дискретних значень ε , заданих з кроком $\Delta\varepsilon$ в інтервалі $0 \leq \varepsilon \leq \varepsilon_{\max}$, визначити оптимальне значення порогу:

$$\varepsilon_0 = \arg \min_{0 \leq \varepsilon \leq \varepsilon_{\max}} \eta(\varepsilon), \quad (4.91)$$

Оскільки $\bar{L}_\varepsilon^{(1)} > 0, \bar{L}_\varepsilon^{(2)} > 0, \bar{L}_\varepsilon^{(1,2)} > 0$ оптимізаційна процедура (4.91) забезпечує мінімальні середні внутрішньокласові відстані $\bar{L}^{(1)}$, $\bar{L}^{(2)}$ і одночасно максимальну середню міжкласову відстань $\bar{L}_\varepsilon^{(1,2)}$.

Достовірність класифікації циклічних сигналів можна підвищити, якщо додатково аналізувати окремі частини кодового слова у вигляді характерних патернів, наприклад, трисимвольних патернів

$$\pi = \lambda \rho \vartheta, \quad \lambda, \rho, \vartheta \in A. \quad (4.92)$$

Для цього обчислимо за навчальною вибіркою середню частоту виникнення патернів (4.92), характерних для кожного класу $\Psi_g \in \{\Psi_1, \dots, \Psi_G\}$:

$$\hat{P}^{(g)}(\pi_l) = \frac{1}{Q_g} \sum_{\mu=1}^{Q_g} \frac{W_{\mu}^{(g)}(\pi_l)}{N_0 - 2}, \quad l = 1, \dots, L, \quad (4.93)$$

де $W_{\mu}^{(g)}(\pi_l)$ кількість входжень l -го трисимвольного патерну π_l в μ -у кодограму класу $V_g \in \{V_1, \dots, V_G\}$; L – кількість варіантів трисимвольних патернів $\pi = \lambda\rho\vartheta$ з елементів алфавіту A ; Q_g – кількість спостережень g -го класу в навчальній вибірці.

Для прискорення процедури (4.93) доцільно використовувати алгоритм пошуку рядків Боєра–Мура (**Boyer-Moore**).

Позначимо $\pi_0^{(g)}$ – домінуючий патерн g -го класу, що задовольняє умову

$$\pi_0^{(g)} = \arg \max_{1 \leq l \leq L} \hat{P}^{(g)}(\pi_l), \quad g = 1, \dots, G. \quad (4.94)$$

Якщо середні частоти $\hat{P}^{(1)}(\pi_0^{(1)})$, $\hat{P}^{(2)}(\pi_0^{(2)})$, ..., $\hat{P}^{(G)}(\pi_0^{(G)})$ мають статистично значимі відмінності, а самі домінуючі патерни $\pi_0^{(1)}$, $\pi_0^{(2)}$, ..., $\pi_0^{(G)}$, визначені за (4.94), неоднакові, то їх можна використовувати як еталони класів у такому вирішальному правилі:

$$\text{Клас } \Psi_n, \text{ якщо } L(\pi_t, \pi_0^{(n)}) = \arg \min_{1 \leq g \leq G} L(\pi_t, \pi_0^{(g)}), \quad n \in [1, G], \quad (4.95)$$

Продемонструємо правило (4.95) на прикладі оброблення ЕКГ, які кодувались з використанням двозначних індикаторних функцій (4.73). За навчальною вибіркою кодограм встановлено (табл. 4.10), що з високою статистичною значущістю ($p < 0,01$) патерн $\pi_0^{(1)} = DAD$ домінує на кодограмах хворих на САД (клас V_1), а патерн $\pi_0^{(2)} = CAA$ домінує на кодограмах здорових пацієнтів (клас V_2).

На рис. 4.33 показано приклади реальних ЕКГ, кодограми яких містять патерни $\pi_0^{(1)} = DAD$ та $\pi_0^{(2)} = CAA$. Незважаючи на те, що зазначені фрагменти сигналу візуально майже невиразні, запропоновані

комп'ютерні процедури забезпечать, що такі фрагменти належатимуть до патерну $\pi_0^{(1)} = DAD$ або $\pi_0^{(2)} = CAA$.

На основі цих досліджень вирішальні правила (4.82), (4.83) розширені таким чином:

$$\begin{aligned} \text{CAD, якщо } L(S_t, S_0^{(1)}) \leq L(S_t, S_0^{(2)}) \wedge G_t(DAD) \geq G_t(CAA), \\ \text{Healthy, якщо } L(S_t, S_0^{(1)}) > L(S_t, S_0^{(2)}) \wedge G_t(DAD) < G_t(CAA), \end{aligned} \quad (4.96)$$

не визначено в інших випадках.

Таблиця 4.10. Оцінювання частот виникнення трисимвольних патернів на кодограмах реальних ЕКГ

Патерн	$\hat{P}^{(Healthy)}(\pi_3)$	$\hat{P}^{(CAD)}(\pi_3)$	p -value
<i>ADA</i>	0,009104	0,029465	0,000006
<i>ADD</i>	0,031545	0,025732	0,017603
<i>BAА</i>	0,007682	0,003707	0,045142
<i>BAD</i>	0,023236	0,012264	0,00159
<i>BCA</i>	0,013526	0,02479	0,003743
<i>BDA</i>	0,021176	0,042309	0,000412
<i>BDC</i>	0,034448	0,025401	0,003428
<i>CAA</i>	0,01891	0,007792	0,002103
<i>CAB</i>	0,034292	0,026427	0,015428
<i>CAD</i>	0,025595	0,051444	0,00015
<i>CBA</i>	0,023889	0,011414	0,000466
<i>CBB</i>	0,012348	0,006269	0,021577
<i>CBD</i>	0,018828	0,028515	0,027272
<i>CDC</i>	0,019029	0,010455	0,009801
<i>DAB</i>	0,028581	0,052221	0,00127
<i>DAD</i>	0,01142	0,031005	0,000021
<i>DBA</i>	0,009673	0,007151	0,11723
<i>DCA</i>	0,062054	0,052256	0,034411

На основі цих досліджень розширені вирішальні правила (4.82), (4.83) матимуть вигляд:

CAD, якщо $L(S_t, S_0^{(1)}) \leq L(S_t, S_0^{(2)}) \wedge G_t(DAD) \geq G_t(CAA)$,
 Healthy, якщо $L(S_t, S_0^{(1)}) > L(S_t, S_0^{(2)}) \wedge G_t(DAD) < G_t(CAA)$, (4.97)
 не визначено в інших випадках.

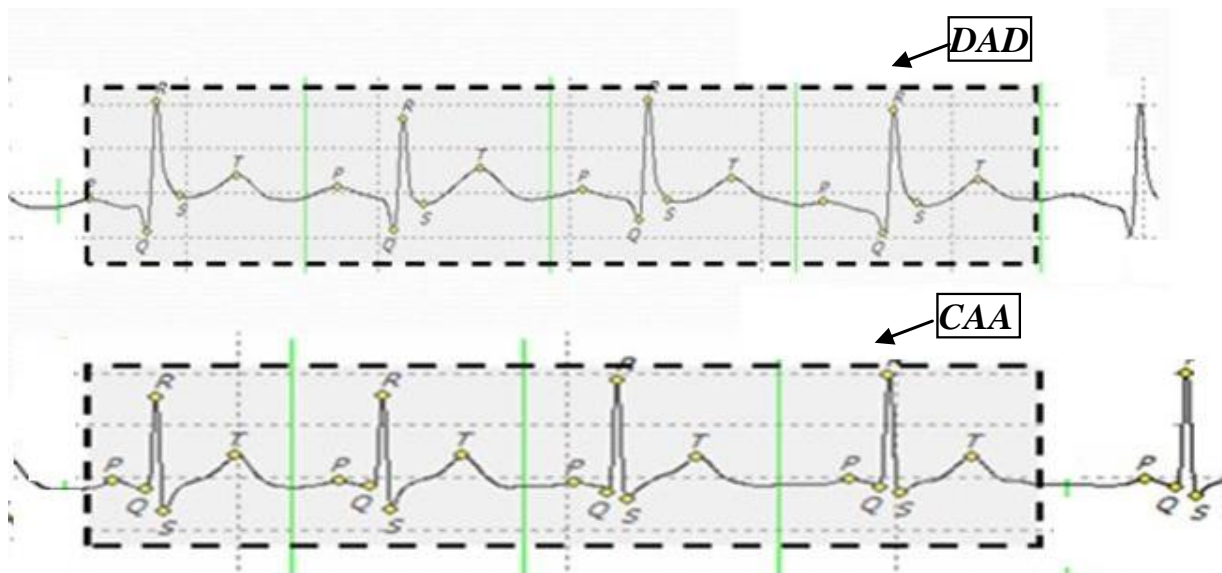


Рис. 4.33. Патерни $\pi_0^{(1)} = DAD$ та $\pi_0^{(2)} = CAA$ на реальних ЕКГ

Під час скринінгових обстежень правило (4.97) дає змогу виявляти пацієнтів із підвищеним ризиком ішемічної хвороби серця: встановлено, що у 87,5 % випадків за правилом (4.97) приймаються однозначні рішення з чутливістю $S_E = 74,7 \%$ та специфічністю $S_p = 79,5 \%$ навіть якщо на ЕКГ відсутні загальноприйняті ознаки ішемії міокарда.

Завдання до комп'ютерних практикумів

Завдання 4.1. Необхідно програмно реалізувати два методи відображення циклічного сигналу на фазовій площині. Основні задачі роботи:

- опанувати процедуру введення інформації з текстових файлів;
- опанувати технологію переходу від скалярного сигналу в часовій області до фазового портрета;
- реалізувати процедуру відображення фазового портрета сигналу $z(t)$ на площині $z(t)$, $z(t - \tau)$;

– реалізувати процедуру відображення фазового портрета сигналу $z(t)$ на площині $z(t)$, dz/dt .

Під час виконання завдання необхідно:

4.1.1. Реалізувати процедуру вводу інформації з файлу ECG.txt. Файл містить дискретні значення ЕКГ, які записані в текстовому вигляді (послідовності чисел) з кроком квантування 500 Гц. Дані розміщено в 7 колонках (зліва направо та зверху донизу). Фрагмент запису у файлі наведено на рис. 4.34.

-88.2811	-57.4347	-50.6638	-33.7054	-29.5587	8.8159	27.5191
97.7407	204.6043	328.1549	465.1085	607.7920	663.4681	666.0235
665.3907	607.1252	513.5504	400.2387	273.6923	141.8301	81.0318
68.1236	30.6428	9.7827	-1.9283	-47.5481	-67.3076	-137.0902
-242.4298	-361.9546	-487.8819	-608.8698	-650.7715	-646.7040	-642.9728
-585.2411	-511.7623	-500.2006	-478.8507	-475.1605	-529.7868	-626.6037
-757.4576	-910.7605	-975.2361	-982.6717	-984.1927	-925.9165	-832.9079
-724.6821	-628.0904	-596.3602	-618.4246	-699.0645	-831.8898	-992.7768
-1154.8394	-1286.8666	-1328.2320	-1297.4200	-1193.1141	-1026.3889	-824.9743
-620.5918	-424.5958	-221.5490	-153.6271	-15.0962	12.7620	225.1145
604.5856	1203.0405	2044.1915	3113.7069	4355.5178	5675.4779	6952.1522
8053.1227	8854.1046	9220.1008	9207.6113	8698.4536	7775.1537	6526.4844
5071.3789	3541.4409	2062.6328	739.2610	-357.2068	-1192.9945	-1770.3062

Рис. 4.34. Фрагмент запису у файлі ECG.txt

4.1.2. Оцінити першу похідну dz/dt масиву даних згідно зі схемою, наведеною на рис. 4.4.

4.1.3. Відобразити фазовий портрет на площині $z(t)$, dz/dt .

4.1.4. Відобразити псевдофазовий портрет на площині $z(t)$, $z(t - \tau)$, де τ – затримка в часі (натуральне число точок), яке задається користувачем (за замовчуванням $\tau = 8$ точок).

Робоче вікно програми має виглядати так, як показано на рис. 4.35. Під час зміни значення параметра τ , на екрані мають відображатися відповідні зміни псевдофазового портрета на площині $z(t)$, $z(t - \tau)$.

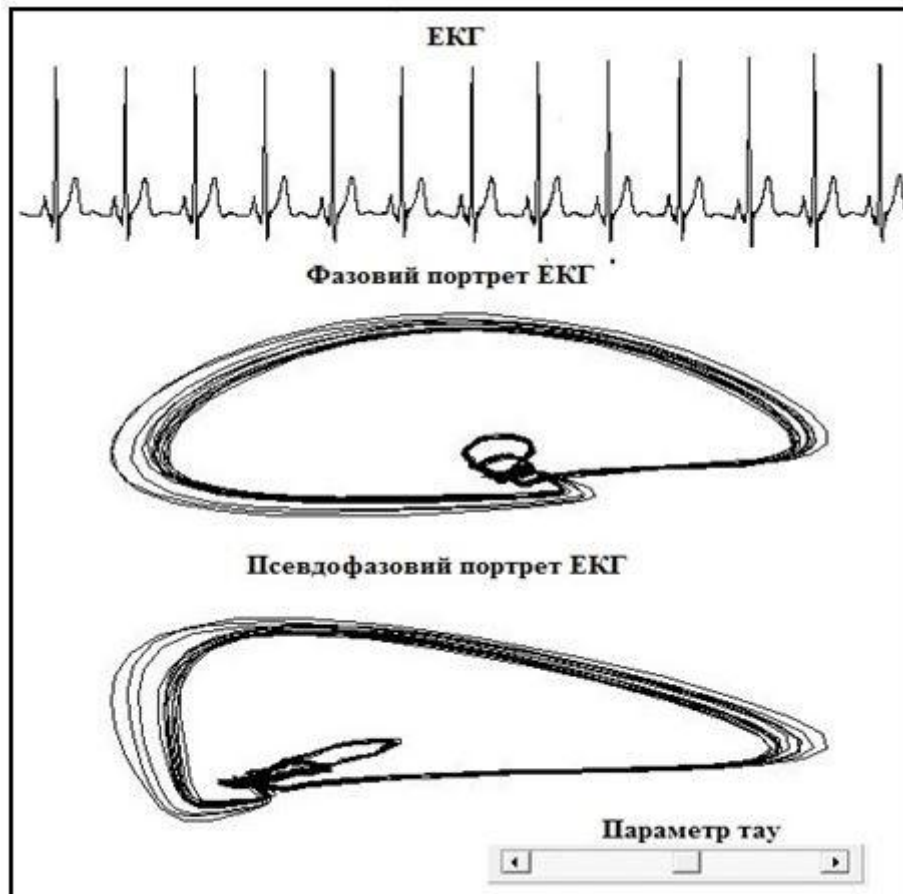


Рис. 4.35. Результати оброблення даних з файлу ECG.txt

Завдання 4.2. Необхідно програмно реалізувати процедуру визначення опорного циклу електрокардіограми за допомогою гаусдорфової відстані. Основні задачі роботи:

- побудувати цикли електрокардіограми, використовуючи числові дані з текстових файлів;
- навчитись робити нормування даних;
- реалізувати процедуру розрахунку відстані Гаусдорфа між нормованими фазовими траєкторіями.

Під час виконання завдання необхідно:

4.2.1. Реалізувати процедуру введення інформації з текстових файлів: 1.txt, 2.txt, 3.txt та 4.txt. Файли містять дискретні відліки $z[k]$ чотирьох циклів реальної ЕКГ. Дані розміщені в 7 колонках (зліва на право, та зверху до низу).

4.2.2. Оцінити першу похідну dz/dt масиву даних згідно зі схемою, наведеною на рис. 4.4.

4.2.3. Для кожного m -циклу, $m=1,2,3,4$, провести нормування даних за формулами:

$$z_m^*[k] = \frac{z_m[k] - \min_{1 \leq k \leq K} z_m[k]}{\max_{1 \leq k \leq K} z_m[k] - \min_{1 \leq k \leq K} z_m[k]},$$

$$\dot{z}_m^*[k] = \frac{\dot{z}_m[k] - \min_{1 \leq k \leq K} \dot{z}_m[k]}{\max_{1 \leq k \leq K} \dot{z}_m[k] - \min_{1 \leq k \leq K} \dot{z}_m[k]}.$$

4.2.2. Відобразити на екрані монітору чотири фазових портрети в координатах $z_m^*[k] \in [0,1]$, $\dot{z}_m^*[k] \in [0,1]$ так, як показано на рис. 4.36.

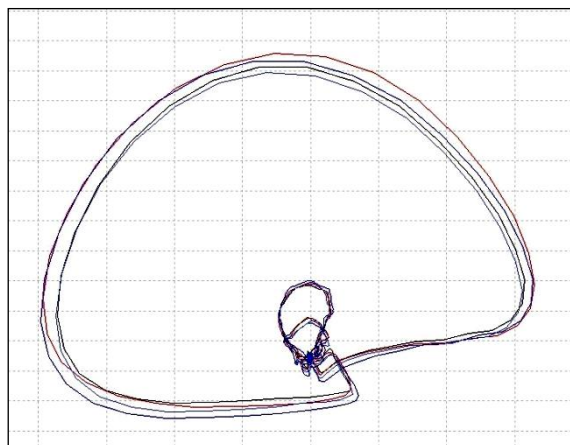


Рис. 4.36. Фазовий портрет ЕКГ з чотирьох циклів

4.2.3. Побудувати матрицю гаусдорфових відстаней між усіма парами чотирьох циклів

$$R = \begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ R_{21} & R_{22} & R_{23} & R_{24} \\ R_{31} & R_{32} & R_{33} & R_{34} \\ R_{41} & R_{42} & R_{43} & R_{44} \end{pmatrix},$$

де

$$R_{ij} = \max \left\{ \max_{q_j \in Q_j} \min_{q_i \in Q_i} \rho(q_i, q_j), \max_{q_i \in Q_i} \min_{q_j \in Q_j} \rho(q_i, q_j) \right\},$$

$q=(z^*[k], \dot{z}^*[k])$ – нормовані фазові координати, а $\rho(q_i, q_j) = \|q_i - q_j\|$ є евклідовою відстанню між векторами $q_i \in Q_i$ та $q_j \in Q_j$.

4.2.4. Знайти рядок матриці R , який визначає домінують траєкторію:

$$Q_0 = \arg \min_{1 \leq j \leq 4} \sum_{i=1}^4 R(Q_i, Q_j)$$

4.2.5. Позначити на рисунку відповідну фазову траєкторію, наприклад, лінією подвійної товщини.

Запитання для самоконтролю

1. Опишіть загальну схему оброблення циклічних сигналів.
2. Охарактеризуйте основну проблему усереднення циклів у часовій області.
3. Що таке фазові координати, фазова траєкторія та фазовий портрет.
4. Як будують фазовий портрет циклічного сигналу?
5. Як будують псевдофазовий портрет циклічного сигналу?
6. У чому полягає метод відновлення еталону у фазовому просторі?
7. У чому полягає проблема оцінювання похідної спотвореного сигналу?
8. Яким чином можна обчислити похідну сигналу?
9. Назвіть основні властивості метрики.
10. Охарактеризуйте метрику Гаусдорфа.
11. Назвіть відомі ентропійні оцінювання хаотичності сигналів.
12. Що таке переставна ентропія?
13. Як будується фазовий портрет ковзної ентропії?
14. Охарактеризуйте лінгвістичний підхід до аналізу та інтерпретації циклічних сигналів.
15. Що визначає відстань Левенштейна?
16. Як будують еталони класів згідно з лінгвістичним підходом?

РОЗДІЛ 5

БІНАРНІ КЛАСИФІКАТОРИ В ЗАДАЧАХ МЕДИЧНОЇ ДІАГНОСТИКИ

5.1. Формальні стратегії прийняття рішень

Можливість класифікувати об'єкти навколишнього середовища одна з важливих властивостей природного і штучного інтелекту. В медичній практиці широко використовують системи автоматичної класифікації медичних зображень та методи скринінгових досліджень для виявлення хворих під час профілактичних обстежень.

Розглянемо формальні стратегії, на яких ґрунтуються системи класифікації.

Нехай об'єкт перебуває в одному з $M \geq 2$ можливих станів множини $V = \{V_1, \dots, V_M\}$. Сукупність характеристик об'єкта

$$x^{(N)} = (x_1, \dots, x_N), \quad (5.1)$$

за якими потенційно можна відрізнити один стан об'єкта від іншого, називають діагностичними ознаками.

Побудуємо вирішальне правило

$$D(x^{(N)}): X^{(N)} \rightarrow M,$$

яке відобразить множину $X^{(N)} \triangleq \{x^{(N)}\}$ можливих значень ознак x_1, \dots, x_N на множину $M \triangleq \{1, \dots, M\}$ номерів можливих діагнозів V_1, \dots, V_M .

Інакше кажучи, потрібно побудувати алгоритм визначення значень індикаторної змінної у вигляді

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

де $\Omega_m, m=1, \dots, M$ – області простору $X^{(N)}$, що не перетинаються, в яких приймають рішення на користь класу V_m .

Математичні методи класифікації можна розділити на дві групи: детерміновані, які в явному вигляді не використовують статистичні властивості спостережуваних образів, і статистичні, які ґрунтуються на методах теорії ймовірностей та математичної статистики.

Статистична постановка задачі виникає у тих важливих випадках, коли за різних станів об'єкта $V_i \neq V_j$ можуть збігатися значення кожної з наявних ознак x_n ($1 \leq n \leq N$), інакше кажучи, допускається, що множини значень кожної з ознак (5.1), відповідні різним класам, перетинаються.

Статистична постановка задачі класифікації ґрунтується на різних формальних стратегіях, однією з яких є байєсівська стратегія.

Згідно з байєсівським підходом стани V_1, \dots, V_M об'єкта розглядають як випадкові події з відомими апіорними ймовірностями

$P(V_m)$, $\sum_{m=1}^M P(V_m) = 1$, а ознаки (5.1) – як випадкові величини, для яких об'єктивно існують умовні розподіли $p(x^{(N)} | V_m)$.

Множини

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

на яких зосереджені ознаки, називають власними областями класів у багатовимірному просторі ознак.

Інтегрування багатовимірного розподілу $p(x^{(N)} | V_m)$ множиною можливих значень усіх ознак x_1, \dots, x_N , окрім ознаки x_n ($1 \leq n \leq N$), дає змогу отримати уявлення про одновимірні (маргінальні) розподіли цієї ознаки $p(x_n | V_1), \dots, p(x_n | V_M)$. Множина $X_{nm} \triangleq \{x_n : p(x_n | V_m)\}$ є носієм маргінального розподілу $p(x_n | V_m)$ ознаки x_n , $1 \leq n \leq N$ у класі $V_m, m=1, \dots, M$.

Часто в задачах діагностики розглядають лише два можливі діагнози, наприклад, «здоровий – хворий». Алгоритми, орієнтовані на такий випадок, називають бінарними класифікаторами.

На рис. 5.1. показано чотири можливі варіанти розміщення носіїв X_1 , X_2 маргінальних розподілів $p(x|V_1)$, $p(x|V_2)$ деякої ознаки x у класах V_1 і V_2 .

Перший випадок ілюструє «ідеалізовану» ситуацію, яка досить рідко трапляється на практиці: носії класів X_1 і X_2 повністю розділяються пороговим значенням x_0 . У такому разі статистична постановка завдання втрачає сенс, оскільки за будь-яким значенням $x < x_0$ приймаються однозначні рішення на користь V_1 , а за будь-яким значенням $x > x_0$ – безпомилкові рішення на користь V_2 .

У випадку 2 множини X_1 і X_2 вже частково перетинаються. У цьому випадку безпомилкові рішення можуть бути прийняті тільки для значень $x \in (X_1 \cup X_2) / (X_1 \cap X_2) \neq \emptyset$, тоді як для значень $x \in X_1 \cap X_2 \neq \emptyset$ не можуть бути прийняті однозначно вірні рішення.

У випадку 3, коли носії розподілів X_1 і X_2 збігаються, але не збігаються умовні розподіли $p(x|V_1)$ і $p(x|V_2)$ вже за будь-якого значення x допускається помилка. Умови, що гарантують діагностичну цінність такої ознаки, будуть досліджені далі.

І, нарешті, у випадку 4, коли збігаються як множини X_1 і X_2 , так і умовні розподіли $p(x|V_1)$ і $p(x|V_2)$, окрема ознака x стає марною. Однак у сукупності дві такі ознаки x_1 і x_2 можуть бути не тільки корисні, а й забезпечити безпомилкову класифікацію (рис. 5.2).

Доведено, що така можливість можлива лише тоді, коли між x_1 і x_2 є умовна статистична залежність одразу в обох класах, тобто

$$p(x_1, x_2 | V_m) \neq p(x_1 | V_m)p(x_2 | V_m), \quad m = 1, 2.$$

Байєсівська стратегія класифікації ґрунтується на відомій теоремі Байєса, яка започаткувала розвиток теорії статистичних рішень.

Оскільки доведення теореми займає всього декілька рядків, нагадаємо цю теорему.

Теорема Байєса. Нехай випадкова подія A може статися одночасно з однією з інших несумісних подій, що утворюють повну групу – гіпотез H_1, \dots, H_N , для яких відомі апріорні ймовірності $P(H_i)$, $i = 1, \dots, N$, такі, що

$$\sum_{i=1}^N P(H_i) = 1.$$

Нехай також відомі умовні ймовірності події A для кожною з гіпотез, тобто числа

$$P(A/H_i) \quad i = 1, \dots, N.$$

Тоді, якщо за результатами експерименту встановлено, що подія A сталася (випадкова подія A могла не статися!), то апостеріорна ймовірність i -ї гіпотези можна визначити за формулою

$$P(H_i | A) = \frac{P(H_i)P(A | H_i)}{\sum_{i=1}^N P(H_i)P(A | H_i)}, \quad i = 1, \dots, N. \quad (5.2)$$

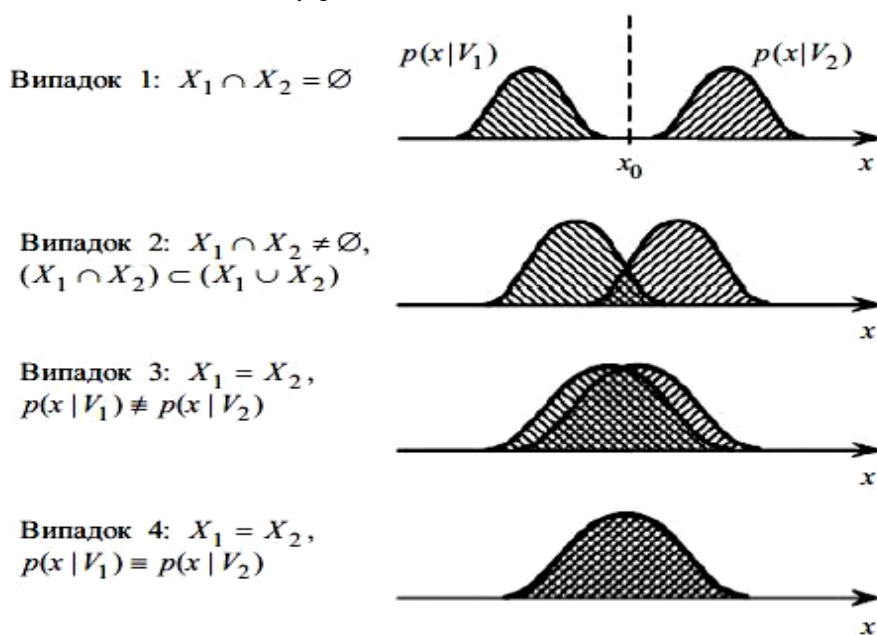


Рис. 5.1. Топологія носіїв умовних розподілів ознаки x

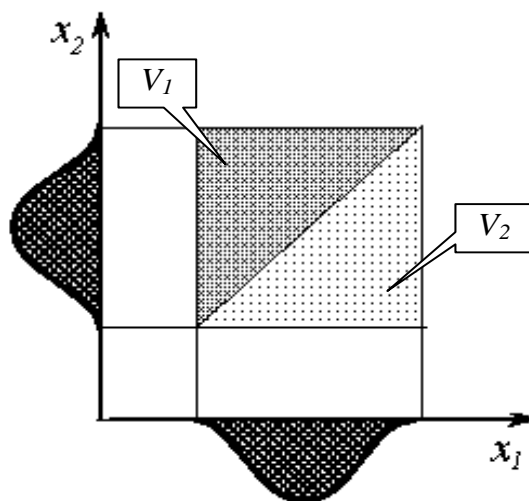


Рис. 5.2. Безпомилкова класифікація V_1 і V_2 за сукупністю окремо не корисних ознак x_1 та x_2

Доведення. За означенням сумісну ймовірність події A з i -ю гіпотезою H_i , $i = 1, \dots, N$ можна записати у вигляді

$$P(AH_i) = P(H_i)P(A | H_i) = P(A)P(H_i | A), \quad i = 1, \dots, N.$$

Звідси в припущенні, що $P(A) = 0$ випливає, що

$$P(H_i | A) = \frac{P(H_i)P(A | H_i)}{P(A)}, \quad i = 1, \dots, N. \quad (5.3)$$

За формулою повної ймовірності маємо

$$P(A) = \sum_{i=1}^N P(A | H_i)P(H_i). \quad (5.4)$$

Підставляючи (5.4) у (5.3), отримаємо (5.2). Теорему доведено.

Застосуємо теорему Байєса для побудови стратегії визначення стану об'єкта $V_j \in \{V_1, V_2\}$ у припущенні, що відомі апіорні ймовірності $P(V_1), P(V_2)$ і багатовимірні умовні розподіли $p(x^{(N)} | V_1)$, $p(x^{(N)} | V_2)$ вектора ознак (5.1) у класах.

Нехай за результатами обстеження визначені поточні значення ознак $x_1 = \hat{x}_1, \dots, x_N = \hat{x}_N$. З відомим вектором $\hat{x}^{(N)} = (\hat{x}_1, \dots, \hat{x}_N)$ ознак можна згідно з формулою Байєса визначити апостеріорні ймовірності станів об'єкта:

$$P(V_1 | \hat{x}^{(N)}) = \frac{P(V_1)p(\hat{x}^{(N)} | V_1)}{\sum_{m=1,2} P(V_m)p(\hat{x}^{(N)} | V_m)} \quad (5.5)$$

та

$$P(V_2 | \hat{x}^{(N)}) = \frac{P(V_2)p(\hat{x}^{(N)} | V_2)}{\sum_{m=1,2} P(V_m)p(\hat{x}^{(N)} | V_m)}. \quad (5.6)$$

За правилом максимуму апостеріорних ймовірностей з формул (5.5) і (5.6) випливає, що рішення на користь V_1 має прийматися в тому випадку, коли

$$P(V_1 | \hat{x}^{(N)}) > P(V_2 | \hat{x}^{(N)}),$$

і на користь V_2 , коли

$$P(V_1 | \hat{x}^{(N)}) < P(V_2 | \hat{x}^{(N)}).$$

Оскільки знаменники правих частин співвідношень (5.5) та (5.6) однакові, то правило максимуму апостеріорних ймовірностей $P(V_1 | \hat{x}^{(N)})$ і $P(V_2 | \hat{x}^{(N)})$ еквівалентно правилу

$$s = \begin{cases} 1, & \text{если } \lambda(\hat{x}^{(N)}) > \lambda_0; \\ 2, & \text{если } \lambda(\hat{x}^{(N)}) < \lambda_0, \end{cases} \quad (5.7)$$

де

$$\lambda(\hat{x}^{(N)}) = \frac{p(\hat{x}^{(N)} | V_1)}{p(\hat{x}^{(N)} | V_2)}, \quad (5.8)$$

є відношенням правдоподібності для поточного значення $x^{(N)} = \hat{x}^{(N)}$ вектора ознак, а λ_0 – поріг, який визначається за відношенням апріорних ймовірностей класів

$$\lambda_0 = \frac{P(V_2)}{P(V_1)}.$$

Для фіксованого значення вектора ознак $x^{(N)} = \hat{x}^{(N)}$ умовна ймовірність помилкових рішень за правилом (5.7) визначає співвідношення

$$P(E | \hat{x}^{(N)}) \triangleq 1 - \max\{P(V_1 | \hat{x}^{(N)}), P(V_2 | \hat{x}^{(N)})\}. \quad (5.9)$$

Звідси випливає, що правило (5.7) забезпечує мінімум середньої ймовірності помилки:

$$P(E) \triangleq \sum_{x^{(N)} \in X^{(N)}} p(x^{(N)}) [1 - \max\{P(V_1 | x^{(N)}), P(V_2 | x^{(N)})\}] \quad (5.10)$$

на множині $X^{(N)}$ можливих значень вектора $x^{(N)}$ ознак.

У теорії статистичних рішень втратами $L(i, j)$ ($i, j = 1, 2$) оцінюють діагноз на користь i -го класу тоді, коли об'єкт перебуває в j -му стані. В такому разі середні втрати (середній ризик) рішень, що приймаються, визначає співвідношення

$$R = \sum_{x^{(N)} \in X^{(N)}} \sum_{j=1}^2 L(i, j) P(V_j) p(x^{(N)} | V_j). \quad (5.11)$$

Можна показати, що за відомим відношенням правдоподібності $\lambda(\hat{x}^{(N)})$ оптимальне рішення, яке забезпечує середній ризик (5.11), має прийматися за правилом

$$s = \begin{cases} 1, & \text{если } \lambda(\hat{x}^{(N)}) > \theta_0; \\ 2, & \text{если } \lambda(\hat{x}^{(N)}) < \theta_0, \end{cases} \quad (5.12)$$

де

$$\theta_0 = \frac{P(V_2)(L_{21} - L_{22})}{P(V_1)(L_{12} - L_{11})}. \quad (5.13)$$

Можна бачити, що правило (5.7) є частковим випадком правила (5.12), коли

$$L_{11} = L_{22} = 0 \text{ та } L_{12} = L_{21} = 1,$$

тобто втрати від правильних рішень дорівнюють нулеві, а втрати від помилкових – дорівнюють одиниці. В такому випадку середній ризик (5.11) вироджується в середню ймовірність помилкових рішень (5.10).

Виникають ситуації, коли байєсівську стратегія не можна застосувати. В такому разі використовують небайєсівські стратегії класифікації.

Розглянемо коротко одну з таких стратегій – стратегію Неймана-Пірсона на прикладі задачі скринінгу, що зводиться до виявлення хворих людей у великій групі обстежуваних, де є і хворі (клас V_1) і умовно здорові особи, у яких немає аналізованого захворювання (клас V_2).

Для спрощення будемо вважати, що бінарний класифікатор має приймати рішення за результатом вимірювання всього одного фізіологічного показника $x \in X$ відповідно до правила:

$$\begin{aligned} & \text{Високий ризик захворювання, якщо } x \in X_1, \\ & \text{Низький ризик захворювання, якщо } x \in X_2, \end{aligned} \quad (5.14)$$

де X_1, X_2 – області відповідних рішень, $X_1 \cup X_2 = X$, $X_1 \cap X_2 = \emptyset$.

Припускається, що під час використання алгоритму (5.14) можуть виникати помилки двох типів:

$$E_1, \text{ коли } x \in X_2, \text{ але особа хвора,} \quad (5.15)$$

$$E_2, \text{ коли } x \in X_1, \text{ але особа здорова.} \quad (5.16)$$

Випадкові події (5.15) і (5.16) в теорії статистичних рішень називають відповідно помилками пропуску цілі та хибної тривоги.

За відомими умовними розподілами $p(x|V_1)$ і $p(x|V_2)$ значень ознаки x у хворих і здорових осіб імовірності випадкових подій E_1 і E_2 визначаються за співвідношеннями:

$$P(E_1) = \int_{x_2} p(x|V_1) dx,$$

$$P(E_2) = \int_{x_1} p(x|V_2) dx.$$

Якщо умовні розподіли $p(x|V_1)$ і $p(x|V_2)$ для хворих і здорових пацієнтів одномодальні, то рішення за (5.11) зводяться до простого правила:

$$\begin{aligned} &\text{Високий ризик захворювання, якщо } x > x_0, \\ &\text{Низький ризик захворювання, якщо } x \leq x_0, \end{aligned} \quad (5.17)$$

де x_0 – порогове значення фізіологічного показника x .

Наприклад, порогове значення глюкози у крові, яке застосовують під час скринінгу сахарного діабету, становить $x_0 = 5$ ммоль/літр (рис. 5.3)

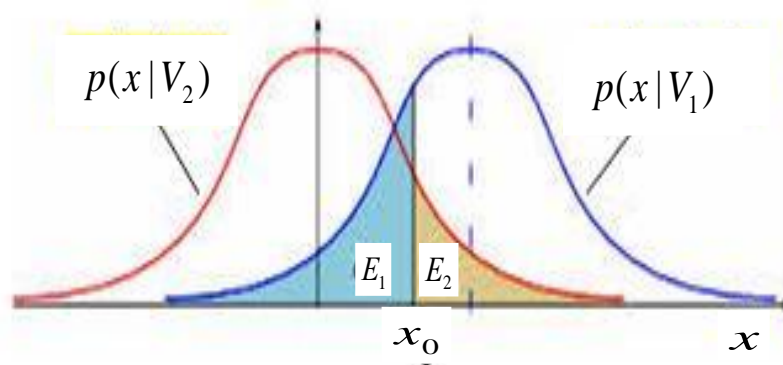


Рис. 5.3. Умовні розподіли $p(x|V_1)$ і $p(x|V_2)$ глюкози у крові

У загальному випадку згідно зі стратегією Неймана-Пірсона області рішень $X_1 \subset X$ і $X_2 \subset X$ обирають так, щоб мінімізувати ймовірність помилки пропуску цілі за обмеженням на ймовірність хибної тривоги, тобто

$$P(E_1) \rightarrow \min, \quad (5.18)$$

$$P(E_2) \leq \varepsilon, \quad (5.19)$$

де $\varepsilon = \text{const}$.

Окрім розглянутих стратегій для розв'язування задач класифікації об'єктів різної типу відомі й інші формальні стратегії. Зокрема, аналіз Вальда передбачає послідовний аналіз компонент вектора ознак $x^{(N)} = (x_1, \dots, x_N)$ у припущенні, що відомі умовні розподіли $p(x_1 | V_m)$, $p(x_1, x_2 | V_m), \dots, p(x_1, \dots, x_N | V_m)$, $m=1,2$.

На кожному $n < N$ кроці оцінюють відношення правдоподібності

$$\lambda_n = \frac{p(x_1, \dots, x_n | V_1)}{p(x_1, \dots, x_n | V_2)}$$

та приймають діагностичні рішення за схемою

$$\begin{aligned} &\text{клас } V_1, \text{ якщо } \lambda_n \geq A, \\ &\text{клас } V_2, \text{ якщо } \lambda_n \leq B, \\ &\text{аналіз наступної ознаки } x_{n+1}, \text{ якщо } A > \lambda_n > B, \end{aligned} \quad (5.20)$$

де A, B – критичні значення, які згідно з теоремою Вальда задовольняють умови

$$\begin{aligned} A &\geq \log_2 \left(\frac{P(E_2)}{1 - P(E_1)} \right), \\ B &\leq \log_2 \left(\frac{1 - P(E_2)}{P(E_1)} \right), \end{aligned}$$

в яких є ймовірності $P(E_1)$ і $P(E_2)$ помилок першого та другого роду.

5.2. Оцінювання ефективності бінарного класифікатора

Відомі різні підходи до оцінювання ефективності бінарних класифікаторів. Зручним засобом є метод, який ґрунтується на аналізі операційної характеристичної кривої (ROC – Receive Operating Characteristic curve). Традиційний ROC-аналіз ґрунтується на оцінюванні двох характеристик діагностичного тесту – чутливості та специфічності, які оцінюють ймовірності помилок пропуску цілі та хибної триво-

ги, прийнятих у теорії статистичних рішень. Для інтегрального оцінювання ефективності бінарного класифікатора, що реалізує діагностичне вирішальне правило, найчастіше визначають площу під ROC-кривою.

Розглянемо детальніше традиційний ROC-аналіз на прикладі задачі скринінгу.

Нехай для виявлення хворих передбачається використовувати бінарний класифікатор, який на підставі доступної інформації приймає одне з двох рішень: пацієнт «Хворий» (позитивний результат тестування) або «Здоровий» (негативний результат тестування). Для оцінювання ефективності такого тесту проводиться його випробування на репрезентативній групі пацієнтів із заздалегідь відомими діагнозами. Результати іспиту можна подати у вигляді матриці (табл. 5.1), в комірках якої наведено числа відповідних спостережень.

Таблиця 5.1. Результати тестування репрезентативної групи

Поточний діагноз	Рішення прийняте за результатами тестування	
	Пацієнт «Хворий»	Пацієнт «Здоровий»
«Хворий» (клас V_1)	Кількість правильних позитивних результатів TP	Кількість помилкових негативних результатів FN
«Здоровий» (клас V_2)	Кількість помилкових позитивних результатів FP	Кількість правильних негативних результатів TN

За даними табл. 5.1 легко оцінити прийняті у медичній діагностиці операційні характеристики тесту, а саме чутливість

$$S_E = \frac{TP}{TP + FN}, \quad (5.21)$$

яка визначає частку правильних позитивних результатів TP , отриманих для всіх хворих осіб $TP + FN$, та специфічність

$$S_P = \frac{TN}{TN + FP}, \quad (5.22)$$

яка визначає частку правильних негативних результатів TN , отриманих для всіх здорових осіб $TN + FP$.

Бінарний класифікатор з фіксованими операційними характеристиками (5.21) та (5.22) відображають точкою в ROC-просторі з координатами S_E та $1 - S_P$. Зауважимо, що величина

$$1 - S_P = 1 - \frac{TN}{TN + FP} = \frac{TN + FP - TN}{TN + FP} = \frac{FP}{TN + FP},$$

визначає частку помилкових позитивних результатів, отриманих для всіх здорових осіб $TN + FP$.

ROC-простір дає змогу графічно зобразити діагностичну цінність тесту та порівнювати ефективність різних тестів.

Ідеальний тест A (рис. 5.4) розміщений у точці з координатами $(0, 1)$. За результатами такого тесту всі хворі належать до класу V_1 і не виникає помилок хибної тривоги. Звідси випливає, що чим ближче тест з операційними характеристиками S_E і S_P до точки A , тим він ефективніший.

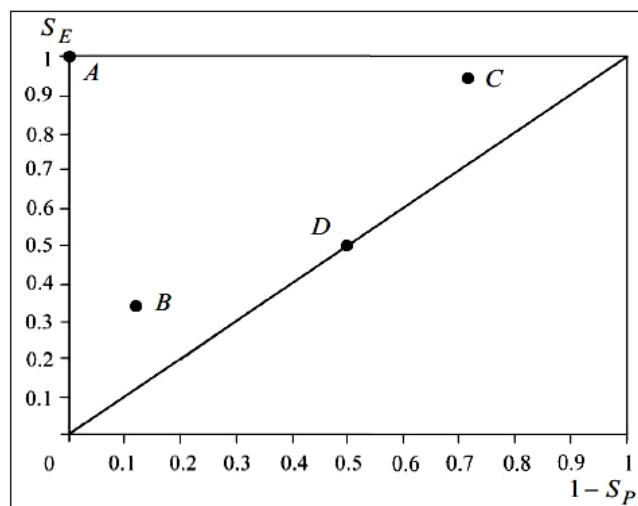


Рис. 5.4. Бінарні класифікатори в ROC-просторі

Тест B , розміщений у лівому нижньому куті, називають «консервативним», оскільки за малого відсотка помилок хибної тривоги $1 - S_P$ він має низьку чутливість S_E . Оскільки тест B має високе значення специфічності S_P , прийняття рішення: пацієнт «Хворий» слід сприймати з високим ступенем довіри. Водночас рішення пацієнт «Здоровий» може бути неправильним через низьку чутливість S_E .

Тест C , розміщений у правому верхньому куті, називають «ліберальним», оскільки за великих значень чутливості S_E він має значний відсоток помилок хибної тривоги $1 - S_P$. Через високу чутливість S_E

рішення: пацієнт «Здоровий», імовірноше правильне, тоді як рішення пацієнт «Хворий» може бути помилковим через низьку специфічність S_p .

Тест D , розміщений на діагоналі $S_E = 1 - S_p$ ROC-простору, є некорисним. Такі тести еквівалентні класифікаторам, які використовують стратегію «випадкового вгадування» класів V_1 та V_2 .

Як вже зазначалось, у медичній діагностиці часто використовують порогові вирішувальні правила вигляду:

$$\begin{aligned} &\text{хворий, якщо } x > x_0. \\ &\text{здоровий, якщо } x \leq x_0, \end{aligned} \quad (5.23)$$

де x – діагностична ознака, а x_0 – порогове значення ознаки x . У такому випадку з'являється можливість визначити пари S_E і S_p для різних значень x_0 , які відповідають послідовності точок бінарного класифікатора у ROC-просторі.

З правила (5.23) впливає простий алгоритм побудови експериментальної ROC-кривої:

- відсортувати значення спостережуваної ознаки x у порядку зменшення;
- послідовно зменшувати (з деяким кроком) поріг x_0 та за екзамінаційною вибіркою спостережень обчислювати відповідні пари значень S_E і S_p ;
- відобразити отриману послідовність пар S_E і S_p у ROC-просторі (рис. 5.5).

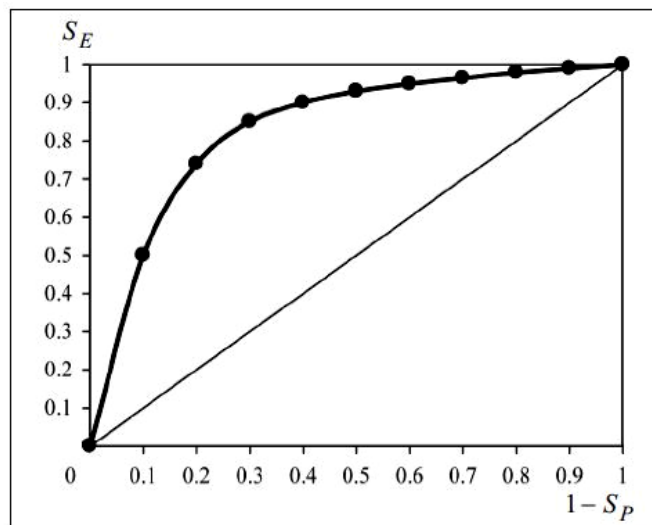


Рис. 5.5. ROC-крива тесту, що ґрунтується на правилі (5.23)

Форма ROC-кривої пов'язана з топологією носіїв класів V_1 і V_2 . Для ілюстрації на рис. 5.6 зображено форми ROC-кривих, що відповідають чотирьом можливим випадкам розміщення носіїв класів (див. рис. 5.1).

Для порівняння ефективності діагностичних тестів зручно використовувати інтегральну характеристику, якою зазвичай вважають площу AUC (**A**rea **U**nder **C**urve), обчислену під ROC-кривою. Величина AUC визначає середню чутливість \bar{S}_E тесту для можливих значеннях специфічності $0 \leq S_p \leq 1$ або середню специфічність \bar{S}_p тесту для можливих значеннях чутливості $0 \leq S_E \leq 1$, причому $AUC = 1$ для ідеального тесту та $AUC = 0,5$ для безкорисного. Тому вважається, що чим більше AUC наближається до одиниці, тим ефективніше тест.

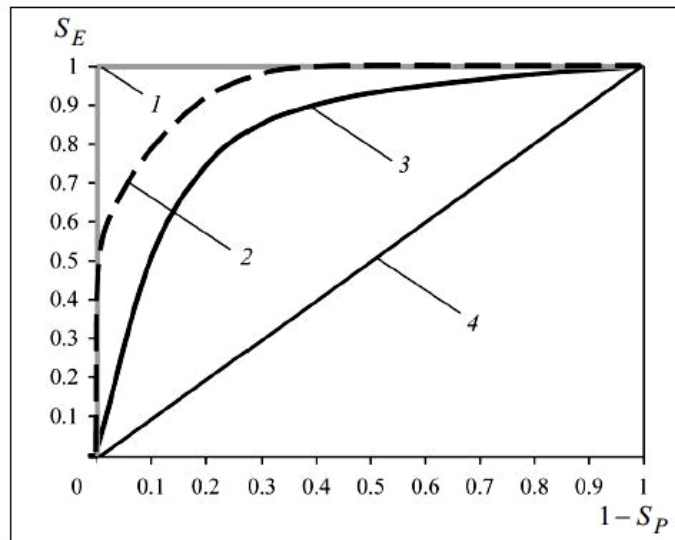


Рис. 5.6. Взаємозв'язок ROC-кривих з топологією носіїв класів
(пояснення в тексті)

Зрозуміло, що для операційних характеристик потенційно корисних тестів, які проходять вище діагоналі ROC-простору, виконується умова

$$S_E > 1 - S_p. \quad (5.24)$$

Розглянемо удосконалення традиційного ROC-аналізу, що дають змогу оцінювати ефективність діагностичного тесту з погляду очікуваних втрат (середнього ризику).

Нехай апіорна ймовірність $P(V_1)$ появи хворого в групі обстежуваних осіб відома за даними медичної статистики і характеризує поширеність (преваленс) тієї чи іншої хвороби. Відповідно величина $P(V_2) = 1 - P(V_1)$ характеризує апіорну ймовірність зустріти у групі людину, яка не має досліджуваного захворювання.

Будемо характеризувати можливі помилки діагностичного тесту величинами L_{12}, L_{21} які визначають втрати за помилки пропуску мети (не виявлено хворого з групи V_1) та помилкової тривоги (здоровий пацієнт, що належить до групи хворих V_1).

Тоді очікувані втрати від неправильної діагностики визначає співвідношення

$$R = P(V_1)(1 - S_E)L_{12} + [1 - P(V_1)](1 - S_P)L_{21}. \quad (5.25)$$

Умова (5.25) дає змогу за відомими $P(V_1)$, L_{12} і L_{21} обчислити середній ризик помилкової діагностики (апостеріорний ризик) для тесту, який має операційні характеристики S_E і S_P . Виникає природне запитання: чи завжди такий тест буде корисним?

Щоб відповісти на це запитання, передусім необхідно сформулювати саме означення корисності діагностичного тесту. Таким означенням може бути умова того, що тест забезпечує менші середні втрати, ніж ті, які спостерігалися би без використання діагностичного тесту. Інакше кажучи, поняття корисності логічно сформулювати у такий спосіб.

Означення 5.1. Діагностичний тест є корисним, якщо виконується строга нерівність

$$R < R_0, \quad (5.26)$$

тобто апостеріорний ризик R менше апіорного ризику R_0 , що реалізує стратегія прийняття рішень лише на підставі відомих значень $P(V_1)$, L_{12} і L_{21} .

Наведемо деякі пояснення. Зрозуміло, якщо бінарний класифікатор реалізує байєсівську стратегію прийняття рішень, яка мінімізує апостеріорний ризик, то виконується умова (5.26) або умова $R = R_0$. Інакше кажучи, формально побудований алгоритм діагностики не може бути «шкідливим».

Однак на практиці часто немає інформації, яка потрібна для реалізації стратегії Байєса, і діагностичний тест будується не формальним способом. У таких важливих для практичного застосування випадках за фіксованих $P(V_1)$, L_{12} і L_{21} і відомих S_E , S_P замість умови (5.26) може виконуватися на перший погляд «парадоксальна» умова $R > R_0$.

Розглянемо таку можливість на модельному прикладі. Нехай для виявлення «Хворих» пацієнтів у групі з 10000 осіб запропоновано використовувати діагностичний тест із достатньо високими операційними характеристиками $S_E = 0,9$ і $S_P = 0,9$. Нехай преваленс захворювання $P(V_1) = 0,02$, а втрати від помилок пропуску цілі та хибної тривоги вважаються однаковими: $L_{12} = L_{21} = 1$.

На рис. 5.7 наведено очікувані результати тестування.

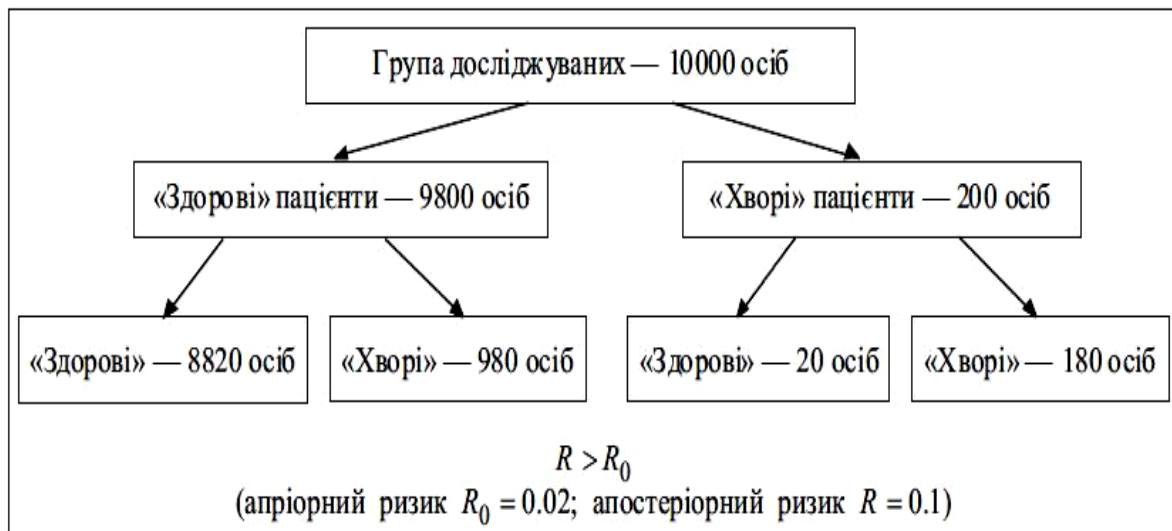


Рис. 5.7. Очікувані результати тестування

Якщо не використовувати діагностичний тест, то оптимальна стратегія має бути такою: всіх обстежуваних визнавати «Здоровими» пацієнтами (інакше втрати R_0 будуть більшими!!!) Тоді середні апріорні втрати на одного обстежуваного, зумовлені тим, що 200 «Хворих» пацієнтів не будуть виявлені, складуть

$$R_0 = (L_{12} \cdot 200) / 10000 = (1 \cdot 200) / 10000 = 0,02. \quad (5.27)$$

У разі використанні діагностичного тесту з чутливістю $S_E = 0,9$ та специфічністю $S_P = 0,9$ середні втрати, пов'язані з помилково по-

зитивними результатами діагностики 980 «Здорових» пацієнтів та не виявленням 20 «Хворих» пацієнтів, становитимуть

$$R = (L_{21} \cdot 980 + L_{12} \cdot 20) / 10000 = (1 \cdot 980 + 1 \cdot 20) / 10000 = 0,1. \quad (5.28)$$

Зі співвідношень (5.27) і (5.28) випливає, що $R > R_0$ і такий тест не можна визнати корисним, оскільки за заданих втрат від помилок пропуску цілі та хибної тривоги він лише збільшує очікувані втрати!

Незважаючи на «парадоксальність» того, що $R > R_0$, отриманий результат цілком обґрунтований. Неважко показати, що за розглянутими зазначеними $P(V_1)$, L_{12} , L_{21} , S_E і S_P рішення: пацієнт «Хворий» не є обґрунтованим для мінімізації середнього ризику. Інакше кажучи, для кожного обстежуваного завжди маємо приймати рішення: пацієнт «Здоровий», яке збігається з апіорними рішеннями.

Покажемо, що умова (5.24) є лише необхідною, але недостатньою умовою, яка гарантує виконання строгої нерівності (5.26).

Твердження 5.1. Діагностичний тест гарантовано корисний у сенсі означення 5.1, якщо його чутливість S_E і специфічність S_P пов'язані співвідношенням

$$S_E > m(1 - S_P), \quad \text{якщо} \quad m \geq 1; \quad (5.29)$$

або

$$S_E > 1 - m + m(1 - S_P), \quad \text{якщо} \quad m < 1; \quad (5.30)$$

де

$$m = \frac{[1 - P(V_1)]}{\omega P(V_1)}, \quad (5.31)$$

є величиною, яка залежить від апіорної ймовірності $P(V_1)$ і співвідношення втрат від помилок пропуску цілі та хибної тривоги

$$\omega = \frac{L_{12}}{L_{21}}. \quad (5.32)$$

Доведення. Якщо не використовувати діагностичного тесту і приймати рішення лише на підставі значень $P(V_1)$, L_{12} та L_{21} , то апіорна стратегія зводиться до одного з двох можливих варіантів:

– усі обстежувані належать до класу «Хворих» V_1 , і тоді апіорний ризик становить

$$R_0^+ = L_{21}[1 - P(V_1)]; \quad (5.33)$$

– приймати рішення про те, що кожен обстежуваний «Здоровий», і тоді апріорний ризик становить

$$R_0^- = L_{12}P(V_1). \quad (5.34)$$

З (5.33), (5.34) з урахуванням (5.31) випливає, що мінімальний апріорний ризик, який використано у правій частині нерівності (5.26), можна записати так:

$$R_0 = \begin{cases} L_{21}[1 - P(V_1)], & \text{якщо } m < 1, \\ L_{12}P(V_1), & \text{якщо } m \geq 1. \end{cases} \quad (5.35)$$

Унаслідок підстановки виразів (5.25) та (5.35) у нерівність (5.26) після елементарних перетворень призводить до співвідношень (5.29) та (5.30). Отже, твердження доведене.

Як видно з виразу (5.28) для перевірки умов (5.26) і (5.27) достатньо мати лише інформацією про співвідношення втрат (5.29) і така інформація більш доступна під час розв'язування прикладних задач, ніж інформація про абсолютні значення втрат від пропуску цілі L_{12} та хибної тривоги L_{21} .

Умови (5.29) та (5.30) дають змогу обґрунтовано підходити до оцінювання корисності нових діагностичних тестів. Для цього доцільно удосконалити традиційний ROC-аналіз, обмеживши відповідний фрагмент ROC-кривої (рис. 5.8).

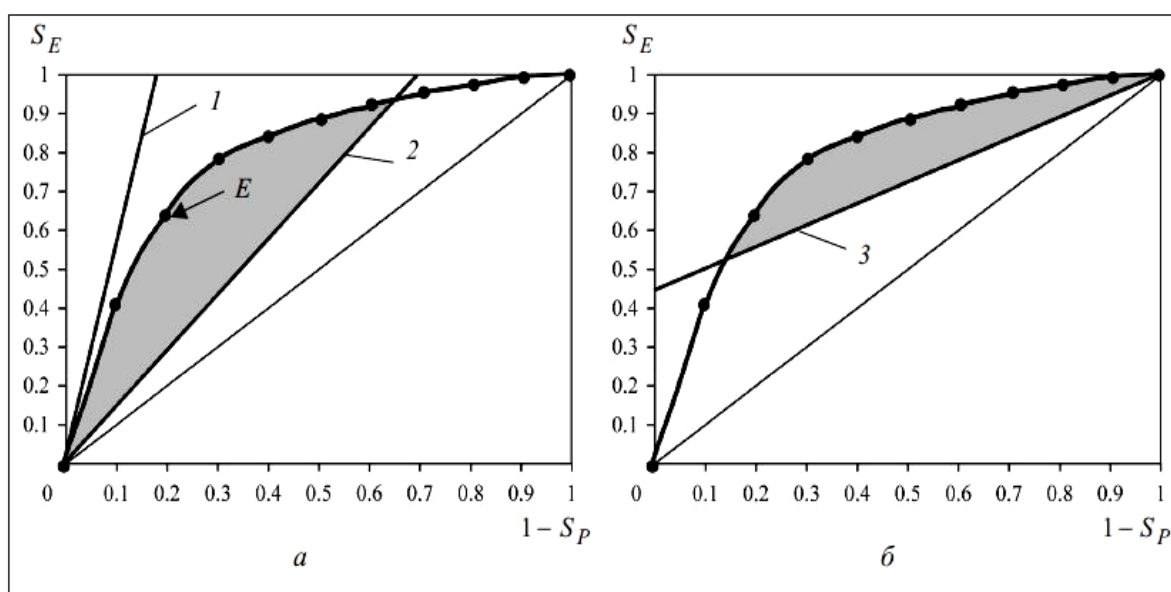


Рис. 5.8. Ілюстрація ідеї удосконаленого ROC-аналізу для випадків $m \geq 1$ (а) та $m < 1$ (б)

Пряма 1 (рис. 5.8, а) відповідає випадку, коли $P(V_1) = 0,15$ і $\omega = L_{12}L_{21}^{-1} = 1$. Отже, тест некорисний, оскільки пряма 1 на перетинає ROC-кривої. Однак якщо припустити, що втрати від пропуску вчетверо вище втрат від хибної тривоги, тобто припустити, що $L_{12}L_{21}^{-1} = 4$, то відповідна пряма 2 вже перетинає ROC-криву. Таким чином, тест, що має чутливість $S_E = 62,5\%$ і специфічність $S_P = 80\%$ виявиться гарантовано корисним згідно з (5.26), оскільки за таких умов точка E , яка відповідає наведеним операційним характеристикам, розміщена на допустимій частині ROC-кривої.

На рис. 5.8, б, коли $m < 1$, обмежувальна пряма 3 починається зі значень $S_P = 0$, $S_E = 1$, і якщо вона перетинає ROC-криву, то на відповідному її фрагменті є значення S_P і S_E , за яких тест буде корисним.

Зауважимо, що для $m = 1$, достатні умови (5.29) і (5.30) збігаються з необхідною умовою (5.24). У цьому випадку пряма, що обмежує фрагмент ROC-кривої, перебуває на діагоналі ROC-простору.

Цікаво порівняти отримані умови з відомими зі статей Томаса Фавсета (**T. Fawcett**) ізоліній ефективності (**Iso-performance lines**). З виразу (5.25) випливає, що рівняння ізолінії з очікуваними втратами $R = R^*$ має вигляд лінійної залежності в координатах $S_E, 1 - S_P$:

$$S_E = m(1 - S_P) + 1 - \frac{R^*}{L_{12}P(V_1)}. \quad (5.36)$$

Рівняння (5.36) збігається з рівняннями прямих (5.29) і (5.30) за виконання умови

$$R^* = \begin{cases} L_{21}P(V_1), & \text{якщо } m \geq 1, \\ L_{12}[1 - P(V_1)], & \text{якщо } m < 1, \end{cases}$$

права частина якої збігається з правою частиною (5.35).

Тобто ділянку корисних тестів обмежує ізолінія ефективності

$$S_E = m(1 - S_P) + 1 - \frac{R_0}{L_{12}P(V_1)}, \quad (5.37)$$

яка відповідає рівню апріорного ризику R_0 та обов'язково проходить через точку з координатами $(0,0)$, якщо $m \geq 1$, або точку з координатами $(1,1)$, якщо $m < 1$.

Зрозуміло, що можливі такі ситуації:

- ізолінія (5.34) перетинає ROC-криву і тоді $R < R_0$;
- ізолінія (5.34) дотична до ROC-кривої і тоді $R = R_0$;
- ізолінія (5.34) не перетинає ROC-криву і тоді $R > R_0$.

Умови (5.29) і (5.30) дають змогу визначити допустимий інтервал співвідношення втрат $\omega = L_{12}L_{21}^{-1}$, за яким тест із чутливістю S_E та специфічністю S_P залишається корисним для скринінгу захворювання з преваленсом $P(V_1)$. Такий інтервал визначається за формулою

$$\frac{1 - P(V_1)}{P(V_1)} \frac{1 - S_P}{S_E} \leq \omega \leq \frac{1 - P(V_1)}{P(V_1)} \frac{S_P}{1 - S_E}. \quad (5.38)$$

За допомогою удосконаленого ROC-аналізу доведена ефективність розглянутого вище інноваційного методу аналізу та інтерпретації електрокардіограм, який ґрунтується на переході від скалярного сигналу $z(t)$, до фазового портрету на площині.

Для обґрунтування діагностичної цінності β , що характеризує симетрію ділянки реполяризації усередненої фазової траєкторії ЕКГ, проводилося статистичне оброблення клінічних даних: 441 запис ЕКГ попередньо верифікованих хворих на ішемічну хворобу серця (ІХС) та 387 записів ЕКГ здорових осіб із контрольної групи.

Результат оброблення даних показав, що середні значення β мають статистично значущі відмінності у групах ($p < 0,01$) і становлять $0,956 \pm 0,43$ і $0,665 \pm 0,12$ у групі ІХС і контрольної групи відповідно.

Для оцінювання корисності діагностичного правила

$$\begin{aligned} & \text{ІХС,} && \text{якщо } \beta > \beta_0, \\ & \text{Норма,} && \text{якщо } \beta \leq \beta_0, \end{aligned} \quad (5.39)$$

що ґрунтується на порівнянні β з пороговим значенням β_0 , використовувався програмний застосунок, в якому реалізовано удосконалений ROC-аналіз. На рис. 5.9 наведено експериментальну ROC-криву. Сірим кольором виокремлено область значень S_E і S_P , які за умов $P(V_1) = 0,1$ та $\omega = 5$ задовольняють умови (5.26).

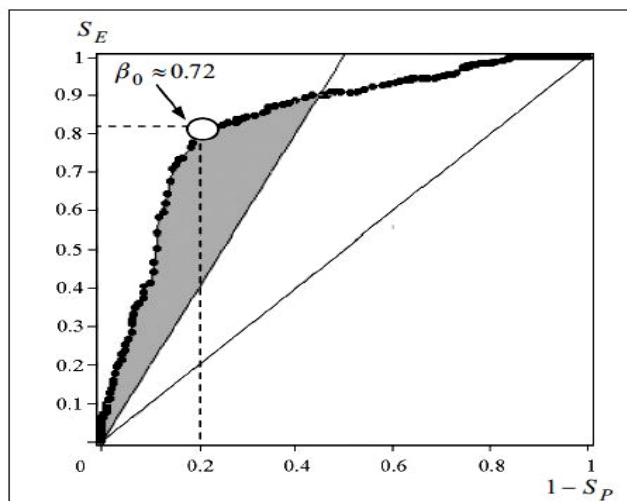


Рис. 5.9. Результат оцінювання корисності діагностичного правила (5.39)

Встановлено, що прийняття рішень за правилом (5.36) забезпечує чутливість $S_E = 81\%$ і специфічність $S_P = 78\%$, якщо $\beta_0 \approx 0,72$ – оптимальне значення порогу, якому відповідає точка, найбільш віддалена від діагоналі (рис. 5.9).

Відповідно до (5.38), якщо $S_E = 81\%$ і $S_P = 78\%$ діагностичне правило (5.39) гарантовано корисно для скринінгу ІХС з преваленсом $P(V_1) = 0,1$ у досить широкому діапазоні значень співвідношення втрат:

$$2,44 \leq \omega \leq 37.$$

Зауважимо, що в групу ІХС були включені лише пацієнти, у яких традиційний ЕКГ аналіз у 12 відведеннях не виявив жодних відхилень від норми. Водночас запропоноване діагностичне правило (5.39) підтвердило достатньо високі показники на такому «складному» клінічному матеріалі, і тому вважатимемо його цілком прийнятним для скринінгу ІХС.

5.3. Допустимі границі ймовірностей пропуску цілі та хибної тривоги

Нагадаємо, що чутливість S_E і специфічність S_P бінарного класифікатора оцінюють ймовірності помилок пропуску цілі $P(E_1)$ та хибної тривоги $P(E_2)$:

$$P(E_1) = 1 - S_E, \quad (5.40)$$

$$P(E_2) = 1 - S_p. \quad (5.41)$$

Співвідношення (5.40), (5.41) дають змогу переформулювати достатні умови (5.29) і (5.30), що гарантують корисність бінарного класифікатора.

Твердження 5.2. Бінарний класифікатор є корисним у тому разі (5.26), якщо

$$[(1 - P(V_1))]P(E_2) < \omega P(V_1)[1 - P(E_1)], \text{ коли } P(V_1)(1 + \omega) < 1 \quad (5.42)$$

або

$$[1 - P(V_1)][1 - P(E_2)] > \omega P(V_1)P(E_1), \text{ коли } P(V_1)(1 + \omega) > 1. \quad (5.43)$$

З твердження 5.2. випливає, що коли не виконується умова (5.42) або умова (5.43), то бінарний класифікатор є некорисним, що дає змогу визначити допустимі границі ймовірностей помилок пропуску цілі та хибної тривоги.

Твердження 5.3. Нехай виконується умова $P(V_1)(1 + \omega) < 1$. Тоді за будь-якої малої ймовірності пропуску цілі $P(E_1)$ бінарний класифікатор є некорисний у сенсі (5.26), якщо ймовірність $P(E_2)$ помилки хибної тривоги така, що

$$P(E_2) \geq \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (5.44)$$

Доведення. Зважаючи на те, що $P(V_1) \neq 1$, запишемо умову (5.42), яка гарантує корисність бінарного класифікатора в області $P(V_1)(1 + \omega) < 1$, в еквівалентній формі

$$P(E_2) < \frac{\omega P(V_1)[1 - P(E_1)]}{1 - P(V_1)}. \quad (5.45)$$

З нерівності (5.45) випливає, що бінарний класифікатор є некорисним, якщо

$$P(E_2) \geq \frac{\omega P(V_1)[1 - P(E_1)]}{1 - P(V_1)}. \quad (5.46)$$

Оскільки $P(E_1) \geq 0$, то посилення (5.46) через підстановку $P(E_1) = 0$ призводить до оцінювання припустимого значення ймовір-

ності $P(E_2)$ помилки хибної тривоги корисного бінарного класифікатора у вигляді умови (5.44). Отже, твердження 5.3 доведено.

З виразу (5.46) випливає важливий наслідок.

Наслідок 5.1. Щоб в області $P(V_1)(1 + \omega) < 1$ бінарний класифікатор, що реалізує стратегію Неймана-Пірсона (5.18), (5.19), був корисний з погляду зменшення середніх втрат, обмеження ε , що використано у правій частині (5.19), має задовольняти умову

$$\varepsilon < \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (5.47)$$

Тут потрібно дати невелике пояснення. Стратегія Неймана-Пірсона орієнтована на випадок, коли невідома апріорна ймовірність $P(V_1)$, наприклад, невідома поширеність маловивченого захворювання. Тому умова (5.47) має лише гносеологічний сенс і її правомірно розглядати як умову існування припустимого обмеження ε на ймовірність помилки помилкової тривоги $P(E_2)$.

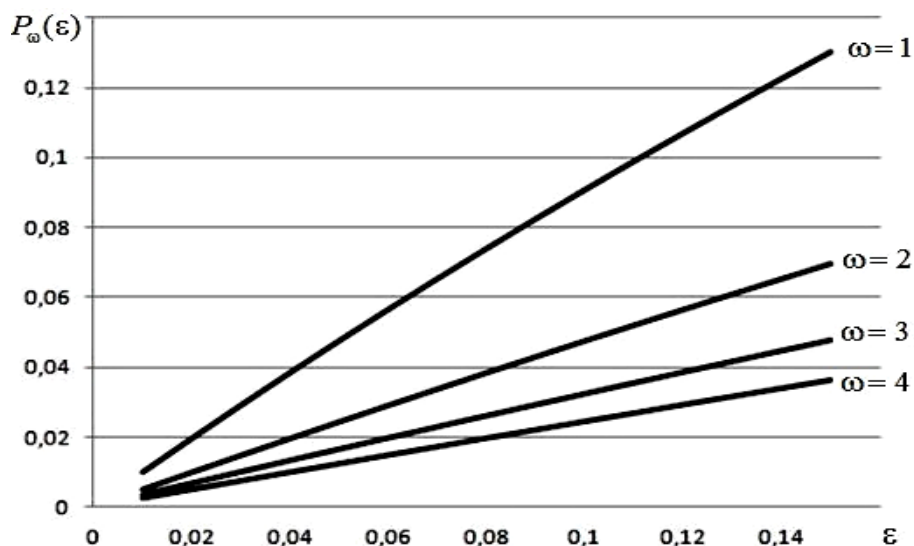
Разом з тим на підставі (5.47) після простих перетворень можна визначити граничну величину апріорної ймовірності $P_\omega(\varepsilon)$, яка залежить від обраного обмеження ε та фіксованого співвідношення ω :

$$P_\omega(\varepsilon) = \frac{\varepsilon}{\omega + \varepsilon}. \quad (5.48)$$

Вираз (5.48) вже має практичну користь. Якщо преваленс досліджуваного захворювання не перевершує величину (5.48), тобто $P(V_1) \leq P_\omega(\varepsilon)$, то стратегія Неймана-Пірсона буде неефективною.

Для ілюстрації на рис. 5.10 показано графіки залежності $P_\omega(\varepsilon)$ від ε , побудовані відповідно до (5.48) для різних значень ω .

Приклад 5.1. Для обстеження групи 10000 пацієнтів передбачається використовувати діагностичний тест, що реалізує стратегію Неймана-Пірсона. Співвідношення втрат (5.29) від помилок пропуску цілі та хибної тривоги обрано $\omega = 3$. Якщо встановити обмеження $\varepsilon = 0,07$ на ймовірність $P(E_2)$ помилки хибної тривоги, то специфічність тесту (відсоток істинно негативних результатів) становитиме $S_p = 93\%$. Припустимо, що тест має ідеальну чутливість (відсоток виявлення істинно хворих) $S_e = 100\%$.

Рис. 5.10. Графіки залежності $P_{\omega}(\epsilon)$ згідно з (5.48)

З виразу (5.48) випливає, що за умов $\epsilon = 0,07$, $\omega = 3$ гранична величина преваленсу становить $P_{\omega}(\epsilon) = 0,0228$. Тоді, якщо преваленс досліджуваного захворювання складе $P(V_1) = 0,02$, тобто $P(V_1) < P_{\omega}(\epsilon)$, то діагностичний тест з операційними характеристиками $S_E = 100\%$ і $S_p = 93\%$ буде некорисним.

Справді, якщо $P(V_1) = 0,02$, то очікувана кількість хворих у групі 10000 пацієнтів становитиме 200 осіб, які без використання тесту не будуть виявлені як хворі. У цьому випадку апіорний ризик на одного обстежуваного становитиме

$$R_0 = (L_{12} \cdot 200) / 10000 = (3 \cdot 200) / 10000 = 0,06. \quad (5.49)$$

Якщо ж використовувати діагностичний тест, то за чутливістю $S_E = 100\%$ всіх 200 хворих пацієнтів будуть виявлені. Однак за специфічністю $S_p = 93\%$ з 9800 здорових осіб буде визнано здоровими лише 9114 осіб, а решта 686 осіб буде помилково належати до групи V_1 хворих пацієнтів. Тобто апостеріорний ризик за рахунок помилково позитивних результатів становитиме

$$R = (L_{21} \cdot 686) / 10000 = (1 \cdot 686) / 10000 = 0,0686. \quad (5.50)$$

На основі порівняння (5.49) та (5.50) робимо висновок, що $R > R_0$, тобто діагностичний тест призвів до збільшення середніх втрат і, звідси, такий тест є некорисним.

Якщо ж в умовах прикладу змінити величину преваленсу, допускаючи, що $P(V_1) = 0,025$, то в цьому випадку $P(V_1) > P_\omega(\epsilon) = 0,0228$. Очікувана кількість хворих у групі 10000 обстежуваних становитиме вже 250 осіб, які не будуть виявлені без використання тесту. Отже, в цьому випадку апріорний ризик становитиме

$$R_0 = (L_{12} \cdot 200) / 10000 = (3 \cdot 250) / 10000 = 0,075. \quad (5.51)$$

Під час використання діагностичного тесту з чутливістю $S_e = 100\%$ всі 250 хворих будуть виявлені. Очікувана кількість здорових осіб становить 9750 осіб. За специфічністю $S_p = 93\%$ до групи V_2 здорових будуть належати 9068 істинно здорових осіб, а інші 682 особи будуть помилково визнані хворими пацієнтами. Отже, за таких умов апостеріорний ризик становитиме

$$R = (L_{21} \cdot 682) / 10000 = (1 \cdot 682) / 10000 = 0,0682. \quad (5.52)$$

З порівняння (5.51) та (5.52) робимо висновок, що тепер вже діагностичний тест є корисним, тому що $R < R_0$.

Розглянемо альтернативний варіант стратегії Неймана-Пірсона:

$$P(E_2) \rightarrow \min, \quad (5.53)$$

$$P(E_1) \leq \delta, \quad (5.54)$$

яка, на відміну від стратегії (5.18), (5.19) передбачає мінімізацію ймовірності $P(E_2)$ помилки хибної тривоги за умови, що на ймовірність помилки $P(E_1)$ пропуску цілі накладають задане обмеження δ .

Зауважимо, що для задач скринінгу характерні низькі значення преваленсу $P(V_1)$. Тому за відносно невеликих значеннях ω буде виконуватися умова $P(V_1)(1 + \omega) < 1$, для якої отримано наслідок 5.1.

Водночас під час діагностики особливо небезпечних захворювань співвідношення втрат ω має набувати дуже великих значень, наприклад, $\omega = 400$ під час діагностики онкологічних захворювань. У цьому випадку, навіть за малих значень преваленсу $P(V_1)$, вже виконується не умова $P(V_1)(1 + \omega) < 1$, а умова $P(V_1)(1 + \omega) > 1$.

Тому для визначення величини δ , за якою бінарний класифікатор, що реалізує стратегію (5.53), (5.54), буде корисний, скористаємося умовою (5.43) і доведемо таке твердження.

Твердження 5.4. Нехай виконується умова $P(V_1)(1 + \omega) > 1$. Тоді за будь-якої малої ймовірності помилки хибної тривоги $P(E_2)$ бінарний класифікатор, що реалізує стратегію Неймана-Пірсона (5.53), (5.54) є некорисним у сенсі (5.26), якщо ймовірність $P(E_1)$ помилки пропуску цілі задовольняє умову

$$P(E_1) \geq \frac{1 - P(V_1)}{\omega P(V_1)}. \quad (5.55)$$

Доведення. Оскільки $P(V_1) \neq 0$, то умова (5.43), яка гарантує корисність тесту для області $P(V_1)(1 + \omega) > 1$, можна подати в еквівалентній формі

$$P(E_1) < \frac{[1 - P(V_1)][1 - P(E_2)]}{\omega P(V_1)}. \quad (5.56)$$

З (5.56) випливає, що діагностичний тест є некорисний, якщо

$$P(E_1) \geq \frac{[1 - P(V_1)][1 - P(E_2)]}{\omega P(V_1)}. \quad (5.57)$$

Оскільки $P(E_2) \geq 0$, то посилення (5.57) за допомогою підстановки $P(E_2) = 0$ призводить до оцінювання припустимого значення ймовірності помилки пропуску цілі $P(E_1)$ у вигляді (5.55). Отже, твердження 5.4 доведено.

З виразу (5.57) може бути сформульовано такий наслідок.

Наслідок 5.2. Щоб в області $P(V_1)(1 + \omega) < 1$ бінарний класифікатор, що реалізує стратегію Неймана-Пірсона (5.53), (5.54) був корисний з погляду зменшення середніх втрат, обмеження δ , що використано у правій частині (5.54), має задовольняти умову

$$\delta \leq \frac{1 - P(V_1)}{\omega P(V_1)}. \quad (5.58)$$

Умова (5.58) після простих перетворень дає змогу отримати допустиме значення преваленсу

$$P_\omega(\delta) = \frac{1}{\delta\omega + 1}, \quad (5.59)$$

за яким діагностичний тест залишається корисним.

Якщо преваленс $P(V_1)$ досліджуваного захворювання перевищує величину (5.59), тобто $P(V_1) > P_\omega(\delta)$, то стратегія Неймана-Пірсона (5.53), (5.54) є неефективною з погляду зменшення апіорного ризику.

Умова $P(V_1) > P_\omega(\delta)$ зумовлена тим, що величина $P_\omega(\delta)$ зменшується зі зростанням обмеження δ на відміну від функції $P_\omega(\varepsilon)$, яка зростає зі збільшенні обмеження ε (рис. 5.10)

На рис. 5.11 показано графіки залежності від $P_\omega(\delta)$ від δ , побудовані відповідно до (5.59) для різних значень ω .

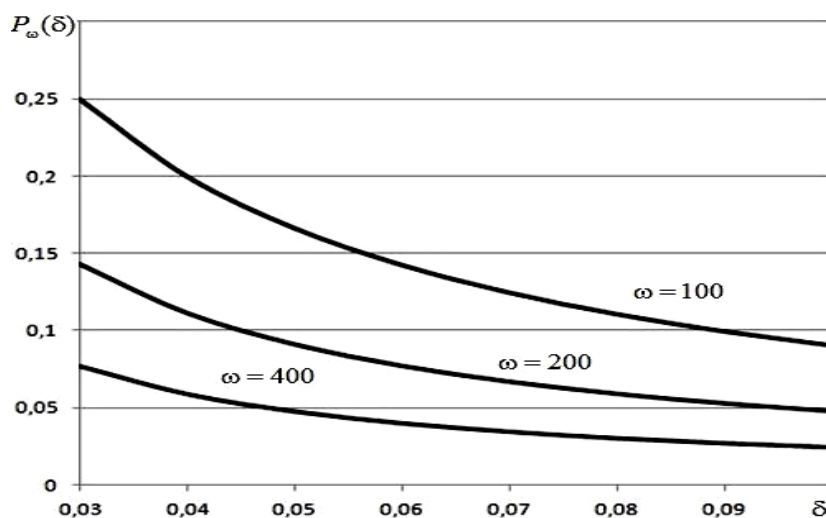


Рис. 5.11. Графіки залежності $P_\omega(\delta)$ від δ згідно з (5.59)

Приклад 5.2. Для обстеження групи 10000 пацієнтів передбачається використовувати діагностичний тест, що реалізує стратегію Неймана-Пірсона (5.50), (5.51). У зв'язку з високою загрозою захворювання співвідношення втрат обрано $\omega = 400$. Якщо встановити обмеження $\delta = 0,07$ та будемо вважати, що $P(E_1) = 0,07$. Звідси чутливість тесту становить $S_E = 93\%$. Припустимо, що тест має ідеально специфічність $S_p = 100\%$, тобто всі зорові особи належать до групи здорових.

З виразу (5.56) випливає, що за умов $\varepsilon = 0,07$ і $\omega = 400$ гранична величина преваленсу становить $P_\omega(\delta) = 0,0345$. Якщо преваленс досліджуваного захворювання становитиме $P(V_1) = 0,04$, тобто $P(V_1) > P_\omega(\delta)$, то діагностичний тест з операційними характеристиками $S_E = 100\%$ і $S_p = 93\%$ не може бути корисним.

Справді, якщо $P(V_1) = 0,04$, то очікувана кількість хворих у групі 10000 пацієнтів становитиме 400 осіб, а здорових – 9600 осіб. Без втрати загальності будемо вважати, що величину $\omega = 400$ визначають такі абсолютні значення втрат $L_{12} = 400$ і $L_{21} = 1$.

Звідси випливає, що для мінімізації апіорного ризику всіх обстежуваних слід визнавати хворими, тому що за інших рішень ризик буде вищим. Тобто апіорний ризик становить

$$R_0 = (L_{21} \cdot 9600) / 10000 = (1 \cdot 9600) / 10000 = 0,96. \quad (5.60)$$

Якщо ж застосовувати діагностичний тест, то за специфічністю $S_p = 100\%$ усі 9600 здорових обстежуваних осіб будуть визнані здоровими.

Водночас за чутливістю $S_e = 93\%$ з 400 хворих пацієнтів лише 372 особи будуть визнані хворими, а решта 28 пацієнтів будуть помилково належати до групи здорових. Тобто в цьому випадку апіорні втрати становитимуть

$$R = (L_{12} \cdot 28) / 10000 = (400 \cdot 28) / 10000 = 1,12. \quad (5.61)$$

З порівняння (5.60) і (5.61) випливає, що $R > R_0$ і такий тест безумовно є некорисним.

Змінимо в умовах прикладу величину преваленсу, припускаючи, що $P(V_1) = 0,03$, тобто $P(V_1) < P_{\omega}(\delta) = 0,0345$. Неважко переконатися, що в такому випадку

$$R_0 = (L_{21} \cdot 9700) / 10000 = (1 \cdot 9700) / 10000 = 0,97, \quad (5.62)$$

$$R = (L_{12} \cdot 21) / 10000 = (400 \cdot 21) / 10000 = 0,84, \quad (5.63)$$

тобто за таких умов діагностичний тест є корисним, тому що виконується строга нерівність $R < R_0$.

Часто діагностика захворювань ґрунтується на аналізі симптомів, наявність чи відсутність яких дозволяє приймати рішення щодо належності пацієнта класу хворих V_1 або здорових V_2 . Розглянемо найпростішу схему прийняття рішення за симптомом $x \in \{x^+, x^-\}$:

$$\begin{aligned} &\text{пацієнт хворий, якщо } x = x^+, \\ &\text{пацієнт здоровий, якщо } x = x^-, \end{aligned} \quad (5.64)$$

де x^+ , x^- – відповідно наявність або відсутність симптому x у конкретного пацієнта.

Здавалося б, якщо за даними медичної статистики відомо, що симптом x характерний для більшості хворих пацієнтів і не характерний для більшості здорових людей, то правило (5.64) забезпечує високу ефективність діагностичних рішень, допускаючи малі значення ймовірностей пропуску цілі $P(E_1)$ та хибної тривоги $P(E_2)$.

Насправді ж це не завжди так! У загальному випадку інформацію про навколишній світ, зокрема інформацію про симптом можна подати у вигляді поля

$$x = x(z, t) \quad (5.65)$$

у точці z простору на момент часу t . З (5.65) випливає, що рішення за правилом (5.30) зводиться до визначення наявності x^+ або відсутності x^- симптому x у деякій області Z тіла конкретного пацієнта для моментів часу $t \in [T_1, T_2]$ в інтервалі спостереження.

Правило (5.64) діагностики конкретного пацієнта з урахуванням (5.65) можна записати так:

$$\text{пацієнт хворий, якщо } \exists z, t : z \in Z, t \in [T_1, T_2], x(z, t) = x^+, \quad (5.66)$$

$$\text{пацієнт здоровий, якщо } \forall z, t : z \in Z, t \in [T_1, T_2], x(z, t) = x^-. \quad (5.67)$$

Розглянемо особливості прийняття таких рішень на прикладі аналізу симптому «Характерні шкірні висипання», який є індикатором багатьох захворювань.

Згідно з (5.66) для прийняття рішення на користь класу V_1 (пацієнт хворий) достатньо в якийсь момент часу $t \in [T_1, T_2]$ виявити цей симптом на локальній ділянці $z_0 \in Z$ тіла пацієнта. Якщо можна припустити, що для конкретного хворого пацієнта ймовірність випадкової події $x(z_0, t_0) = x^+$ висока, наприклад, $P(x^+ | V_1) = 0,99$, то з цього випливає, що ймовірність пропуску цілі становить $P(E_1) = 0,01$.

Водночас згідно з (5.67) достовірність прийняття протилежного рішення на користь класу V_2 (пацієнт здоровий) характеризує ймовірність складної події

$$A = \bigcup_{i,j} [x(z_i, t_j) = x^-], \quad (5.68)$$

яка полягає в тому, що в кожний момент спостереження $t_i \in [T_1, T_2]$ симптом не виявлений на всіх локальних ділянках z_i області Z .

Допускаючи незалежність елементарних подій, що утворюють складну подію (5.68), можна оцінити її ймовірність величиною

$$P(A) = \prod_{i,j} P[x(z_i, t_j) = x^-]. \quad (5.69)$$

Якщо в деякій точці z_i тіла здорової людини дуже висока ймовірність відсутності симптому x у фіксований момент часу t_j , з цього не впливає висока ймовірність складної події $P(A)$. Навіть якщо $P[x(z_i, t_j) = x^-] = 0,99$, то згідно з (5.69) маємо $P(A) \ll 0,99$ і для ймовірності хибної тривоги може виконуватись умова $P(E_2) \gg 0,01$.

Запитання для самоконтролю

1. Назвіть відомі вам формальні стратегії прийняття рішень.
2. У чому полягає байєсівська стратегія прийняття рішень?
3. Що таке ризик помилкової класифікації?
4. Що таке помилки пропуску цілі та хибної тривоги?
5. У чому полягає стратегія Неймана-Пірсона?
6. У чому полягає послідовний аналіз Вальда?
7. Дайте визначення чутливості та специфічності бінарного класифікатора.
8. За яких умов бінарний класифікатор можна вважати некорисним?
9. Для чого застосовують традиційний ROC-аналіз.
10. У чому полягає відмінність вдосконаленого ROC-аналізу від традиційного.

РОЗДІЛ 6

МАШИННЕ НАВЧАННЯ

6.1. Змістовна та формальна постановки задачі

Машинним навчанням називають побудову алгоритму класифікації, в окремому випадку, побудову поверхні в деякому просторі, що розділяє класи, які треба розпізнавати. Навчання здійснюється на основі прецедентів – прикладів окремих об'єктів із зазначенням їх приналежності до того чи іншого класу.

Мета навчання – забезпечити здатність технічної системи однаково реагувати на всі об'єкти одного класу і по-різному реагувати на об'єкти різних класів. Зазвичай процес навчання має завершитися на основі показу кінцевої кількості об'єктів різних класів з вказівками «вчителя».

Розглянемо приклад типової задачі навчання – навчитися класифікувати три типи квітки іриса (сформульована Фішером у 1936 році): щетинистий (***Iris setosa***), віргінський (***Iris virginica***) та різнокольоровий (***Iris versicolor***) (рис. 6.1).



Рис. 6.1. Три типи квіток ірису

Ґрунтуючись на багаторічному досвіді ботаніків можна припустити, що прийнятними характеристиками для класифікації квітів є такі чотири ознаки (рис. 6.2):

- x_1 – довжина чашолистки;

- x_2 – ширина чашолистки;
- x_3 – довжина пелюстки;
- x_4 – ширина пелюстки.



Рис. 6.2. Ознаки – довжина та ширина чашолистки та пелюстки

Припустимо, що ми маємо кінцеву навчальну вибірку реальних спостережень у вигляді таблиці зі значення ознак x_1, x_2, x_3, x_4 (табл. 6.1).

Таблиця 6.1. Фрагмент навчальної вибірки квітів ірису

x_1	x_2	x_3	x_4	Тип ірису
Чашолистик		Пелюстка		
Довжина	Ширина	Довжина	Ширина	
7,7	3,8	6,7	2,2	Віргінський
4,9	3,0	1,4	0,2	Щетинистий
5,9	3,0	4,2	1,5	Різнокольоровий
6,0	2,2	4,0	1,0	Різнокольоровий
6,3	3,3	6,0	2,5	Віргінський
5,2	3,5	1,5	0,2	Щетинистий
5,7	2,9	4,2	1,3	Різнокольоровий
...

Результати спостережень, наведені в таблиці 6.1, можна візуалізувати так, як показано на рис. 6.3. Для наочності точки, що відповідають конкретному типу квітки, доцільно пофарбувати різними ко-

льорами, і тоді на рисунку буде видно, як окремі класи групуються у просторі ознак.

Якість навчання суттєво залежить від кількості спостережень. Наприклад, для прийнятної класифікації квітів ірису застосовували 150 спостережень, по 50 спостережень для кожного типу.

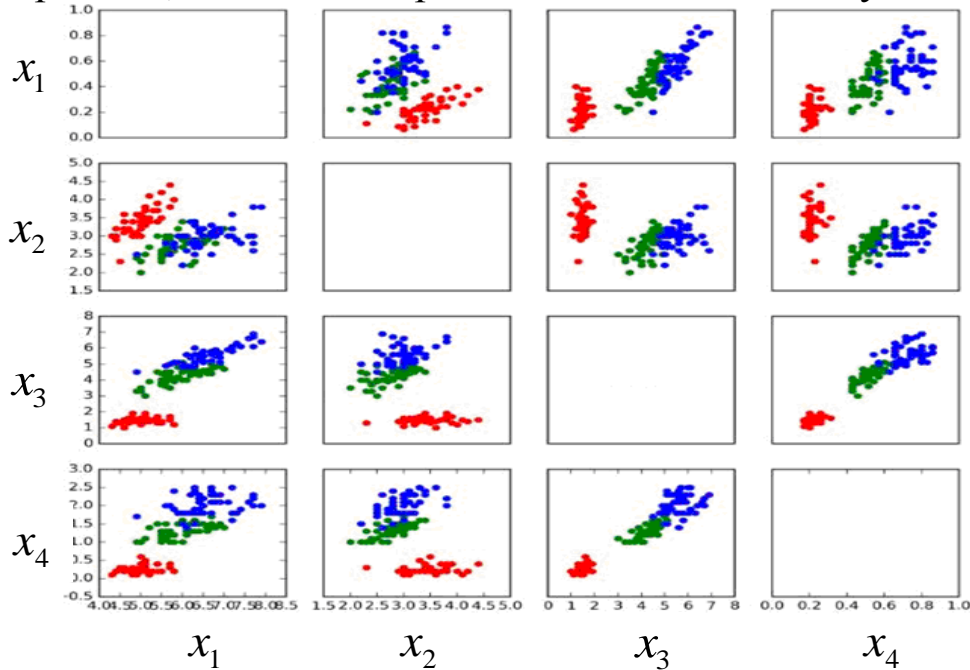


Рис. 6.3. Візуалізація даних навчальної вибірки

Формально задачу машинного навчання можна сформулювати так. Потрібно за навчальною вибіркою побудувати правило класифікації

$$D = D(x, w), \quad (6.1)$$

де $D: X \rightarrow Y$ фіксована функція, задана з точністю до невідомого вектора параметрів $w \in W$, яка відображає множину спостережень X (множину ознак) на множину рішень Y – міток класів.

6.2. Лінійні класифікатори

Найпростіше правило класифікації (6.1) – лінійне правило з вектором параметрів

$$w = (w_1, \dots, w_N).$$

Для двох класів V_1 і V_2 таке правило має вигляд:

$$D(x, w) = \text{sign} \sum_{j=1}^N w_j f_j(x), \quad (6.2)$$

тобто

$$\begin{aligned} &\text{рішення на користь } V_1, \text{ якщо } D(x, w) > 0 \\ &\text{рішення на користь } V_2, \text{ якщо } D(x, w) < 0 \end{aligned} \quad (6.3)$$

Навчальна вибірка – це сукупність K спостережень $x_i^{(N)} = (x_{i1}, \dots, x_{iN})$ з відомою належністю до класів

$$(x_1^{(N)}, U_1), (x_2^{(N)}, U_2), (x_3^{(N)}, U_3), \dots, (x_K^{(N)}, U_K),$$

де U_j – вказівки «вчителя», які задані в формі

$$U_j = \begin{cases} 1, & \text{якщо } x_j^{(N)} \text{ об'єкт класу } V_1 \\ 2, & \text{якщо } x_j^{(N)} \text{ об'єкт класу } V_2 \end{cases}, \quad j = 1, \dots, K, \quad (6.4)$$

За умови, що число K спостережень мале, навчання може бути неефективним. Однак ефективність навчання залежить не лише від кількості спостережень.

Відомо жарт, коли геолог приніс до спеціаліста всього зразка каменів і попросив навчити систему їх класифікувати. Коли ж від дізнався, що два зразка замало, він взяв молоток та розбив камені на велику кількість маленьких зразків. Зрозуміло, що не буде поліпшення якості навчання від такого підвищення кількості спостережень.

Слід розрізняти ситуації, коли в у просторі ознак спостереження класів V_1 і V_2 можна розділити лінійною функцією

$$D(x^{(N)}) = \sum_{i=1}^N w_i x_i + w_0, \quad (6.5)$$

і коли такої можливості немає (рис. 6.4).

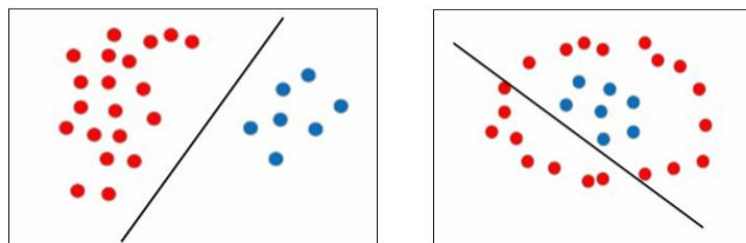


Рис. 6.4. Спостереження, які можливо (ліворуч) і не можливо (праворуч) розділити лінійною функцією (6.5)

Щоб формалізувати умови лінійного розділення точок спостережень навчальної вибірки, нагадаємо такі означення.

Означення 6.1. Опуклою називають множину X , яка включає всі точки відрізка, що з'єднує будь-які дві внутрішні точки $x \in X$ і $y \in Y$ (рис. 6.5).

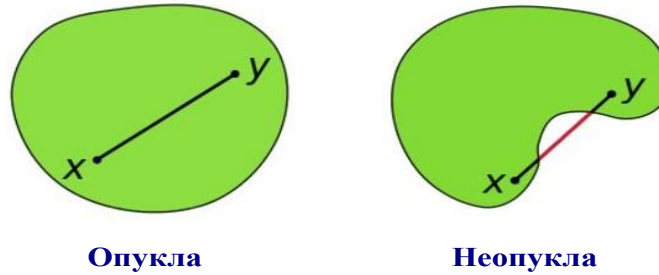


Рис. 6.5. Опукла (ліворуч) та неопукла (праворуч) множини

Означення 6.2. Опукла оболонка $Co(X)$ множини X – це мінімальна опукла множина (багатогранник), який включає множину X (рис. 6.6.).



Рис. 6.6. Опукла множина і опукла оболонка множини

Означення 6.3. Множина X є опуклою, якщо $X = Co(X)$.

Твердження 6.1. Множини спостережень X^+ і X^- навчальної вибірки лінійно роздільні, якщо

$$Co(X^+) \cap Co(X^-) = \emptyset.$$

Є багато способів провести пряму між лінійно роздільними спостереженнями навчальної вибірки, наприклад (рис. 6.7):

- будують найширшу смугу між спостереженнями класів V_1 і V_2 , центр якою визначає дискримінантну пряму (метод опорних векторів);
- максимізують суму відступів між спостереженнями класів V_1 і V_2 .

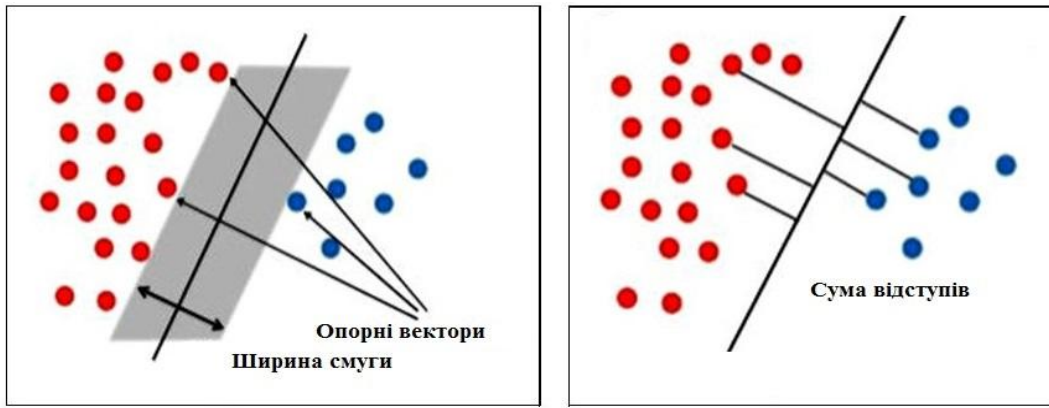


Рис. 6.7. Ідеї, які покладені в основу вибору найкращої прямої

Під час розв'язуванні практичних задач часто застосовують перехід за допомогою деякого перетворення

$$\Psi = X \rightarrow \tilde{X},$$

від початкового простору ознак X , в якому спостереження навчальної вибірки лінійно нерозділені, до нового простору \tilde{X} в якому спостереження можуть бути лінійно розділені. Такий простір \tilde{X} називають спрямовуючим простором.

У загальному випадку методологія машинного навчання за прецедентами передбачає такі етапи:

Етап 1 – навчання (**Training**), коли за кінцевою навчальною вибіркою x_k , $k = 1, \dots, K$ та відповідними вказівками «вчителя» y_k , $k = 1, \dots, K$ будують дискримінантну функцію $D(x)$, яка дає змогу класифікувати об'єкти:

$$\begin{pmatrix} f_1(x_1), & \dots, & f_N(x_1) \\ \dots & \dots & \dots \\ f_1(x_K), & \dots, & f_N(x_K) \end{pmatrix}, \begin{pmatrix} y_1 \\ \dots \\ y_K \end{pmatrix} \rightarrow D(x).$$

Етап 2 – перевірка (**Testing**), коли за незалежною вибіркою спостережень x_k^* , $k = 1, \dots, K$ оцінюють функцію втрат $L(\cdot)$, що характеризує ефективність алгоритму (помилки класифікації):

$$\begin{pmatrix} f_1(x_1^*), & \dots, & f_N(x_1^*) \\ \dots & \dots & \dots \\ f_1(x_K^*), & \dots, & f_N(x_K^*) \end{pmatrix} \rightarrow \begin{pmatrix} D(x_1^*) \\ \dots \\ D(x_K^*) \end{pmatrix} \rightarrow L(\cdot).$$

Функція втрат (**Loss function**) має характеризувати помилку класифікації спостереження x^* , наприклад так:

$$L(D, x^*) = \begin{cases} 0, & \text{если } D(x^*) = y(x^*), \\ 1, & \text{если } D(x^*) \neq y(x^*), \end{cases} \quad (6.6)$$

де $y(x^*)$ – правильна відповідь для спостереження x^* .

За функцією втрат $L(D, x_i^*)$ можна визначити емпіричний ризик – середні втрати на вибірці об'ємом K :

$$R = \frac{1}{K} \sum_{i=1}^K L(D, x_i^*). \quad (6.7)$$

З погляду мінімізації середніх втрат задача класифікації має принципові відмінності від задачі побудови лінії регресії (рис.6.8).

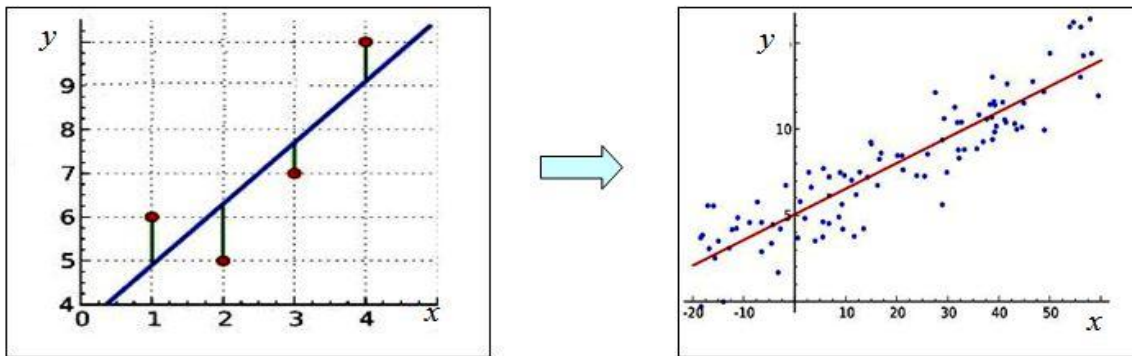


Рис. 6.8. Ілюстрація методу побудови лінії регресії

У задачі побудови лінії регресії функція втрат визначається квадратичною помилкою – квадратом відхилення значення $D(x)$, обчисленого в точці x , від значення $y(x)$, що спостерігається в цій точці, тобто

$$L(D, x) = (D(x) - y)^2. \quad (6.8)$$

За такою функцією втрат середній ризик

$$R(w) = \frac{1}{K} \sum_{i=1}^K L(D(x), x_i) = \frac{1}{K} \sum_{i=1}^K (y_i - w_1 x_i - w_0)^2 \quad (6.9)$$

є опуклою функцією.

Тому для знаходження оптимального значення вектора параметрів $w = (w_1, w_0)$ можна використовувати аналітичний метод – метод найменших квадратів (МНК):

$$w_{\text{опт}} = \arg \min_{w \in W} R(w),$$

тобто визначати $w_{\text{опт}}$ на основі розв'язування системи рівнянь, побудованої на умовах

$$\frac{\partial R(w)}{\partial w_1} = 0, \dots, \frac{\partial R(w)}{\partial w_0} = 0.$$

У задачі класифікації середній ризик (6.7), в якому є функція втрат (6.6), не є опуклою функцією.

Тому, на відміну від задачі пошуку оптимальних параметрів лінії регресії, в цьому випадку застосовують не аналітичний, а алгоритмічний метод оптимізації.

Один з таких методів – метод градієнтного спуску, суть якого показано на рис. 6.9.

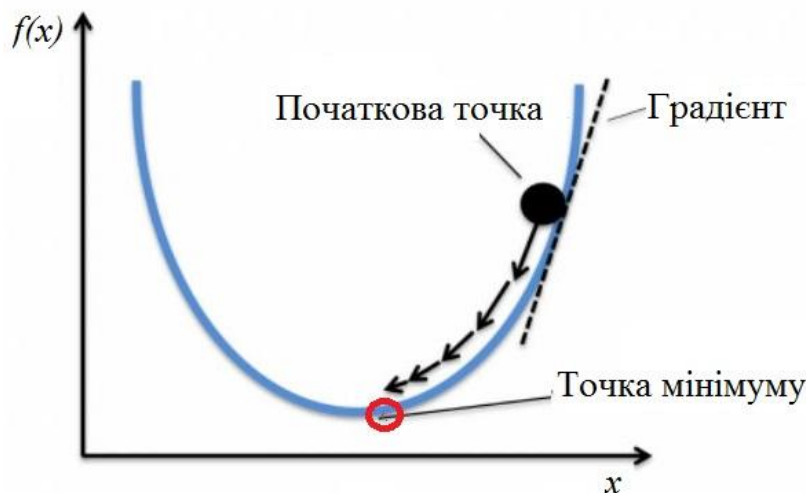


Рис. 6.9. Візуалізація методу градієнтного спуску

Градієнтний спуск (**Gradient descent**) – це ітераційний алгоритм, в якому для знаходження локального мінімуму функції здійснюються кроки, пропорційні протилежному значенню градієнта (або наближеного градієнта) функції в поточній точці.

Якщо спадна функція має мінімум, то під час переміщення від довільної точки у напрямку, протилежному значенню похідної, ми зменшуємо значення функції:

$$x_{i+1} = x_i - \gamma f'(x_i). \quad (6.10)$$

Зауважимо, що для визначення, в який бік рухатися, потрібен лише знак похідної $f'(x_i)$.

Програмна реалізація градієнтного методу мінімізації функції $f(x)$ однієї змінної x має вигляд:

- 1) ініціалізувати початкове x_i випадковим значенням;
- 2) $i := 1$;
- 3) $x_{i+1} := x_i - \gamma f'(x_i)$;
- 4) $i++$;
- 5) if $f(x_i) - f(x_{i+1}) > C$ goto 3.

Константа γ визначає крок спуску, а константа C визначає наскільки близько ми хочемо наблизитися до ідеального розв'язку.

В основу методів машинного навчання покладено гіпотезу компактності, яка полягає в тому, що об'єкти одного класу перебувають у компактній області простору ознак. Такій гіпотезі відповідає розглянута задача класифікації квітів ірису (рис. 6.3).

Близькість об'єктів оцінюють відстанню (метрикою), наприклад евклідовою відстанню.

Для довільного $u \in X$ відсортуємо об'єкти:

$$\rho(u, x_u^{(1)}) \leq \rho(u, x_u^{(2)}) \leq \dots \leq \rho(u, x_u^{(K)}), \quad (6.11)$$

де $\rho(u, x_u^{(i)})$ – відстань об'єкта $u \in X$ до i -го сусіда серед об'єктів x_1, \dots, x_K .

На підставі аналізу відстаней (6.11) реалізують такі методи:

- найближчого сусіда: довільному об'єкту $u \in X$ приписують клас об'єкта $x_u^{(1)}$;
- найближчих сусідів: об'єкту $u \in X$ приписують той клас, до якого належить більшість об'єктів

$$\rho(u, x_u^{(1)}) \leq \rho(u, x_u^{(2)}) \leq \dots \leq \rho(u, x_u^{(n)});$$

- класифікація за відстанню до еталонів

$$X_{01} = (x_{11}^0, x_{12}^0, \dots, x_{1N}^0) \quad \text{та} \quad X_{02} = (x_{21}^0, x_{22}^0, \dots, x_{2N}^0) \quad (6.12)$$

класів V_1 і V_2 через усереднення відповідних векторів вибірки

$$x_{mi}^0 = \frac{1}{K_m} \sum_{i=1}^{K_m} x_{mi} \quad i = 1, 2, \dots, N, \quad m = 1, 2, \quad (6.13)$$

де K_m – кількість об'єктів m -го класу в навчальній вибірці, а i – номер ознаки.

Зауважимо, що еталони (6.12) – це абстрактні класи, які можуть не збігатися з жодним об'єктом навчальної вибірки і навіть жодним з об'єктів генеральної сукупності.

Для нового об'єкта, який не входив у навчальну вибірку, рішення приймається за мінімумом відстаней до еталонів X_{01} і X_{02} (рис. 6.10).

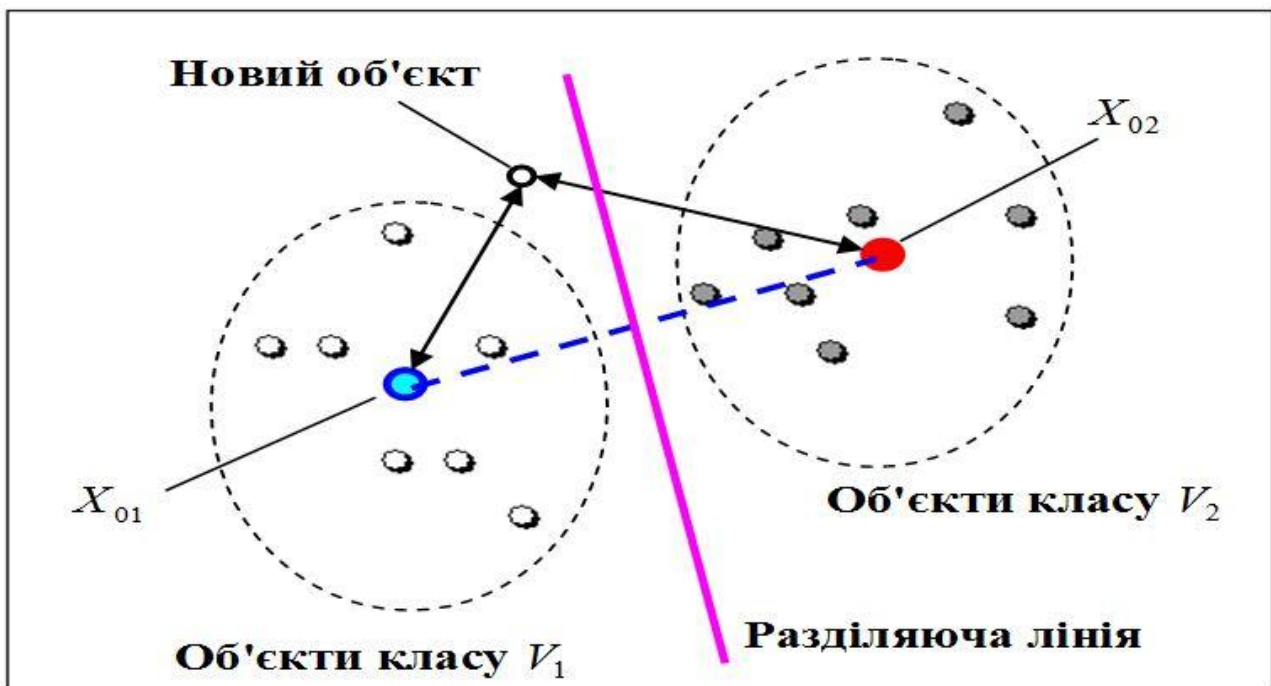


Рис. 6.10. Класифікація об'єктів за відстанню до еталонів X_{01} і X_{02}

Такий класифікатор є лінійним класифікатором, оскільки за умови двох ознаках геометричне місце точок на площині, рівновіддалених від X_{01} і X_{02} , є пряма лінія, яка під прямим кутом проходить через середину відрізка, що з'єднує X_{01} і X_{02} .

Покажемо, що коли власні області $X_1^{(N)}$ і $X_2^{(N)}$, на яких зосереджені умовні розподіли класів, не є опуклими множинами, то лінійний класифікатор може призводити до абсурдних результатів навіть коли виконується умова $X_1^{(N)} \cap X_2^{(N)} = \emptyset$.

Жартівливий приклад. Нехай для класифікації якісних (клас V_1) та неякісних (клас V_2) телевізорів пропонується використовувати дві нормовані ознаки: якість звуку $x_1 \in [0,1]$ і якість зображення $x_2 \in [0,1]$.

Оптимальними рішеннями може слугувати таке правило:

$$\begin{aligned} &\text{клас } V_1, \text{ якщо } (x_1 > 0,5) \wedge (x_2 > 0,5); \\ &\text{клас } V_2 \text{ в протилежному випадку.} \end{aligned} \quad (6.14)$$

Будемо вважати, що на етапі «навчання» точки якісних і неякісних телевізорів рівномірно заповнили відповідні до (6.14) власні області класів у просторі ознак.

Згідно з (6.13) X_{01} та X_{02} будуть визначати центри мас вказаних областей (рис. 6.11):

$$X_{01} = (0,75 ; 0,75) \text{ і } X_{02} \approx (0,417 ; 0,417).$$

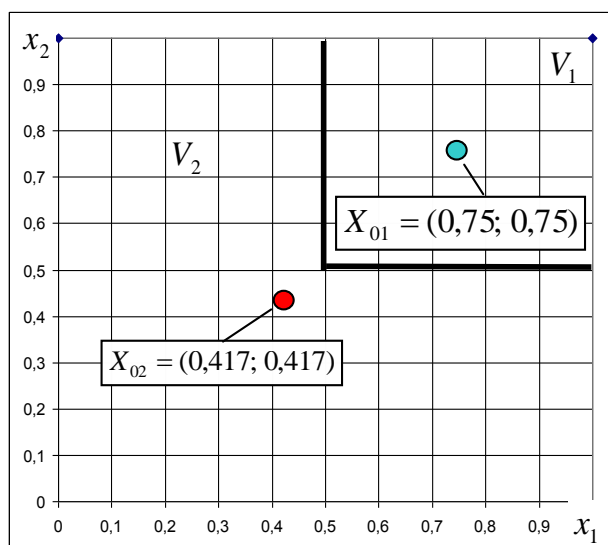


Рис. 6.11. Еталони якісних та неякісних телевізорів у просторі ознак

Геометричне місце точок, рівновіддалених від X_{01} і X_{02} , визначає пряма β , яка під прямим кутом проходить через середину відрізка α , що з'єднує точки X_{01} і X_{02} . За наявними даними рівняння прямої β має вигляд $x_2 + x_1 - 1,17 = 0$. Така пряма розділяє простір ознак x_1 і x_2 на дві області Ω_1 і Ω_2 рішень відносно класів V_1 і V_2 (рис. 6.12):

$$\begin{aligned} &V_1, \text{ якщо } x=(x_1, x_2) \in \Omega_1 \\ &V_2, \text{ якщо } x=(x_1, x_2) \in \Omega_2 \end{aligned}$$

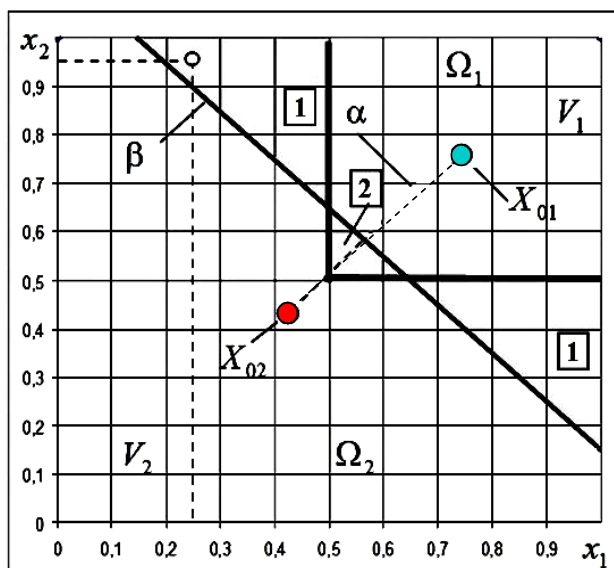


Рис. 6.12. Побудовані області Ω_1 і Ω_2 рішень відносно класів V_1 і V_2

Можна бачити, що в областях 1 неякісні телевізори будуть вважатися якісними, а в області 2 якісні телевізори будуть помилково належати до неякісних.

Наприклад, телевізор з огидним звуком $x_1 = 0,25$ буде визнаний якісним тільки тому, що він має гарне зображення $x_2 = 0,95$.

Оцінимо чутливість S_E і специфічність S_P отриманих рішень

$$S_E = \frac{W_B - W_1}{W_B}; \quad S_P = \frac{W_G - W_2}{W_G}, \quad (6.15)$$

де W_G, W_B , – площі власних областей якісних (клас V_1) і неякісних (клас V_2) телевізорів згідно з (6.14), а W_1, W_2 , – площі областей 1 і 2 відповідно.

Відповідно до наявних даних маємо:

$$S_E \approx \frac{0,75 - 0,1225}{0,75} \approx 0,837; \quad S_P \approx \frac{0,25 - 0,00125}{0,25} \approx 0,995.$$

Таким чином, правило мінімуму евклідових відстаней до побудованих еталонів X_{01} і X_{02} майже у 20 % випадків призводить до помилкових результатів.

Зауважимо, що згідно з класичним дискримінантним аналізом у випадках, коли власні області класів перетинаються, тобто

$$X_1 \cap X_2 \neq \emptyset,$$

застосування лінійної дискримінантної функції обґрунтовано лише за виконання двох важливих обмежень:

- умовні розподіли $p(x^{(N)} | V_m)$, $m = 1, 2$ – багатовимірні нормальні розподіли, що потребує перевірки за критерієм Шапіро-Вілка;
- рівні коваріаційні матриці цих розподілів (властивість гомескедастичності), що вимагає перевірки за критерієм Бокса.

6.3. Одношаровий персептрон

Сформулюємо ще одне твердження, яке визначає умови лінійної роздільності двох множин X^+ і X^-

Нехай $x^{(N)} = (x_1, \dots, x_N)$ – N -вимірний вектор ознак. Для зручності запису введемо розширений вектор

$$x = (x^{(N)}, 1) \triangleq (x_1, \dots, x_N, 1).$$

Тоді рівняння лінійної дискримінантної функції, яка розділяє класи V_1 і V_2 у N -вимірному просторі, можна записати в компактному вигляді

$$D(w, x) = \langle w, x \rangle \triangleq \sum_{i=0}^N w_i x_i, \quad (6.16)$$

де $w = (w_0, \dots, w_N)$ – вектор параметрів (вагових коефіцієнтів), а $\langle w, x \rangle$ означає скалярний добуток векторів w і x .

Твердження 6.2. Множини спостережень X^+ і X^- навчальної вибірки лінійно роздільні, якщо є вектор параметрів

$$w = (w_0, \dots, w_N),$$

що є розв'язком системи лінійних нерівностей

$$\begin{aligned} \langle w, x[j] \rangle &> 0, & \text{якщо } x^{(N)}[j] \text{ – об'єкт класу } V_1, \\ \langle w, x[j] \rangle &< 0, & \text{якщо } x^{(N)}[j] \text{ – об'єкт класу } V_2, \end{aligned} \quad \forall j = 1, 2, \dots, K. \quad (6.17)$$

Один з методів побудови лінійного вирішувального правила запропонував Френк Розенблат у технічному пристрої персептрон (рис.

6.13), який моделює нервову клітину (модель нейрона), а сукупність перцептронів утворює нейронні мережі.

Перцептрон з функцією активації $f = \text{sign } D(x)$ реалізує модель

$$D_{12}(w, x) \triangleq \text{sign} \sum_{i=0}^N w_i x_i = \text{sign}(w_0 + w_1 x_1 + \dots + w_N x_N), \quad (6.18)$$

яка дає змогу розділити простір ознак на дві області рішень: Ω_1 на користь класу V_1 і Ω_2 на користь класу V_2

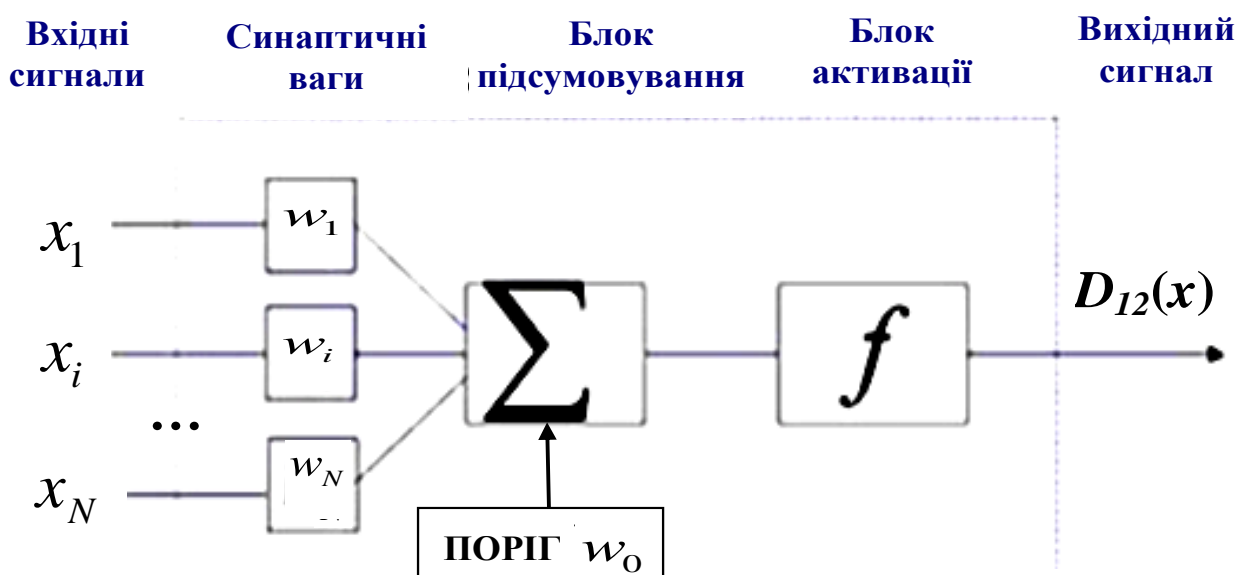
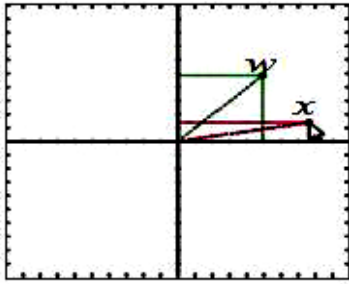


Рис. 6.13. Структура одношарового перцептрона

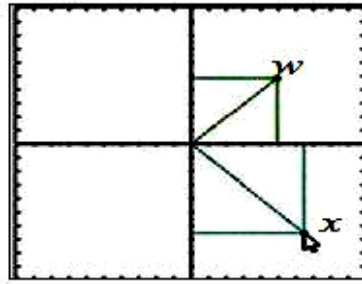
Розглянемо частковий випадок двовходового перцептрону з функцією активації, який реалізує модель

$$D_{12}(w, x) = \text{sign}(w_1 x_1 + w_2 x_2) = \text{sign} \langle w, x \rangle, \quad (6.19)$$

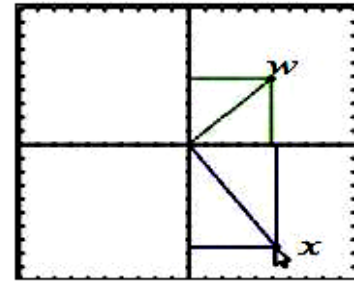
де $\langle w, x \rangle$ – скалярний добуток вектора ваг $w = (w_1, w_2)$ та вектора сигналів (ознак) $x = (x_1, x_2)$. На рис. 6.14 показано, як змінюється знак функції $D_{12}(w, x)$ під час зміни сигналів x_1, x_2 за сталих значень параметрів w_1, w_2 .



$$\begin{aligned} w &= (4,9; 4,8) \\ x &= (7,6; 1,4) \\ \langle w, x \rangle &= 43,96 \\ D_{12}(w, x) &> 0 \end{aligned}$$



$$\begin{aligned} w &= (4,9; 4,8) \\ x &= (6,5; -6,5) \\ \langle w, x \rangle &= 0,65 \\ D_{12}(w, x) &\approx 0 \end{aligned}$$



$$\begin{aligned} w &= (4,9; 4,8) \\ x &= (5,2; -7,5) \\ \langle w, x \rangle &= -10,52 \\ D_{12}(w, x) &< 0 \end{aligned}$$

Рис. 6.14. Залежність знака функції $D_{12}(w, x)$ від параметрів w_1, w_2 .

Структуру персептрона можна інтерпретувати таким чином. Будемо умовно вважати, що

- кожний i -й вхід персептрона – радник;
- вага w_i – ступінь довіри до i -го радника;
- $w_i = 1$ – висока довіра до i -го радника;
- $w_i = 0,5$ – часткова довіра до i -го радника;
- $w_i = -1$ – висока недовіра до i -го радника;
- $|w_i|$ характеризує важливість i -го радника.

За таких умов вихід персептрона можна трактувати як результат колективної думки радників, узгоджений за методом голосування.

Можна інтерпретувати персептрон по-іншому. Припустимо, що на вхід персептрона надходять дві ознаки x_1 і x_2 , які набувають тільки два можливі значення:

$$x_1 \in \{-1, +1\} \text{ та } x_2 \in \{-1, +1\}.$$

Будемо вважати, що такі значення можна інтерпретувати так:

$$\begin{aligned} \text{ІСТИНА,} & \quad \text{якщо } x = +1, \\ \text{ХИБНІСТЬ,} & \quad \text{якщо } x = -1. \end{aligned}$$

Зауважимо, що в булевій алгебрі «ХИБНІСТЬ» ототожнюють зі значенням $x = 0$.

Будемо також вважати, що ваги набувають двох значень $w_i \in \{-1, +1\}$:

$$\begin{aligned} &\text{ЗБУДЖЕННЯ, якщо } w_i = +1, \\ &\text{ГАЛЬМУВАННЯ, якщо } w_i = -1. \end{aligned} \quad (6.20)$$

За допомогою персептрона можна реалізувати логічні функції. На рис. 6.15 показано схему реалізації кон'юнкції (логічного «І»):

$$\text{sign}(w_0 + w_1x_1 + w_2x_2) = x_1 \wedge x_2 \quad (6.21)$$

На схемі наведено таблицю істинності (рис. 6.15, а), система рівнянь, що відповідає цій таблиці (рис. 6.15, б), відповідний розв'язок системи нерівностей (рис. 6.15, г), а також графічна інтерпретація результату (рис. 6.15, г).

Аналогічно реалізується функція диз'юнкції (логічне «АБО»):

$$\text{sign}(w_0 + w_1x_1 + w_2x_2) = x_1 \vee x_2, \quad (6.22)$$

що зображено схемою, наведеною на рис. 6.16.

Однак одношаровий персептрон не здатний реалізувати функцію XOR («виключне АБО»):

$$\text{sign}(w_0 + w_1x_1 + w_2x_2) = x_1 \oplus x_2, \quad (6.23)$$

тому що система нерівностей

$$\begin{cases} w_0 - w_1 - w_2 < 0, \\ w_0 - w_1 + w_2 > 0, \\ w_0 + w_1 - w_2 > 0, \\ w_0 + w_1 + w_2 < 0, \end{cases} \quad (6.24)$$

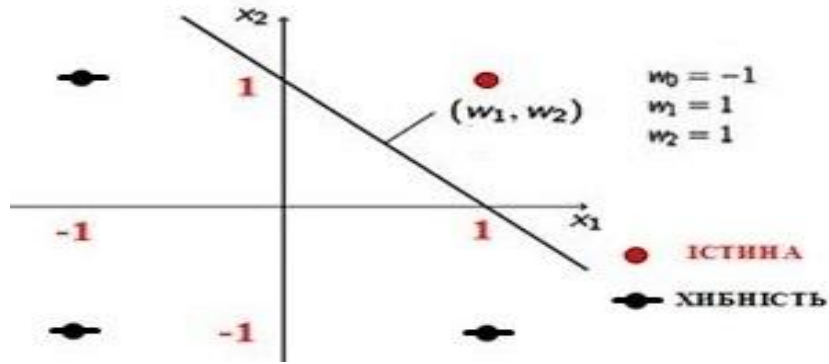
до якої зводиться функція XOR (6.23), не має розв'язку. Це показано на рис. 6.17, де подано геометричну інтерпретацію функції (6.23).

x_1	x_2	$x_1 \wedge x_2$
-1	-1	-1
-1	1	-1
1	-1	-1
1	1	1

а

б

в



г

Рис. 6.15. Реалізація кон'юнкції (логічне «І»)

x_1	x_2	$x_1 \vee x_2$
-1	-1	-1
-1	1	1
1	-1	1
1	1	1

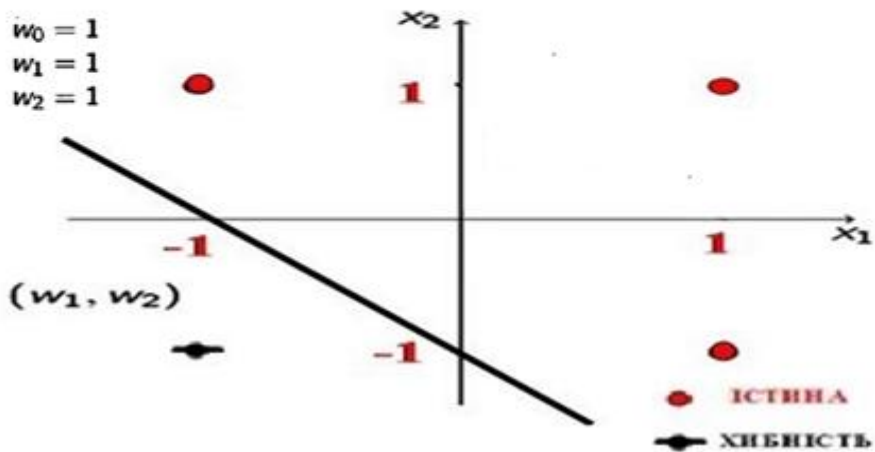


Рис. 6.16. Реалізація кон'юнкції (логічне «АБО»)

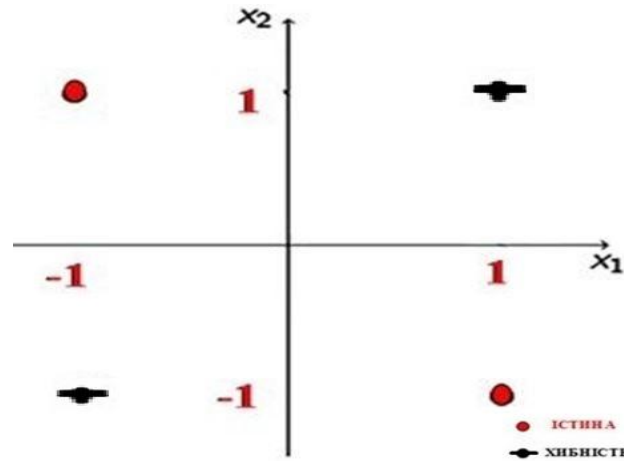


Рис. 6.17. Графічна інтерпретація функції XOR

Водночас функцію XOR можна реалізувати на трьох одношарових перцептронах за схемою, наведеною на рис. 6.18.

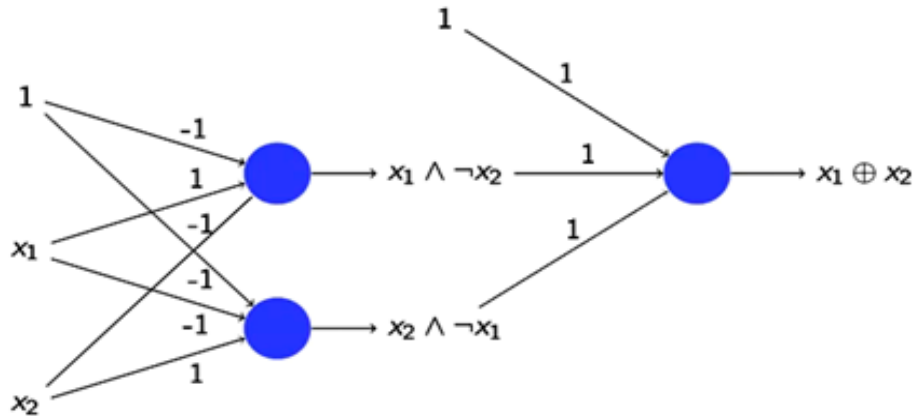


Рис. 6.18. Можливість реалізації функції XOR (6.23) на трьох зв'язаних одношарових перцептронах

6.4. Алгоритми навчання Розенблата і Козинця

Нехай у N -вимірному просторі ознак задано вибірку спостережень з відомою приналежністю до класів

$$X = \{(x_1^{(N)}, y_1), (x_2^{(N)}, y_2), \dots, (x_n^{(N)}, y_n)\}, \quad (6.25)$$

де n – кількість елементів у вибірці, $x_j^{(N)} \triangleq (x_1, \dots, x_N)$ – точки (N -вимірні вектора), а y_j – індикаторна змінна така, що

$$y_j = \begin{cases} +1, & \text{якщо } x_j^{(N)} \in V_1, \\ -1, & \text{якщо } x_j^{(N)} \in V_2. \end{cases} \quad j = 1, \dots, n. \quad (6.26)$$

Передбачається, що спостереження класів V_1 і V_2 можуть бути розділені лінійною дискримінантною функцією

$$D(x) = \langle w, x \rangle \stackrel{\Delta}{=} \sum_{i=0}^N w_i x_i, \quad (6.27)$$

де, як і раніше, запис $\langle w, x \rangle$ позначає скалярний добуток $(N+1)$ -вимірних векторів параметрів $w = (w_0, w_1, \dots, w_N)$ та розширених векторів $x = (1, x_1, \dots, x_N)$.

Задача навчання перцептрона полягає в тому, щоб за скінченною навчальною вибіркою (6.25) з відомими значеннями індикаторної змінної (6.26) визначити вектор параметрів $w = (w_0, w_1, \dots, w_N)$ дискримінантної функції (6.27), яка надасть змогу розділити спостереження вибірки.

Тобто, знайти такий вектор $w = (w_0, w_1, \dots, w_N)$, який надасть змогу реалізувати таке правило:

$$\text{рішення на користь } V_1, \text{ якщо } \langle w, x \rangle > 0, \quad (6.28)$$

$$\text{рішення на користь } V_2, \text{ якщо } \langle w, x \rangle < 0. \quad (6.29)$$

Розглянемо два альтернативні алгоритми розв'язування такої задачі: алгоритм Розенблата та алгоритм Козинця.

Алгоритм Розенблата передбачає виконання таких кроків:

Крок 1. Довільним чином обирають початкові значення вектора $w^{(0)}$. Наприклад, для двовимірного випадку $N = 2$ можна задати $w^{(0)} = (0, 0, 1)$.

Крок 2. Послідовно вибирають спостереження $x_\alpha^{(t)} = (x^{(N)}, 1)$, $\alpha = 1, \dots, n$ з навчальної вибірки (6.25) і відповідно до (6.27) визначають значення дискримінантної функції $D(w^{(t-1)}, x_\alpha^{(t)})$ для поточного значення вектора $w^{(t-1)}$, $t = 1, 2, \dots$.

Крок 3. Якщо виконується умова

$$D(w^{(t-1)}, x_\alpha^{(t)}) y_t > 0, \quad (6.30)$$

тобто знаки дискримінантної функції $D(w^{(t-1)}, x_\alpha^{(t)})$ і індикаторної змінної y_{t-1} збігаються, то корекція значення вектора $w^{(t-1)}$ не здійснюється. Якщо ж виконується умова

$$D(w^{(t-1)}, x_\alpha^{(t)})y_t < 0, \quad (6.31)$$

це свідчить, що спостерігається помилка для класифікації поточного значення $x_\alpha^{(t)}$.

Крок 4. За наявності помилки вектор параметрів модифікується:

$$w^{(t)} = \begin{cases} w^{(t-1)} + \gamma x_\alpha^{(t)}, & \text{якщо } D(w^{(t-1)}, x_\alpha^{(t)}) < 0 \text{ та } y_t = +1, \\ w^{(t-1)} - \gamma x_\alpha^{(t)}, & \text{якщо } D(w^{(t-1)}, x_\alpha^{(t)}) > 0 \text{ та } y_t = -1, \end{cases} \quad (6.32)$$

де $0 < \gamma < 1$ – константа, що визначає темп корекції.

5. Кроки 2–4 повторюються доки всі точки вибірки (6.25) будуть класифіковані правильно.

На рис. 6.19 показано принцип корекції параметрів дискримінантної функції на площині двох ознак x_1, x_2 .

Маємо поточне спостереження з розширеним вектором $x_\alpha^{(t)} = (1, 2,5, 1)$, для якого $y_t = +1$. Нехай вектор параметрів дискримінантної функції становить

$$w^{(t-1)} = (1, -2, 2). \quad (6.33)$$

Це свідчить про помилкову класифікацію $x_\alpha^{(t)}$, оскільки виконується умова (6.31).

Згідно з (6.32) у такому випадку, якщо $\gamma = 1$, корекція параметрів дискримінантної функції має вигляд:

$$w^{(t)} = w^{(t-1)} + x_\alpha^{(t)}, \quad (6.34)$$

тобто

$$w^{(t)} = (2, 0,5, 3). \quad (6.35)$$

У результаті такої корекції вже виконується умова (6.30) і помилка усувається.

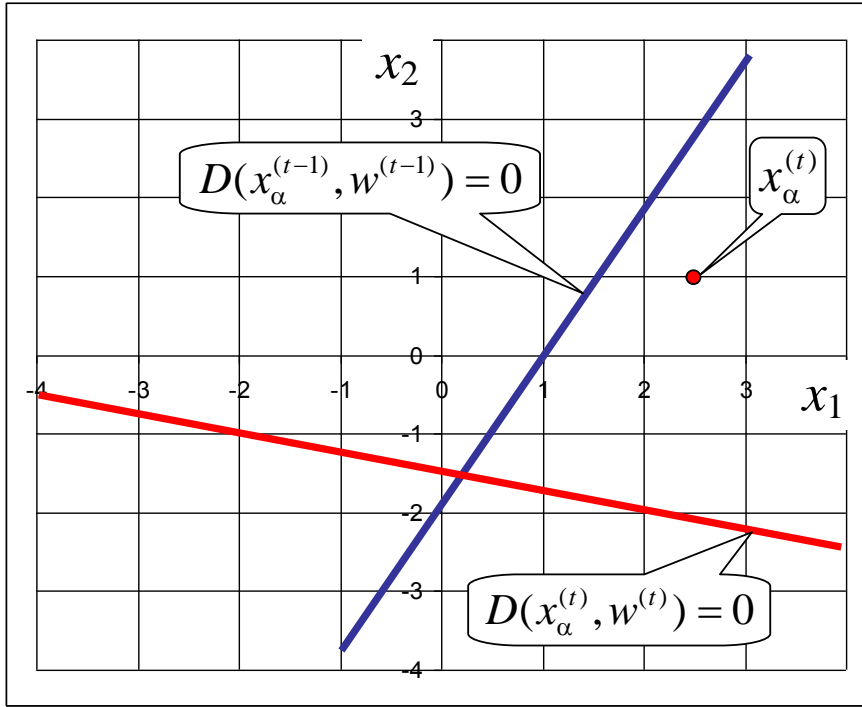


Рис. 6.19. Приклад корекції вектора параметрів $w^{(t)}$ за (6.34)

Теорема Новікова. Для кінцевої вибірки, що може бути лінійно розділена, алгоритм навчання Розенבלата збігається за кінцевої кількості ітерацій (кроків) t^0 , причому

$$t^0 \leq \frac{Q^2}{\varepsilon^2}, \quad (6.36)$$

де $Q = \max_{i \in [1, n]} |x_i^{(N)}|$, $\varepsilon = \min_{x^{(N)} \in \text{Co}(X)} |x^{(N)}| > 0$, $\text{Co}(X)$ – опукла оболонка множини X .

Розглянемо тепер алгоритм навчання Козинця.

Основна ідея алгоритму полягає в тому, що на кожному кроці ітерації $t = 1, 2, \dots$ проводиться пошук такого спостереження $x_\alpha^{(t)} = x^{(N)}$ навчальної вибірки (6.25), яке неправильно класифікується за поточним значенням $w^{(t-1)}$ вектора параметрів дискримінантної функції.

Якщо таких векторів немає для всіх точок навчальної вибірки, алгоритм завершує свою роботу. Якщо ж знайдено вектор $x_\alpha^{(t)}$, який неправильно класифікується, то проводиться корекція вектора параметрів таким чином:

$$w^{(t)} = (1 - \gamma^{(t)}) \cdot w^{(t-1)} + \gamma^{(t)} \cdot x_\alpha^{(t)}, \quad (6.37)$$

де

$$\gamma = \arg \min \left| (1 - \gamma^{(t)}) \cdot w^{(t-1)} + \gamma^{(t)} \cdot x_{\alpha}^{(t)} \right|. \quad (6.38)$$

Алгоритм Козинця зводиться до таких кроків.

Крок 1. Випадково вибираються дві точки навчальної вибірки, що належать різним класам, між яким проводиться пряма AB . Параметри серединного перпендикуляра W_0 до відрізка AB визначають початкове наближення вектора параметрів $w^{(0)}$ дискримінантної функції, що шукається (рис. 6.20, а).

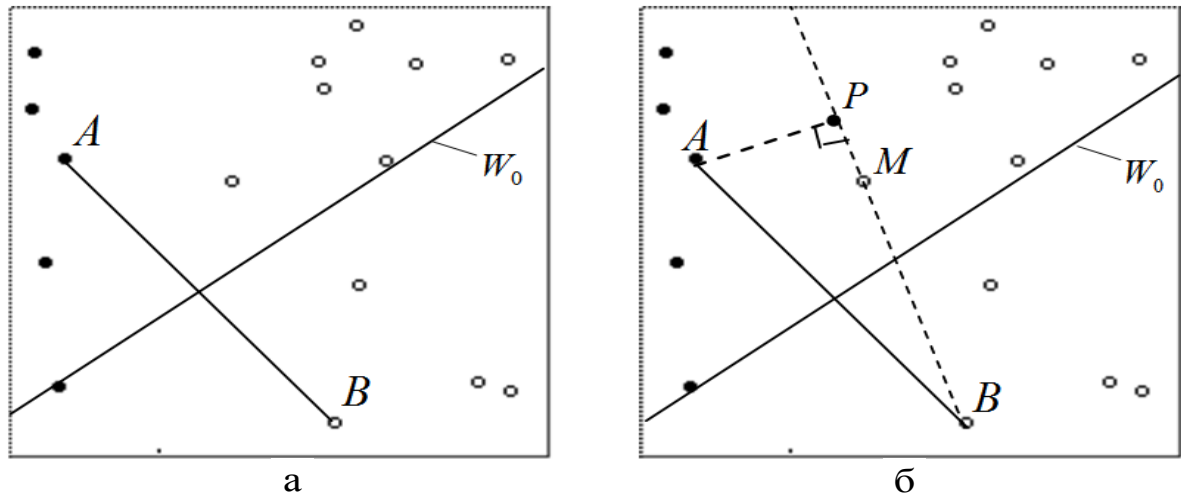


Рис. 6.20. Перші два кроки алгоритму Козинця:
а – крок 1; б – крок 2

Крок 2. Обирається довільна точка M і за знаком індикаторної змінної (6.26) визначається її приналежність до одного з класів. Точка M з'єднується прямою з точкою B того ж класу, а з точки A протилежного класу опускається перпендикуляр AP на вказану пряму (рис. 6.20, б).

Крок 3. Якщо основа перпендикуляра AP виходить за межі прямої BM (рис. 6.21, а), то точка M визначає нове положення прямої AB (рис. 6.21, б), за допомогою якої знаходять параметри відкоригованої дискримінантної функції.

Крок 4. Якщо ж основа перпендикуляра AP перебуває у межах відрізка BM (рис. 6.22, б), то нове положення прямої AB (параметри скоригованої дискримінантної функції $w^{(1)}$) визначає вже точка P .

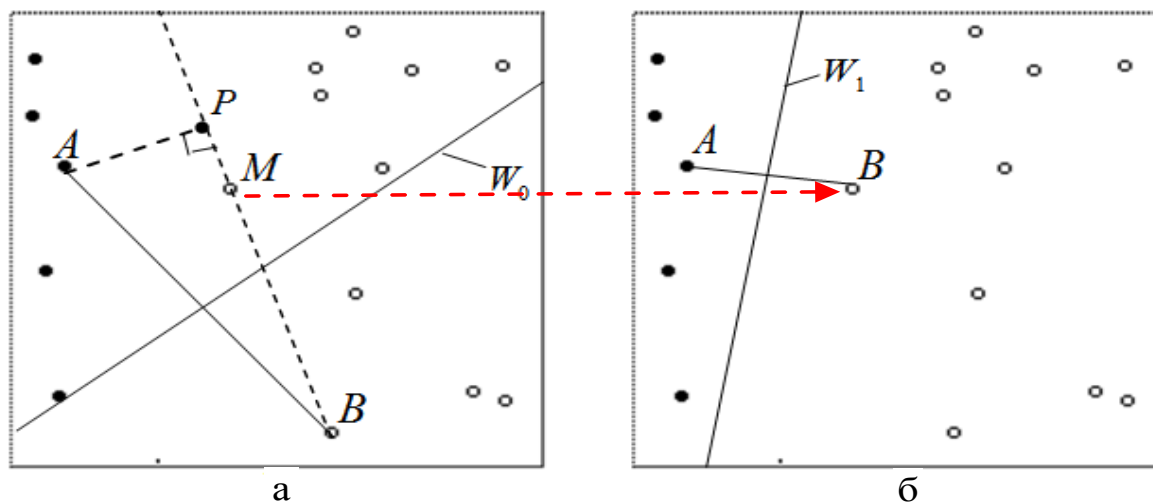


Рис. 6.21. Третій крок алгоритму Козинця:
а – до корекції параметрів; б – після корекції параметрів

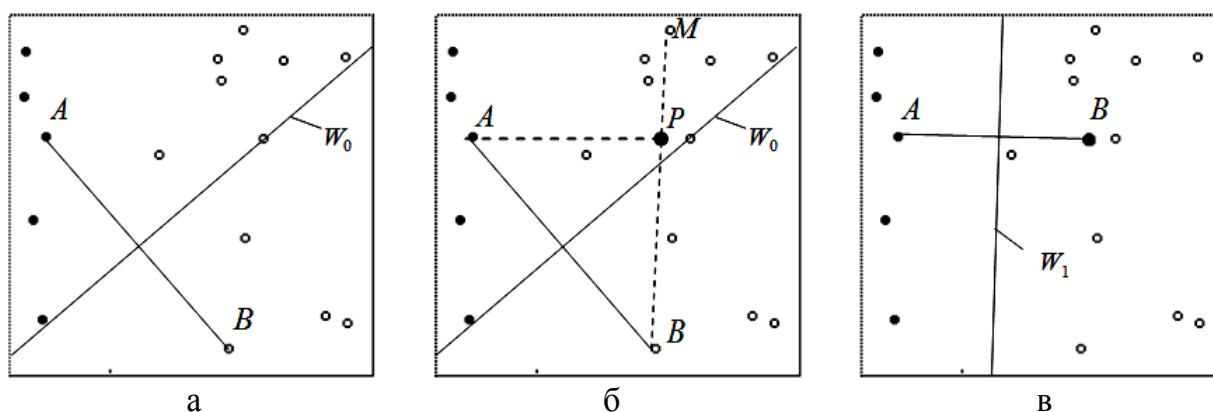


Рис. 6.22. Четвертий крок алгоритму Козинця:
а – початкове положення прямої AB ; б – визначення нового положення точки B ; в – нове положення прямої AB

Кроки 1–4 повторюються доки всі точки навчальної вибірки будуть правильно класифіковані.

Доведено, що алгоритм Козинця дає змогу:

- побудувати опуклі оболонки множин X^+ і X^- ;
- визначити найближчі пари точок $x^+ \in \text{Co}(X^+)$ і $x^- \in \text{Co}(X^-)$;
- побудувати серединний перпендикуляр до відрізка, що з'єднує ці точки.

Зауважимо, що алгоритм Козинця визначає не найближчі пари точок між множинами X^+ і X^- , а найближчі пари точок x^+ і x^- між опуклими оболонками $\text{Co}(X^+)$ і $\text{Co}(X^-)$ цих множин, що не одне й те саме! (рис. 6.23).

Самі ж точки x^+ і x^- можуть і не належати множинам X^+ і X^- .

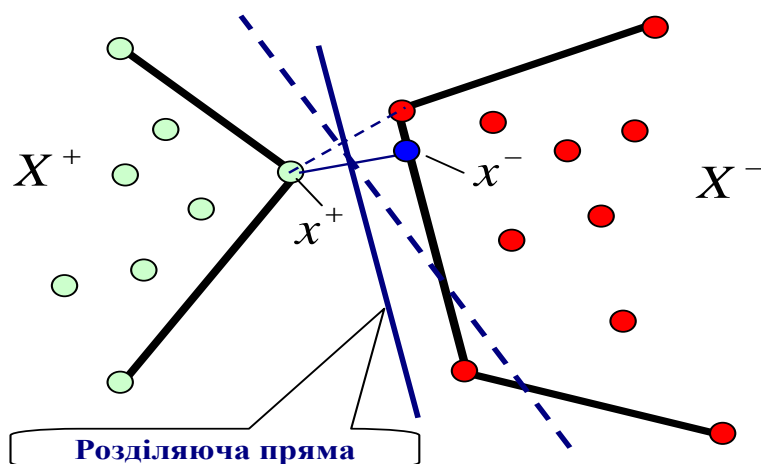


Рис. 6.23. Пояснення до алгоритму Козинця

Доведено теорему, відповідно до якої алгоритм Козинця сходиться за кінцеву кількість ітерацій

$$t_0 \leq \frac{Q^2}{\varepsilon^2} \ln \frac{Q^2}{\varepsilon^2}. \quad (6.39)$$

З порівняння оцінювання (6.36) і (6.39) може здатися, що алгоритм Розенблата завжди збігається за меншої кількості ітерацій, проте оцінювання (6.36) та (6.39) досить грубі. Отже, висновок про перевагу алгоритму Розенблата, зважаючи на це оцінювання, не правомірний.

Тому проводились експериментальне порівняння швидкості збігання алгоритмів на основі статистичного експерименту під час класифікації генерованих даних у навчальній вибірці (метода Монте-Карло).

Експерименти показали, що (рис. 6.24):

- під час навчання за малим обсягом даних (кількість спостережень $K \leq 50$) приблизно 20 % випадків алгоритми Розенблата і Козинця мали однакові швидкості збіжності;
- зі збільшенням кількості спостережень алгоритм Козинця виявлявся абсолютним лідером і коли $K > 100$ у 90 % випадків навчався швидше, ніж алгоритм Розенблата;
- швидкість збіжності алгоритму навчання Козинця менш чутлива до розміщення точок у просторі ознак.

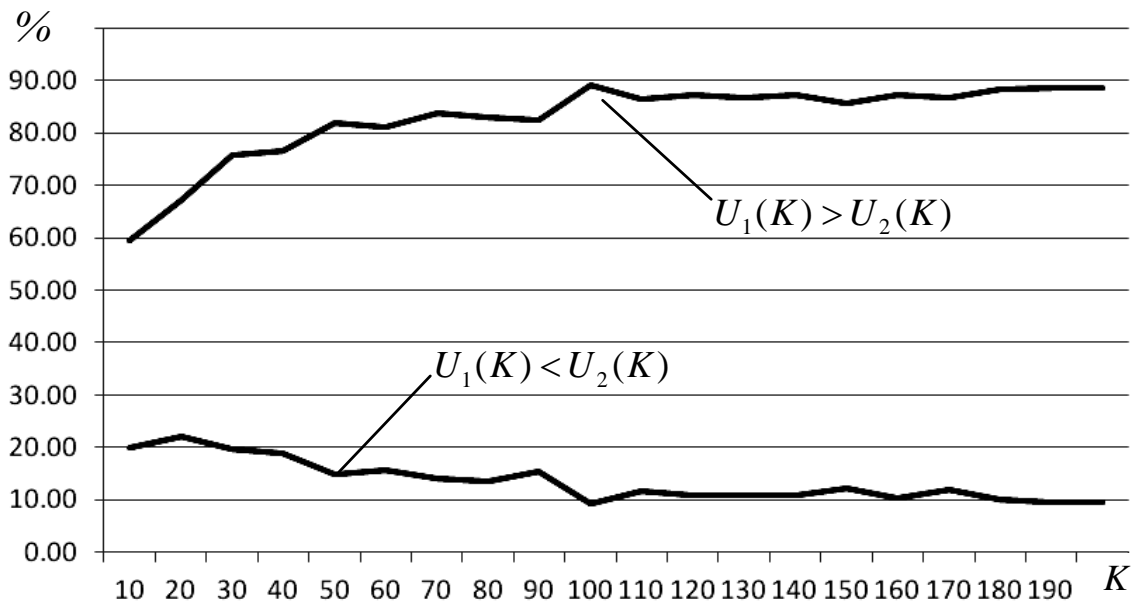


Рис. 6.24. Порівняння відсотка лідерства алгоритмів:
 $U_1(K) < U_2(K)$ – лідер алгоритм Розенблата;
 $U_1(K) > U_2(K)$ – лідер алгоритм Козинця

Слід зазначити, що вимога лінійної роздільності класів є дуже жорстким апіорним припущенням про структуру даних. Можна припустити, що на практиці ця вимога часто виявиться невиконаною. Доведення збіжності алгоритмів Розенблата та Козинця за кінцеву кількість ітерацій стверджує лише те, що в разі лінійної роздільності даних кількість коригувань вирішального правила завершується.

Однак без залучення додаткової інформації, яка характеризує поведінку алгоритму, нічого не можна сказати про те, зациквився алгоритм або довжина навчальної вибірки все ще недостатня для його зупинки.

6.5. Навчання штучної нейронної мережі

На рис. 6.25 показано найпростішу структуру штучної мережі прямого поширення (**feed-forward net**), яка містить три шари: вхідний, прихований (проміжний) і вихідний. У такій мережі інформація розповсюджується тільки в одному напрямку. У вузлах кожного з шарів розміщена певна кількість штучних нейронів, що реалізують математичну функцію (6.18).

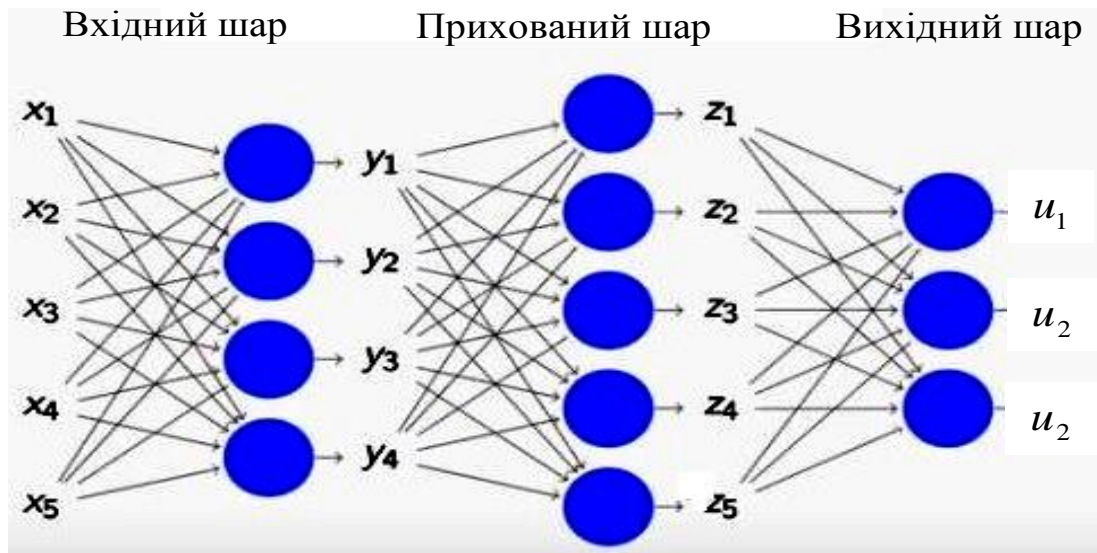


Рис. 6.25. Тришарова нейронна мережа прямого поширення

Вхідний шар застосовується для прийому та передачі вхідних даних x_1, x_2, \dots, x_N . На виході кожного з вузлів цього шару формуються сигнали

$$y_j = f^{(1)}(\mu_1 x_1 + \mu_2 x_2 + \dots + \mu_N x_N + \mu_0), \quad j = 1, \dots, J \quad (6.40)$$

у вигляді зваженої суми x_1, x_2, \dots, x_N з ваговими параметрами $\mu_1, \mu_2, \dots, \mu_N$ і, за необхідності, зсувом μ_0 , де $f^{(1)}(\cdot)$ – функції активації, яка нормує отриману суму.

Аналогічно прихований шар формує сигнали

$$z_m = f^{(2)}(v_1 y_1 + v_2 y_2 + \dots + v_J y_J + v_0), \quad m = 1, \dots, M. \quad (6.41)$$

Може бути не один, а кілька прихованих шарів. Кількість прихованих шарів називають глибиною мережі. Певні особливості прихованих шарів характеризують тип нейронної мережі.

І, нарешті, вихідний шар, з входу якого надходять сигнали z_1, z_2, \dots, z_M від попереднього шару, реалізує функцію

$$u_g = w_1 z_1 + w_2 z_2 + \dots + w_M z_M, \quad g = 1, \dots, G, \quad (6.42)$$

що визначає відповідь нейронного класифікатора як номер класу з цієї множини $\Psi = \{\Psi_1, \dots, \Psi_G\}$.

Можливу структуру штучної нейронної мережі для класифікації квітів ірису наведено на рис. 6.26.

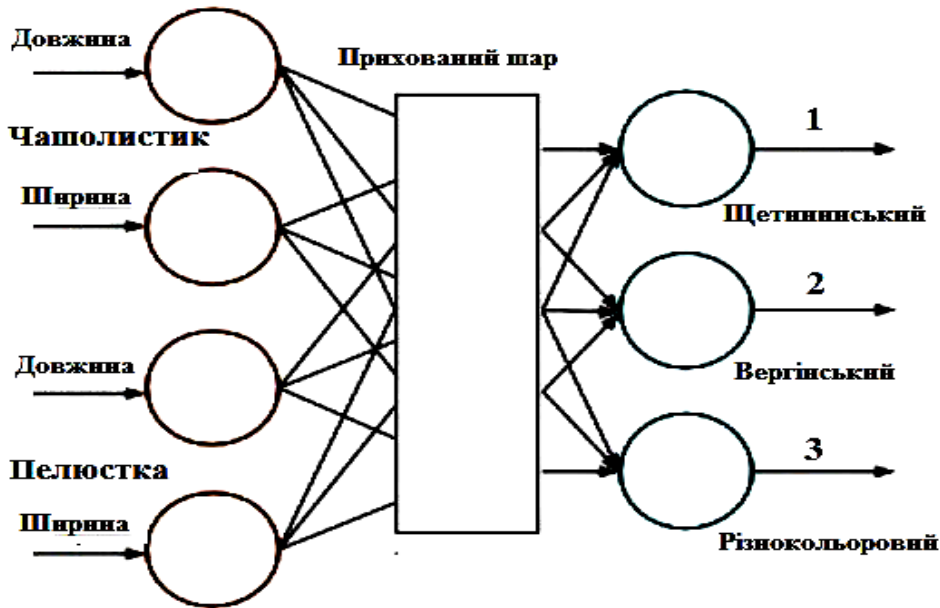


Рис. 6.26. Структура нейронної мережі для класифікації квітів ірису

Потенційна здатність нейронної мережі розв'язувати інтелектуальну задачу класифікації об'єктів ґрунтується на гіпотезі, що в багатовимірному просторі ознак образи розміщені компактно і можуть бути розділені деякою функцією $F(x)$, яка в загальному випадку нелінійна.

Теоретичне обґрунтування можливості апроксимації прийнятної $F(x)$ за допомогою навчання штучної нейронної мережі дає така теорема.

Теорема Цибенко (1989 р.). Для будь-якої неперервної функції $F(x)$ від вектора ознак $x = (x_1, \dots, x_N)$ існує такий вектор $\omega = (\omega_1, \dots, \omega_N)$ вагових коефіцієнтів і числа α_i, ω_{0i} , що

$$\left| F(x) - \sum_{i=1}^N \alpha_i \varphi(\omega^T x + \omega_{0i}) \right| < \varepsilon, \quad (6.43)$$

де $\varphi(\eta)$ – неперервна функція активації, наприклад, функція $\varphi(\eta) = 1/(1 + \exp^{-\eta})$, а $\varepsilon > 0$.

З теореми Цибенко випливає, що штучна нейронна мережа прямого розповсюдження з одним прихованим шаром може апроксимувати будь-яку неперервну функцію $F(x)$ багатьох змінних за достат-

ньою кількістю нейронів прихованого шару та вдалого вибору вагових коефіцієнтів.

У загальному випадку нейронні мережі з більшою кількістю прихованих шарів (більш глибокі мережі) можуть апроксимувати більш складні функції $F(x)$.

Для налаштування багатошарової нейронної мережі зазвичай використовують метод зворотного поширення помилок, який дає змогу відрегулювати кожний ваговий коефіцієнт мережі залежно від того, як цей параметр впливає на спільну помилку.

За рахунок ітеративного градієнтного спуску реалізується алгоритм навчання з вчителем, спрямований на зменшення загальної помилки за допомогою налаштування кожної ваги $\omega_i, i = 1, \dots, K_\omega$. У результаті матимемо оптимальний вектор вагових коефіцієнтів ω_{opt} , який задовольняє умову

$$\omega_{opt} = \arg \min_{\omega \in \Omega} \frac{1}{K} \sum_{i=1}^{K_\omega} (c_i - D(x_i, \omega_i))^2, \quad (6.44)$$

де K_ω – кількість вагових коефіцієнтів в усіх шарах нейронної мережі.

Згідно з (6.40)–(6.42) для тришарової мережі, показаної на рис. 6.25, можна записати

$$y = \langle x, \mu \rangle; \quad z = \langle y, \nu \rangle; \quad u = \langle z, w \rangle. \quad (6.45)$$

З (6.45) випливає, що загальна помилка нейронної мережі залежить від сигналів z_i , які надходять на вихідний шар, та відповідних ваг w_i . Зі свого боку сигнали z_i залежать від сигналів y_j , які надходять від прихованого шару, і відповідних ваг ν_j . І нарешті, сигнали y_j залежать від вхідних сигналів x_k і відповідних ваг μ_k .

Отже, загальну помилку нейронної мережі можна подати як складну функцію

$$E = E(u(z(y(x, \mu), \nu), w)) \quad (6.46)$$

і для визначення оптимальних значень вагових коефіцієнтів обчислити часткові похідні E за μ, ν, w за правилом диференціювання складних функцій.

Пояснимо процедуру навчання на фрагменті нейронної мережі (рис. 6.27).

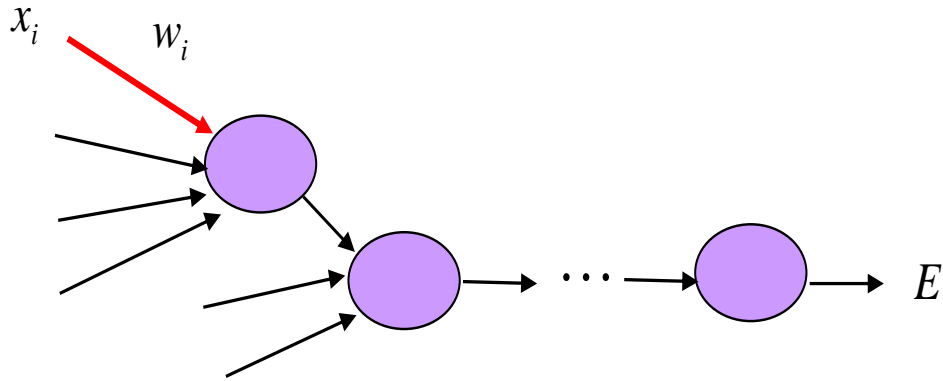


Рис. 6.27. Пояснення до методу навчання нейронної мережі

Часткова похідна $\frac{\partial E}{\partial w_i}$ визначає вклад (ступень довіри) до сигналу

x_i («градника») під час формування загальної помилки E колективного рішення. Тим самим природним чином розв'язується задача оцінювання розподілу сумарної помилки по всіх вагах нейронної мережі.

Водночас слід враховувати, що метод градієнтного спуску не дає можливість провести глобальну оптимізацію. Тому процедура навчання нейронної мережі може завершитися в точці локального мінімуму, яка далека від справжнього оптимального значення. Спроби подолати цю проблему на основі евристичних методів не завжди приводять до бажаного результату.

Слід також враховувати, що для ефективного проведення глибокого навчання необхідно визначити:

- оптимальну кількість прихованих шарів;
- прийнятну кількість нейронів у кожному шарі;
- обсяг об'єму вибірки, за якою проводять навчання.

Засновником ідеї глибокого навчання вважають англо-канадського вченого Джефрі Гінтона (**Geoffrey Everest Hinton**). Однак досі немає строгого наукового обґрунтування практичних результатів, отриманих за допомогою цього методу і фахівці всього світу працюють над вирішенням цієї проблеми.

Наприклад, професор Тель-Авівського університету Нафталі Тішбі (**Naftali Tishby**), працюючи на стику фізики, інформатики та обчислювальної нейробіології, запропонував концепцію інформаційної теорії глибокого навчання. Його результати ґрунтуються на теорії інформації та експериментів, запозичених зі статистичної фізики.

Головний висновок цих результатів полягає в тому, що основна мета глибинного навчання – забування непотрібної інформації під час проходження прихованих шарів.

На завершення нагадаємо, що ідеї глибинного навчання багатошарових нейронних мереж дуже близькі до ідеї багаторівневого алгоритму МГУА, які ще в 60-х роках минулого століття запропонував український вчений, академік О. Г. Івахненко.

Цей алгоритм пропонує процедуру побудови складної апроксимуючої функції на основі оптимального налаштування часткових залежностей

$$u_1 = u_1(x_1, x_2); \quad u_2 = u_2(x_1, x_3); \quad \dots; \quad u_M = u(x_{M-1}, x_M), \quad (6.47)$$

значення яких застосовують як аргументи моделей наступного рівня

$$v_1 = u(u_1, u_2); \quad v_2 = u(u_1, u_3); \quad \dots; \quad v_s = u_s(u_{s-1}, u_s) \quad (6.48)$$

і т. д.

6.6. Генетичний алгоритм

Генетичний алгоритм – один з різновидів еволюційних алгоритмів, в основу якого покладено теорію еволюції природних систем (дарвінізм).

Основні механізми еволюції:

- схрещування особин (породження нащадків), які наслідують властивості батьків (рис. 6.28);
- змінність (мутація) особин у популяції;
- природний відбір кращих особин.

Існує також додатковий механізм ізоляції – заборона певним особам породжувати нащадків.

Цікаво, що дарвінізм – емпірична наука, яка ґрунтувалась лише на спостереженнях її засновника. І тільки через багато років основні положення теорії еволюції теоретично підтвердилися у новій науці – генетиці, яка вивчає закономірності спадковості та мінливості.

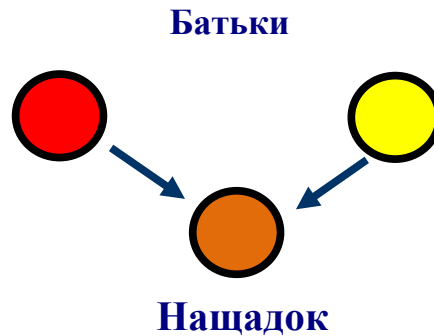


Рис. 6.28. Породження нащадків за допомогою схрещування особин

Спадкоємність властивостей батьків забезпечується передачею генетичної інформації, матеріальними носіями яких є гени, локалізовані в хромосомах батьків (рис. 6.29).

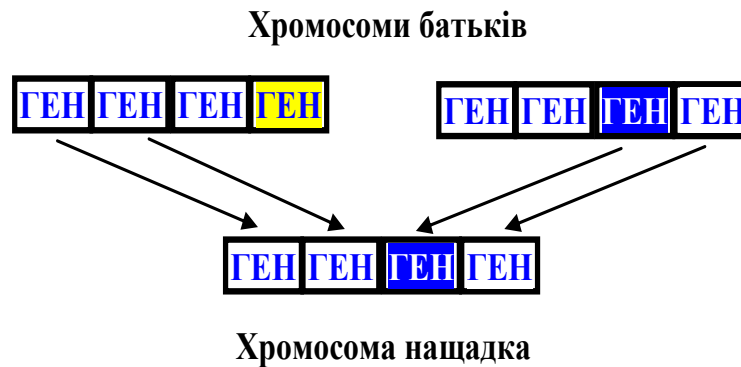


Рис. 6.29. Передача генетичної інформації нащадка

Розрізняють генотип індивіда, коли його властивості зумовлені лише генетичним матеріалом, успадкованим від батьків і фенотип індивіда, коли його властивості набуті у процесі розвитку та взаємодії з навколишнім середовищем.

Здатність нащадка набувати властивостей, що відрізняються від батьківських (мутації індивіда), здійснюються за допомогою змін генетичного матеріалу. Такі зміни можуть бути спричинені помилками копіювання генетичного матеріалу або зовнішніми факторами (радіація, хімічні речовини та ін.).

Мутації вважаються основною рушійною силою еволюції, коли менш сприятливі або шкідливі властивості видаляються з генофонду природним відбором, тоді як сприятливі (вигідні) прагнуть накопичуватись.

Саме такі механізми покладено в основу генетичного алгоритму, який реалізує така послідовність кроків (рис. 6.30):

- кодування особин та генерація популяції;
- відбір особин, що мутують;
- мутація;
- відбір особин, які схрещуються;
- схрещування;
- оцінювання результатів;
- рішення.



Рис. 6.30. Загальна схема генетичного алгоритму

Процес схрещування та відбору кращих особин триває багаторазово до досягнення необхідної якості розв’язування задачі або досягнення заданого користувачем обмеження на кількість поколінь.

Розглянемо два приклади.

Приклад 6.1. Дано множину цілих чисел:

$$A_0 = \{a_1, a_2, \dots, a_N\}. \quad (6.49)$$

Потрібно знайти підмножину $A \subset A_0$ таку, що сума всіх елементів в A дорівнюватиме сумі всіх елементів, що не належать A :

$$\sum_{a_i \in A} a_i = \sum_{a_j \in A_0 \setminus A} a_j, \quad i, j \in [1, \dots, N], \quad i \neq j. \quad (6.50)$$

Можна показати, що така задача NP повна, тобто з великими N повний перебір варіантів стає фактично неможливим. Розглянемо порядок розв'язку цієї задачі за допомогою генетичного алгоритму.

Для кодування можливих розв'язків задачі оберемо бітовий масив S довжиною N , в якому входження елементу у множину A буде кодувати 1, а не входження – 0. Наприклад, за умови $N = 7$ одне з таких рішень можна подати так:

$$A = \{a_2, a_3, a_5, a_6, a_7\}, \text{ якщо } S = 0110111. \quad (6.51)$$

Можливі різні способи генерації популяцій за такою схемою, наприклад, заповнення позицій S випадковими нулями та одиницями. Кількість елементів у початковій популяції обирає користувач.

На сформованій популяції можливих рішень реалізуємо механізми еволюції – мінливість, успадкування та відбір.

Механізм мутації $m(S)$ особини S можна реалізувати різними способами, наприклад так:

- з множини генерованих особин випадковим чином обирається та, яка піддається мутації;
- в обраній особині випадковим чином обираються біти (гени), які піддаються мутації;
- обрані біти інвертуються, тобто 1 змінюється на 0, а 0 на 1 (рис. 6.31).

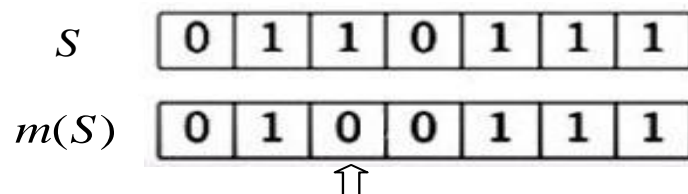


Рис. 6.31. Реалізація механізму мутації особини

Механізм схрещування двох особин, за допомогою якого породжується їх нащадок (кросовер), можна реалізувати різними способами, наприклад одним з таких способів.

Спосіб 1: випадковим чином обирають значення бітів (генів) з першої S_1 та другої S_2 особин, які породжують нащадка $C(S_1, S_2)$ (рис. 6.32).

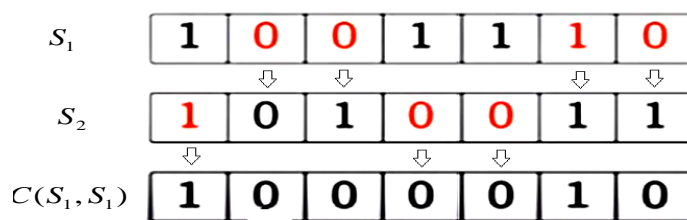


Рис. 6.32. Реалізація першого способу кросовера

Спосіб 2: Випадковим чином обирається індекс, до якого у нащадка успадковуються біти першої особи, а після цього індексу успадковуються біти другої особи (рис. 6.33).

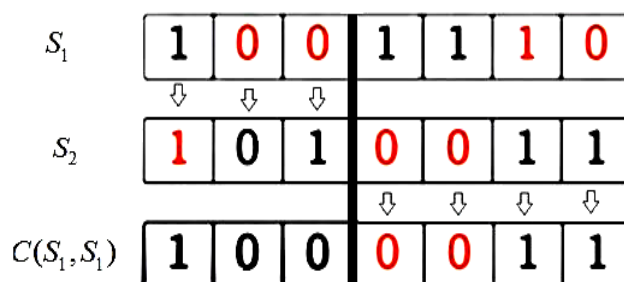


Рис. 6.33. Реалізація другого способу кросовера

Можливі більш «досконалі алгоритми» схрещування та успадкування генів батьків.

Після реалізації механізмів мутації та схрещування проводиться відбір сформованих особин. Конструювання механізму відбору один з головних етапів побудови генетичного алгоритму, тому що саме від нього суттєво залежить якість остаточних рішень

Для оцінювання конкретного розв'язку можна, наприклад, використовувати функцію

$$f_1(A) = \left| \sum_{a_i \in A} a_i - \sum_{a_j \in A_0 \setminus A} a_j \right|, \quad i, j \in [1, \dots, N], \quad i \neq j. \quad (6.52)$$

Однак краще замість $f_1(A)$ застосовувати зворотну функцію, яка тим більша, чим краща конкретна особина:

$$f_2(A) = \left| \sum_{a_i \in A} a_i - \sum_{a_j \in A_0 \setminus A} a_j \right|^{-1}, \quad i, j \in [1, \dots, N], \quad i \neq j. \quad (6.53)$$

Щоб уникнути проблеми поділу на нуль, доцільно модифікувати вираз (6.53), наприклад так:

$$f_3(A) = \left(1 + \left| \sum_{a_i \in A} a_i - \sum_{a_j \in A_0 \setminus A} a_j \right| \right)^{-1}, \quad i, j \in [1, \dots, N], \quad i \neq j. \quad (6.54)$$

Вираз (6.54) має такі властивості:

- нормована в інтервалі $[0, 1]$;
- чим гірше розв'язок, тим вираз ближче до 0 (якщо абсолютна різниця сум велика, то додавання 1 не змінить цього великого числа);
- чим краще знайдений розв'язок, тим вираз ближче до 1.

Експерименти показують, що застосування виразу (6.54) дає змогу за прийнятну кількість ітерацій знайти розв'язок задачі.

Приклад 6.2. Потрібно знайти цілі додатні числа a, b, c, d , які задовольняють діамантовому рівнянню

$$a + 2b + 3c + 4d = 30. \quad (6.55)$$

Оскільки

$$a, b, c, d \in [1, 30], \quad (6.56)$$

то повний перебір для визначення a, b, c, d потребує $30^4 = 810000$ варіантів.

Генетичний алгоритм дає змогу швидше знайти розв'язок за допомогою цілеспрямованого перебору.

Беручи до уваги (6.56), виберемо спочатку п'ять випадкових рішень – хромосом першого покоління (табл. 6.2) та обчислимо нев'язки (табл. 6.3) після підстановки генів конкретної хромосоми (значень a, b, c, d) в у рівняння (6.55):

$$\Delta_i = |a_i + 2b_i + 3c_i + 4d_i - 30|, \quad i = 1, \dots, 5. \quad (6.57)$$

Таблиця 6.2. Хромосоми першого покоління

Хромосома	(a, b, c, d)
№ 1	(1, 28, 15, 3)
№ 2	(14, 9, 2, 4)
№ 3	(13, 5, 7, 3)
№ 4	(23, 8, 16, 19)
№ 5	(9, 13, 5, 2)

Таблиця 6.3. Нев'язки початкових розв'язків

Хромосома		Δ_i
№ 1	(1, 28, 15, 3)	$ 114-30 =84$
№ 2	(14, 9, 2, 4)	$ 54-30 =24$
№ 3	(13, 5, 7, 3)	$ 56-30 =26$
№ 4	(23, 8, 16, 19)	$ 163-30 =133$
№ 5	(9, 13, 5, 2)	$ 58-30 =28$
Середня врозв'язку рівняння (6.55)		59

Вимагаємо, щоб якісніші хромосоми, які мають менші невязки Δ_i , мали більшу ймовірність P_i виявитися батьками. Для цього згідно з даними таблиці табл. 6.3, обчислимо ймовірності

$$P_i = \frac{1}{\Delta_i} / \left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2} + \frac{1}{\Delta_3} + \frac{1}{\Delta_4} + \frac{1}{\Delta_5} \right), \quad (6.58)$$

та запишемо їх у табл. 6.4.

Як видно з таблиці найбільш «привабливими» батьками мають бути хромосоми № 2, № 3 та № 5, для яких більш висока ймовірність P_i порівняно з хромосомами № 1 та № 4. Ґрунтуючись на цьому, сформуємо таблицю «привабливих» пар батьків (табл. 6.5.)

Таблиця 6.4. Оцінювання привабливості батьків

Хромосома		Ймовірність P_i бути батьком
№ 1	(1, 28, 15, 3)	$(1/84)/0,135266 = 8,80 \%$
№ 2	(14, 9, 2, 4)	$(1/24)/0,135266 = 30,8 \%$
№ 3	(13, 5, 7, 3)	$(1/26)/0,135266 = 28,4 \%$
№ 4	(23, 8, 16, 19)	$(1/133)/0,135266 = 5,56 \%$
№ 5	(9, 13, 5, 2)	$(1/28)/0,135266 = 26,4 \%$

Реалізуємо механізм схрещування батьків, що сформовані у табл. 6.5, таким чином. Нехай батьки мають такий набір хромосом (розв'язків рівняння (6.55)):

$$1\text{-й батько} - a_1, b_1, c_1, d_1; \quad (6.59)$$

$$\text{2-й батько} - a_2, b_2, c_2, d_2. \quad (6.60)$$

Таблиця 6.5. Найбільш привабливі батьки та їх хромосоми

Хромосоми першого батька		Хромосоми другого батька	
№ 3	(13, 5, 7, 3)	№ 1	(1, 28, 15, 3)
№ 5	(9, 13, 5, 2)	№ 2	(14, 9, 2, 4)
№ 3	(13, 5, 7, 3)	№ 5	(9, 13, 5, 2)
№ 2	(14, 9, 2, 4)	№ 5	(9, 13, 5, 2)
№ 5	(9, 13, 5, 2)	№ 3	(13, 5, 7, 3)

Грунтуючись на (6.59), (6.60) можна побудувати шість варіантів схрещування хромосом першого та другого батьків (табл. 6.6). У таблиці знаком | відмічено поділ між генами батьків, які передаються нащадкам (кросоверам).

На основі табл. 6.5 та 6.6 сформуємо табл. 6.7, в якій наведено хромосоми можливих нащадків, що породжені найбільш привабливими батьками.

Таблиця 6.6. Можливі варіанти кросоверів

Хромосома першого батька	Хромосома другого батька	Хромосома нащадка
$a_1 / b_1, c_1, d_1$	$a_2 / b_2, c_2, d_2$	a_1, b_2, c_2, d_2 або a_2, b_1, c_1, d_1
$a_1, b_1 / c_1, d_1$	$a_2, b_2 / c_2, d_2$	a_1, b_1, c_2, d_2 або a_2, b_2, c_1, d_1
$a_1, b_1, c_1 / d_1$	$a_2, b_2, c_2 / d_2$	a_1, b_1, c_1, d_2 або a_2, b_2, c_2, d_1

Таблиця 6.7. Результати схрещування привабливих батьків

Хромосоми 1-батька	Хромосоми 2-го батька	Хромосоми нащадків
(13, 5, 7, 3)	(1, 28, 15, 3)	(13, 28, 15, 3)
(9, 13, 5, 2)	(14, 9, 2, 4)	(9, 13, 2, 4)
(13, 5, 7, 3)	(9, 13, 5, 2)	(13, 5, 7, 2)
(14, 9, 2, 4)	(9, 13, 5, 2)	(14, 13, 5, 2)
(9, 13, 5, 2)	(13, 5, 7, 3)	(13, 5, 5, 2)

Обчислимо нев'язки розв'язків, які отримано в результаті схрещування привабливих батьків (табл. 6.8).

Таблиця 6.8. Нев'язки розв'язків нащадків

Хромосоми нащадків	Не Δ_i
(13, 28, 15, 3)	$ 126-30 =96$
(9, 13, 2, 4)	$ 57-30 =27$
(13, 5, 7, 2)	$ 57-30 =22$
(14, 13, 5, 2)	$ 63-30 =33$
(13, 5, 5, 2)	$ 46-30 =16$
Середня нев'язка розв'язку рівняння (6.55)	38,8

З порівняння результатів обчислень, наведених у табл. 6.3 і 6.8, випливає, що середня непогодженість розв'язків діофантового рівняння (6.55), знайдених за допомогою механізму схрещування, становить 38,8, тоді як середня непогодженість батьків становила 59.

Процес схрещування може бути продовжений із генерацією нащадків наступних поколінь. Для цих нащадків може бути застосовано механізм мутації, наприклад заміною одного з розв'язків випадковим цілим числом від 1 до 30.

Продовжуючи процес таким чином, одна з породжених хромосом може навіть забезпечити не0, тобто виявитися дійсним розв'язуванням рівняння (6.55). Експерименти показують, що системи з великою початковою популяцією (наприклад, з 50 початковими розв'язками замість 5) достатньо швидко сходяться до бажаного розв'язку.

Можна сформулювати таку практичну рекомендацію.

Нейронні мережі та генетичні алгоритми слід застосовувати, якщо

точний алгоритм розв'язання задачі невідомий;

алгоритм відомий, але зводиться до *NP*-повної задачі;

нераціонально шукати точний алгоритм, коли його практична реалізація у великому проекті займає лише малу частину загального завдання, а отже, витрати на пошук такого алгоритму не можна порівняти з вигодою.

Запитання для самоконтролю

1. Охарактеризуйте змістовну постановку задачі машинного навчання.
2. Що таке навчальна вибірка?
3. Що таке лінійний класифікатор?
4. Охарактеризуйте задачу навчання з учителем.
5. Дайте визначення опуклої оболонки.
6. В якому випадку дві множини спостережень можуть бути лінійно розділені?
7. Що означає спрямовуючий простір?
8. Які етапи передбачає методологія машинного навчання?
9. Що таке метод градієнтного спуску?
10. Що означає гіпотеза компактності?
11. Охарактеризуйте метод найближчого сусіда.
12. Наведіть коротку характеристику одношарового персептрона?
13. Які логічні функції може реалізувати одношаровий персептрон?
14. Дайте коротку характеристику алгоритму навчання Розенблата.
15. Наведіть основний висновок теореми Новікова.
16. Дайте коротку характеристику алгоритму навчання Козинця.
17. Наведіть основний висновок теореми Цибенко.
18. У чому полягає метод зворотного поширення помилок?
19. Які основні механізми реалізує генетичний алгоритм?

РОЗДІЛ 7

ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ПРЕВЕНТИВНОЇ МЕДИЦИНИ

7.1. Загальні характеристики превентивної медицини

Превентивна медицина – це напрям у сучасній медичній науці та практиці, головною метою якої є збереження здоров'я людини. Принцип «хворобі легше запобігти, ніж лікувати» відомий багатьом людям, проте мало хто дотримується його на практиці. І тут на перший план виходить превентивна діагностика, основана на виявленні початкових ознак розвитку загрозливих захворювань, які за несвоєчасного початку лікування можуть призводити до серйозних ускладнень, інвалідності і навіть смерті пацієнта.

Саме на мінімізацію ризиків розвитку захворювання та їх ускладнень, збереження здоров'я та молодості людини на довгі роки спрямовані засоби превентивної діагностики.

Діагностика захворювань як наука пройшла декілька стадій свого розвитку (рис. 7.1).



Рис. 7.1. Стадії розвитку діагностики як науки

Засновник клітинної патології, німецький вчений XIX століття Рудольф Вірхов (**Rudolf Virchow**) пов'язував діагностичне рішення лише з виявленням специфічного патофізіологічного субстрату в ор-

ганізмі людини. Наприклад, наявності виразкового ураження, стенозу коронарних судин або виявлення ракових пухлин. Така діагностика можлива лише на фінальній стадії розвитку хвороби (рис. 7.1, а).

У ХХ сторіччі сформувалося уявлення про функціональні порушення організму, які проявляються вже на початкових стадіях хвороби (рис. 7.1, б). Наприклад, на початкових стадіях виразкова хвороба проявляється у вигляді різних порушень функцій шлунка (роздратованого шлунка), які передують виразковій хворобі.

Наступний етап – формування уявлення про донозологічну діагностику і передхворобу (рис. 7.1, в). Вважається, що перехід від фізіологічної норми, коли організм має задовільну адаптацію до впливу зовнішніх факторів, проходить стадії напруги та перенапруги механізму адаптації і завершується хворобою (зривом адаптації).

Під час проведення профілактичних обстежень діагностичні рішення спираються на поняття медичної норми, яка зазвичай зводиться до референтних інтервалів фізіологічних показників. Наприклад, вважається, що у здорової людини систолічний артеріальний тиск не повинен перевищувати 140 мм рт. ст., а частота серцевих скорочень має належати інтервалу 60–90 уд/хв.

Однак клінічна практика показує, що у багатьох людей перебіг захворювань виходить за загальноприйняті межі. Значною мірою це пояснюється тим, що відповідно до наявних стандартів для визначення референтних інтервалів використовують значення, отримані лише у 95 % обстежуваної групи здорових осіб.

Звідси випливає, що у 5 % здорових людей (кожного двадцятого) фізіологічні показники можуть не «потрапляти» у встановлені межі референтного інтервалу. Інакше кажучи, прийняття діагностичних рішень за правилом

$$\begin{aligned} &\text{Норма, якщо } x_t \in X_0, \\ &\text{Увага, якщо } x_t \notin X_0, \end{aligned} \tag{7.1}$$

що ґрунтується на аналізі відповідності поточного результату вимірювання x_t загальноприйнятої медичної норми X_0 , можуть призводити до помилково позитивних і/та помилково негативних результатів діагностики.

Крім того, відомо, що більшість фізіологічних показників схильні до значних спонтанних коливань (**True biological variability**) і це вважається фізіологічною нормою. Тому епізодичний контакт пацієн-

та з лікарем навіть під час використання найдосконалішої діагностичної системи може призводити до неправильного оцінювання ризику розвитку захворювання, якщо приймати рішення тільки за правилом (7.1).

Розглянемо інший підхід до прийняття діагностичних рішень, який спирається на головний принцип персоніфікованої медицини – лікувати хворого, а не хворобу, враховуючи індивідуальні особливості організму.

Припустимо, що пацієнт має можливість самостійно на досить великому проміжку часу оцінювати фізіологічні показники, що мають інформацію про поточний стан організму. Результати $x_1, x_2, x_3 \dots$ таких спостережень утворюють множину (навчальну вибірку)

$$X = \{x_1, \dots, x_N\}. \quad (7.2)$$

Елементи множини (7.2) можуть мати різну форму. В одних випадках $x_n \in X$, $n = 1, \dots, N$ – скалярна величина, наприклад, рівень глюкози в крові. В інших випадках $x_n \in X$ – вектор, компоненти якого є значеннями сукупності фізіологічних показників, наприклад, показників варіабельності серцевого ритму. Нарешті, результат тестування може мати більш складну форму, наприклад у вигляді фоноспірограми, що характеризує спектральні компоненти респіраторного шуму під час дихання пацієнта.

Незалежно від форми подання результатів будемо зображати $x_n \in X$ як точку в деякому метричному просторі, вважаючи що для будь-яких двох елементів $x_i \in X$ і $x_j \in X$ можна обчислити величину $S(x_i, x_j)$, яка має властивості метрики:

- $S(x_i, x_j) \geq 0$,
- $S(x_i, x_j) = S(x_j, x_i)$;
- $S(x_i, x_j) \leq S(x_i, x_z) + S(x_z, x_j)$,

де $x_z \in X$.

Величина $S(x_i, x_j)$ характеризує близькість x_i і x_j у метричному просторі, в якому прийматимемо персоніфіковані рішення.

Зобразимо результати багаторазових вимірювань (7.2), проведених i -м пацієнтом, областю Ω_i з центром C_i . Подамо також у цьому самому просторі область Ω_0 з центром C_0 результати популяційних обстежень репрезентативної групи здорових добровольців.

Розглянемо чотири варіанти взаємного розміщення областей Ω_0 та Ω_i у просторі ознак (рис. 7.2):

- $\Omega_i \subset \Omega_0$, тобто індивідуально область Ω_i повністю входить до області Ω_0 популяційної норми (рис. 7.2, а);
- $(\Omega_i \cap \Omega_0) \neq \emptyset$, $\Omega_i / (\Omega_i \cap \Omega_0) \neq \emptyset$, тобто індивідуальна область Ω_i лише частково входить до області Ω_0 популяційної норми, причому $C_i \in \Omega_0$ (рис. 7.2, б);
- $(\Omega_i \cap \Omega_0) \neq \emptyset$, $\Omega_i / (\Omega_i \cap \Omega_0) \neq \emptyset$, тобто індивідуальна область Ω_i також частково входить до області Ω_0 популяційної норми, але $C_i \notin \Omega_0$ (рис. 7.2, в);
- $(\Omega_i \cap \Omega_0) = \emptyset$, тобто індивідуальна область Ω_i виходить за межі області Ω_0 популяційної норми Ω_0 (рис. 7.2, г).

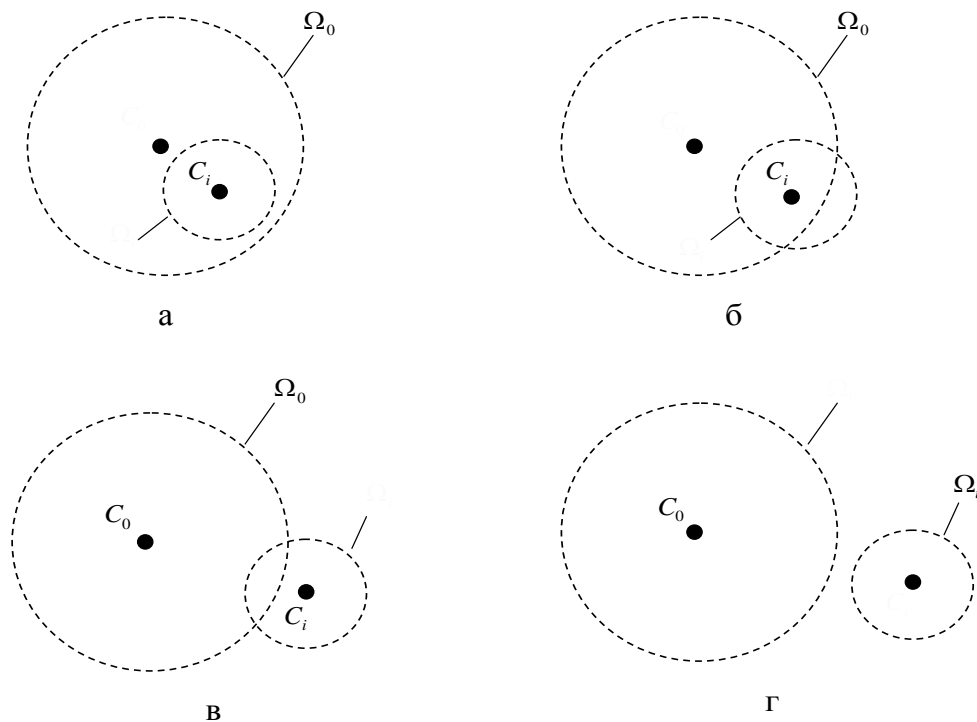


Рис. 7.2. Популяційна Ω_0 та індивідуальна Ω_i області значень фізіологічних показників з центрами C_0 і C_i

Звідси у першій ситуації (рис. 7.2, а) пацієнта слід визнати здоровим. Оскільки в другій ситуації (рис. 7.2, б) середнє значення C_i також належить області популяційної норми Ω_0 , то тоді треба визначити пацієнта, як умовно здорового, незважаючи на те, що результати окремих вимірювань не відповідали референтним величинам.

У третій ситуації (рис. 7.2, в) і особливо в четвертій (рис. 7.2, г) не залишається нічого іншого, як визначити пацієнта до групи хворих або групи високого ризику захворювання, що вимагає поглибленого обстеження.

Аналіз взаємного розміщення індивідуальних областей Ω_i , Ω_j різних пацієнтів відносно області популяційної норми (рис. 7.3, а) дає змогу оцінювати ступені ризику патології у i -го та j -го пацієнтів за відстанями s_i і s_j між відповідними центрами C_i , C_j та еталоном C_0 (рис. 7.3, а).

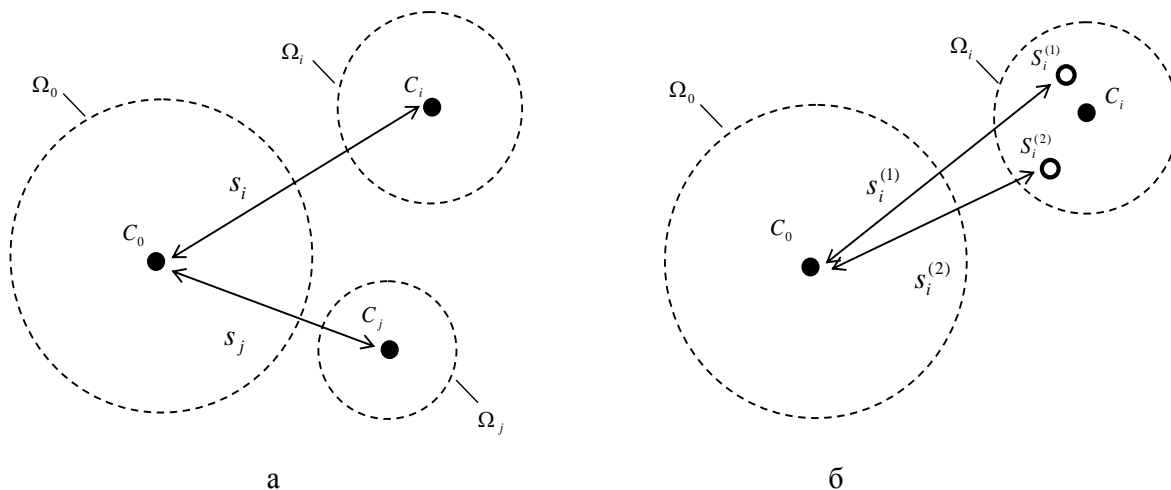


Рис. 7.3. Ідея методу прийняття персоналізованих рішень:
а – для двох пацієнтів; б – для одного пацієнта

Результати кожного нового вимірювання показників i -го пацієнта можуть постійно уточнювати його персоналізовану норму Ω_i , а порівняння відстаней $s_i^{(1)}$ і $s_i^{(2)}$ між результатами вимірювань у станах $S_i^{(1)}$, $S_i^{(2)}$ відносно еталона C_0 дають змогу зробити висновок, що стан $S_i^{(2)}$ кращий за стан $S_i^{(1)}$, якщо $s_i^{(2)} < s_i^{(1)}$ (рис. 7.3, б).

I, нарешті, персоніфікована область Ω_i показників i -го пацієнта може бути розділена на окремі зони персоніфікованих рішень (рис. 7.4), а саме:

- зона I – персональна норма;
- зона II – помірних відхилень від персональної норми;
- зона III – виражених відхилень від персональної норми.

За допомогою лінії (поверхні), що визначається за рівнянням $s = s_i$, де s_i – відстань між центром C_i і еталоном C_0 , зони II і III можна розділити на дві підобласті: помірних і виражених погіршень, коли $s > s_i$, і помірних і виражених поліпшень, коли $s < s_i$.

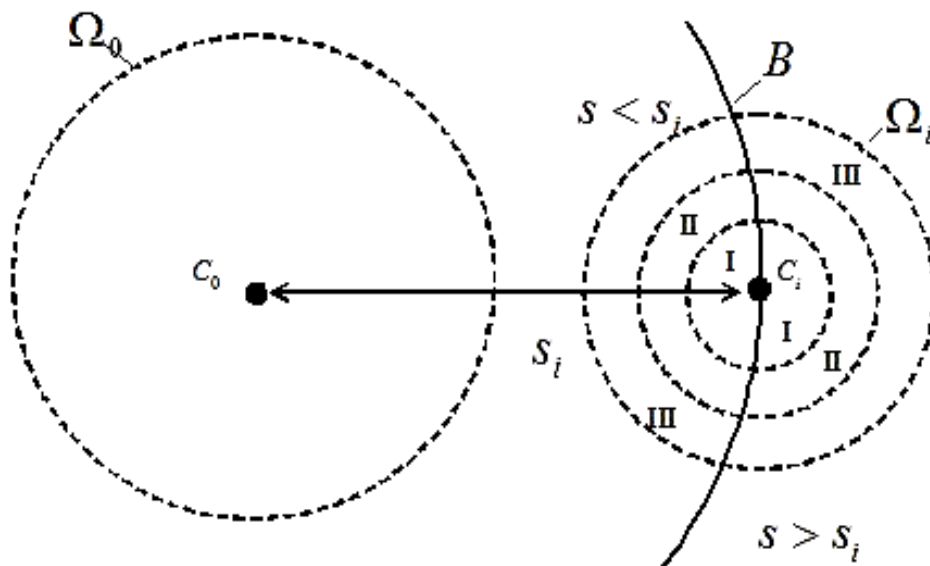


Рис. 7.4. Зони персоніфікованих рішень

Розглянемо формальні положення, на яких буде ґрунтуватися метод персоніфікованих діагностичних рішень за масивом спостережень (7.2), які користувач самостійно накопичує за певний проміжок часу.

Сформуємо квадратну матрицю відстаней $S(x_i, x_j)$, $i = 1, \dots, N$, $j = 1, \dots, N$ між усіма парами елементів множини (7.2)

$$R = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1N} \\ S_{21} & S_{22} & \dots & S_{2N} \\ \dots & \dots & \dots & \dots \\ S_{N1} & S_{N2} & \dots & S_{NN} \end{pmatrix}. \quad (7.3)$$

Матриця R дає змогу визначити дві інтегральні величини: еталонний результат $x_0 \in X$, який найближчий до всіх інших спостережень з навчальної вибірки (7.2), тобто

$$x_0 = \arg \min_{1 \leq j \leq N} \sum_{i=1}^N S_{ij}, \quad (7.4)$$

та середнє відхилення \bar{S} результатів спостережень, яке обчислюється за формулою

$$\bar{S} = \frac{1}{N^2 - N} \sum_{i=1}^N \sum_{j=1}^N S_{ij}. \quad (7.5)$$

Зауважимо, що умова (7.5) відрізняється від умови, яку задовольняє геометричний центр

$$\tilde{x}_0 = \arg \min_{x \in R^N} \sum_{i=1}^N S(x_n, x), \quad (7.6)$$

який є точкою \tilde{x}_0 у N -вимірному просторі, від якої сума евклідових відстаней до всіх точок $x_n \in X$ мінімальна.

Незважаючи на зовнішню подібність (7.4) і (7.6), величини x_0 і \tilde{x}_0 різні: в загальному випадку величина \tilde{x}_0 може не збігатися з жодним результатом навчальної вибірки X , тоді як $x_0 \in X$ є найбільш характерним результатом наявних спостережень.

Це значно спрощує процедуру визначення еталонного результату x_0 порівняно з оптимізаційною процедурою (7.6) обчислення геометричного центру \tilde{x}_0 , практична реалізація якої викликає відомі обчислювальні труднощі.

Якщо спостереження x_1, \dots, x_N – скалярні величини, то процедура визначення еталонного результату x_0 стає ще простіше: достатньо обчислити медіану в одновимірній вибірці спостережень x_1, \dots, x_N . Якщо ж до того x_1, \dots, x_N розподілені симетрично щодо моди (наприклад, за нормальним законом), то величину можна x_0 оцінити арифметичним середнім значенням.

Для зручності можна скористатися рекурентним співвідношенням

$$x_0^{(n)} = x_0^{(n-1)} + \frac{1}{n} [x_n - x_0^{(n-1)}], \quad n = 1, 2, \dots, \quad (7.7)$$

з початковою умовою $x_0^{(0)} = 0$, що дає змогу постійно уточнювати еталонний результат x_0 у процесі накопичення даних, не зберігаючи весь масив спостережень x_1, \dots, x_N . Аналогічно можна обчислювати величину $\bar{S}^{(n)}$, яка так само, як і $x_0^{(n)}$, прямує до стійкого значення зі зростанням кількості спостережень.

У випадках, коли кожен із результатів $x_n = (x_{n1}, x_{n2}, \dots, x_{nM}) \in X$, $n = 1, \dots, N$ навчальної вибірки – вектор (упорядкований набір M діагностичних показників), еталон x_0 можна визначити як M -вимірний вектор $x_0 = (x_{01}, x_{02}, \dots, x_{0M})$, компоненти якого є медіани чи середні значення відповідних компонент векторів з навчальної вибірки. В цьому випадку вектор-еталон x_0 не обов'язково належить множині X .

Інтегральні величини x_0 і \bar{S} характеризують індивідуальні особливості організму конкретного пацієнта: еталон x_0 відображає центральну тенденцію (**central tendency**) результатів спостережень x_1, \dots, x_N , а величина \bar{S} – ступінь мінливості цих спостережень.

Така інформація дає змогу додатково до (7.1) приймати персоніфіковане рішення про поточний стан пацієнта за правилом:

$$\begin{aligned} &\text{Персональна норма, якщо } S(x_t, x_0) \leq \lambda \bar{S}, \\ &\text{Увага, якщо } S(x_t, x_0) > \lambda \bar{S}, \end{aligned} \quad (7.8)$$

де x_t – результат поточного спостереження, а $\lambda \geq 1$ – коефіцієнт, який забезпечує бажаний компроміс між чутливістю і специфічністю діагностичних рішень.

Крім того, аналіз належності x_0 популяційної норми X_0 дає змогу оцінити ризик можливого розвитку патології за правилом.

$$\begin{aligned} &\text{Низький ризик патології, якщо } x_0 \in X_0, \\ &\text{Високий ризик патології, якщо } x_0 \notin X_0. \end{aligned} \quad (7.9)$$

Правило (7.9) принципово відрізняється від правила (7.1), оскільки діагностичне рішення приймається не за поточним спостереженням x_t , а за еталоном x_0 , обчисленим за множиною спостережень.

Тим самим запропонований похід, на відміну від традиційного, забезпечує адаптацію діагностичних правил до індивідуальних особ-

ливостей організму конкретного користувача, що підвищує достовірність прийнятих рішень.

Оскільки приклади часто бувають переконливішими, ніж формальні викладки, розглянемо деталі запропонованого підходу на прикладах побудови конкретних комп'ютерних засобів превентивної діагностики.

7.2. Інтелектуальний електрокардіограф ФАЗАГРАФ®

Ішемічна хвороба серця (ІХС) залишається найпоширенішим кардіологічним захворюванням, яке часто призводить до інвалідності та смерті пацієнта. Дані останніх епідеміологічних досліджень свідчать про підвищення серцевої недостатності, спричиненої ІХС.

Однією з суб'єктивних ознак ІХС є за груди́нні болі, викликані невідповідністю коронарного кровотоку потребам серцевого м'яза в кисні. Однак навіть під час поглибленого опитування не завжди вдається виявити напади стенокардії у значної кількості хворих на ІХС, або такі напади мають атиповий характер.

Зрозуміло, що для масових профілактичних обстежень неможливо використовувати метод коронарографії (**Coronary Angiography**). Хоча цей метод визнаний «золотим стандартом» під час виявлення стенозу судин та діагностиці ІХС, коронарографія – доволі коштовний і, найголовніше, небезпечний метод обстеження.

Відомо також, що ЕКГ спокою, що оцінюється за загальноприйнятими критеріями, залишається нормальною у половини хворих із хронічною формою ІХС. Тому розроблення безпечних та надійних методів діагностики ІХС на ранніх стадіях – важлива проблема, вирішення якої може сприяти своєчасно призначеному лікуванню та зниженню смертності населення.

Наочним прикладом вдалого застосування інтелектуальних ІТ у цифровій медицині є інноваційний метод у кардіології – метод фазаграфії, створений Міжнародним науково-навчальним центром інформаційних технологій та систем НАН України та МОН України. Метод реалізовано в програмно-технічному комплексі ФАЗАГРАФ®, який отримав:

- свідоцтво № 14006/2014 про Державну реєстрацію з необмеженим строком дії та дозвіл на застосування в медичній практиці;

- державний сертифікат № UA-MI/1-2558-2009 затвердження засобу вимірювальної техніки;
- рекомендацію МОЗ України для скринінгу ішемічної хвороби серця.

ФАЗАГРАФ[®] (рис. 7.5) складається з мікропроцесорного сенсора з пальцевими електродами, який забезпечує реєстрацію ЕКГ з першого стандартного відведення та введення цифрового сигналу в персональний комп'ютер через стандартні порти USB або Bluetooth.

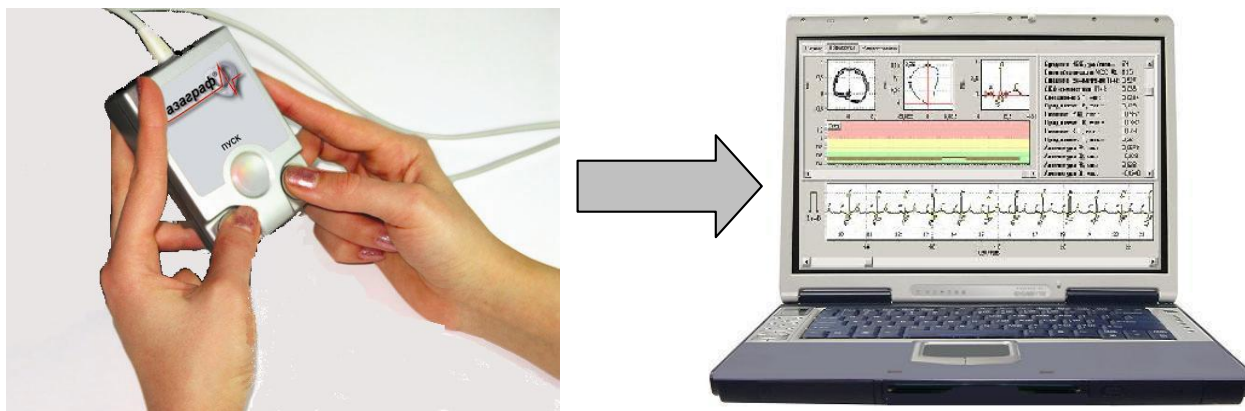


Рис. 7.5. Програмно-технічний комплекс ФАЗАГРАФ[®]

Комп'ютерна програма реалізує всі стадії інтелектуальної інформаційної технології оброблення сигналу, про які йшлося в розділі 4, зокрема:

- відновлення еталонного кардіоциклу у часовій області й аналіз традиційних ЕКГ ознак;
- виділення фрагмента усередненої фазової траєкторії, що відповідає стадії реполяризації (зубцю T);
- оцінювання додаткового діагностичного показника $\beta = D_2 / D_1$, що характеризує симетричність форми зазначеного фрагмента усередненої фазової траєкторії (рис. 7.6);
- реалізацію діагностичного правила, що ґрунтується на порівнянні β з пороговим значенням.

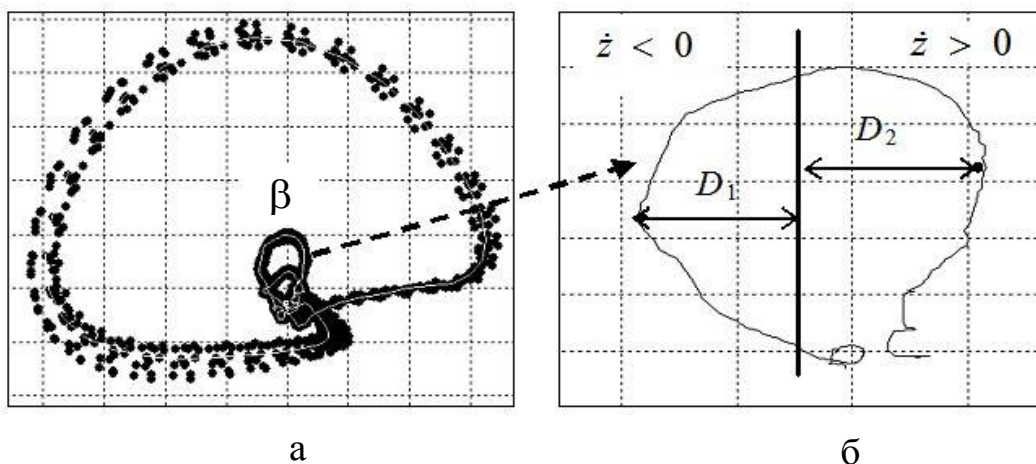


Рис. 7.6 Фазова траєкторія ЕКГ (а) та її фрагмент реполяризації (б)

Головне вікно комп'ютерної програми має декілька сторінок. На першій сторінці в реальному часі відображаються графіки вихідного сигналу першого стандартного відведення та фільтрований сигнал, а також результат автоматичного поділу ЕКГ на окремі серцеві цикли, виявлені екстрасистоли (відзначаються сірим кольором) та мітки зубців P , Q , R , S , T (рис. 7.7).

Лікар має можливість задати потрібний масштаб відображення графіків за рівнями сигналу (5, 10, 20, 40 мВ/мВ) та часом (10, 25, 50, 100, 250 мм/с). За допомогою кнопок меню можна автоматично переміщатися вздовж сигналу до ділянок, на яких виявлені атипівні цикли (екстрасистоли чи артефакти).

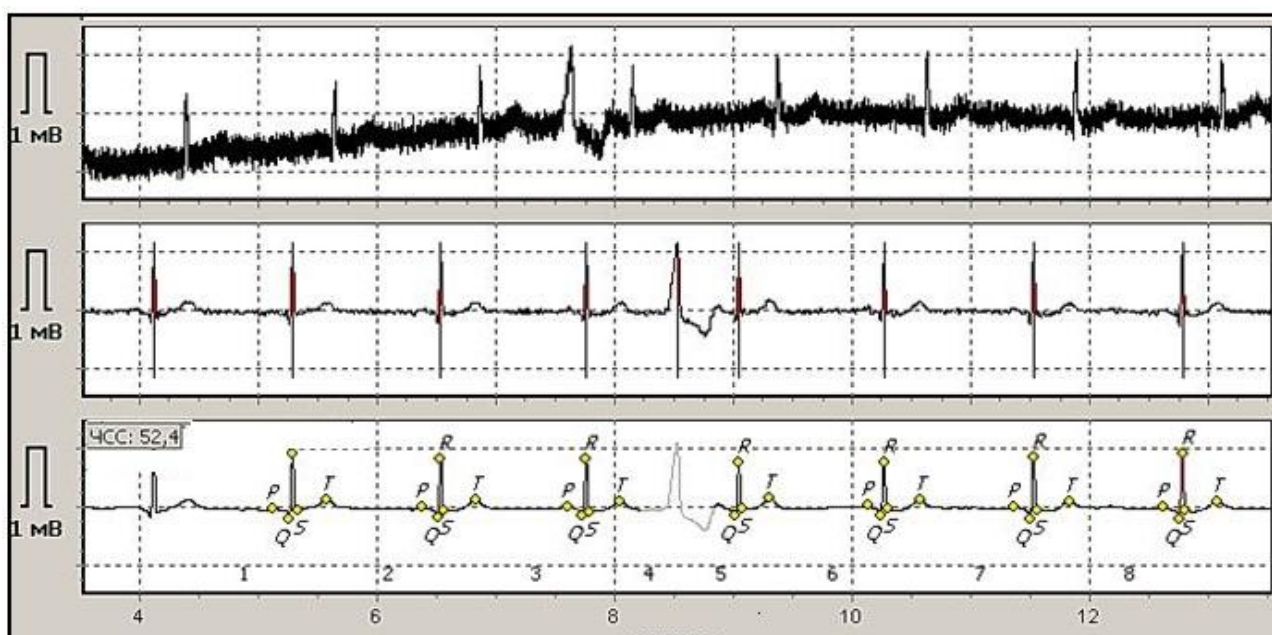


Рис. 7.7. Перша сторінка робочого вікна комплексу ФАЗАГРАФ®

У верхній частині другої сторінки (рис. 7.8) додатково відображаються графіки фазового портрету ЕКГ, фрагмента реполяризації усередненої фазової траєкторії та відновлений еталонний кардіоцикл.

На цій самій сторінці наводяться результати автоматичного обчислення амплітудно-часових параметрів зареєстрованої ЕКГ. У правій частині другої сторінки відображаються середні значення ЧСС (уд/хв), загальна кількість зареєстрованих циклів, кількість виявлених атипичних циклів, а також амплітуди та тривалості всіх виявлених зубців усередненого кардіоциклу.

Окрім значень традиційних показників відображається також середнє значення показника β та його середньоквадратичне відхилення під час спостереження (СКВ β), яке також має діагностичну інформацію про ризик розвитку ішемії міокарду.

Для наочності в середній частині вікна відображається лінія поточного значення β , що характеризує динаміку зубця T протягом реєстрації ЕКГ.

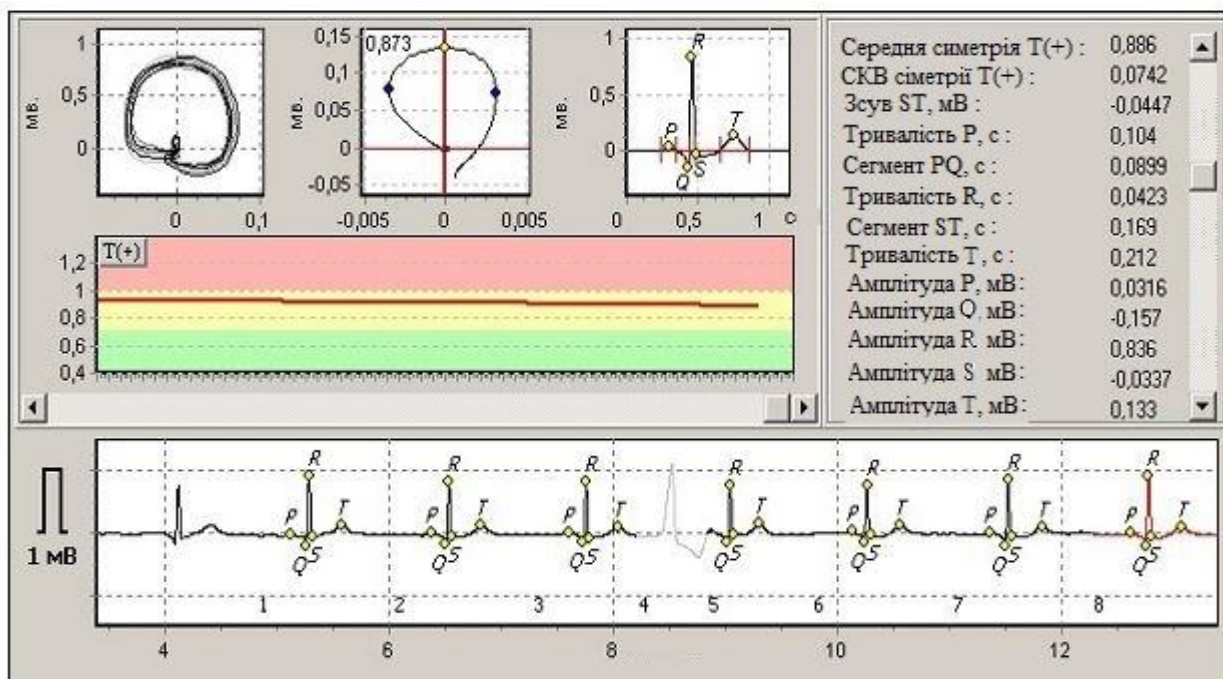


Рис. 7.8. Друга сторінка робочого вікна комплексу ФАЗАГРАФ®

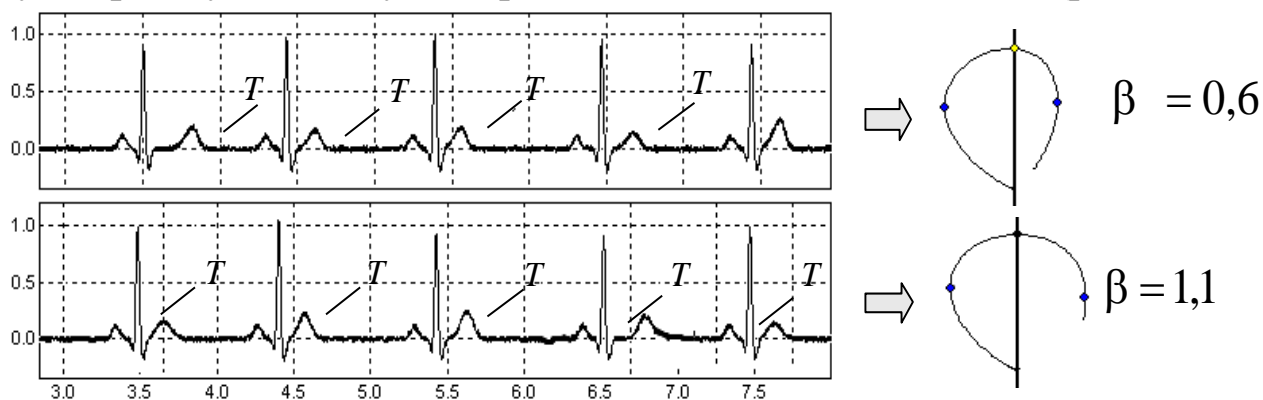
Клінічні дослідження, проведені у великих групах здорових добровольців і верифікованих хворих на ІХС показали, що оцінювання показника β підвищує чутливість і специфічність виявлення ішемії міокарда навіть у тих випадках, коли аналіз ЕКГ у 12 традиційних відведеннях є неінформативним. Експерименти на тваринах, проведені

в умовах штучної ішемії, також підтвердили високу чутливість показника до ішемії міокарда, яка більш ніж у п'ятеро перевищувала чутливість традиційного показника – депресії сегмента ST .

На підставі цих досліджень встановлено референтні інтервали показника β та реалізовано діагностичне правило:

$$\begin{aligned} & \text{НОРМА, якщо } \beta < 0,7, \\ & \text{ЗАДОВІЛЬНО, якщо } 0,7 \leq \beta \leq 1,05, \\ & \text{УВАГА, якщо } \beta > 1,05. \end{aligned} \quad (7.10)$$

Зауважимо, що ЕКГ з нормальними та патологічними значеннями показника β візуально однакові у часовій області, але під час переходу на фазову площину їхнє розпізнавання стає можливим (рис. 7.9).



ЕКГ в часовій області з нормальними (вгорі) та патологічними (внизу) зубцями T

Фрагменти реполяризації на фазовій площині

Рис. 7.9 Аналіз показника β у часовій області та на фазовій площині

Використовуючи правило (7.10), можна за допомогою ФАЗАГРАФ[®] у домашніх умовах оцінювати поточний функціональний стан серцевої діяльності та оптимізувати спосіб життя, раціонально розподіляючи режими навантажень та відпочинку.

Ще більш важливу інформацію ФАЗАГРАФ[®] надає на основі інтегральних характеристик $x_0^{(\beta)}$ і $\bar{S}^{(\beta)}$ обчислених за результатами самостійно зроблених багаторазових спостережень відповідно до запропонованого підходу. Відхилення $\Delta_\beta = \beta_T - x_0^{(\beta)}$ поточного результату β_T від еталонного результату $x_0^{(\beta)}$ конкретного користувача, яке автоматично обчислює ФАЗАГРАФ[®], дає змогу приймати персоналізовані рішення за схемою, наведеною в табл. 7.1.

Таблиця 7.1. Схема персоніфікованих рішень

Знак Δ_β	Умова	Персоніфіковані рішення	Повідомлення користувачу
+	$ \Delta_\beta > 1,5 \bar{S}^{(\beta)}$	Суттєве погіршення	Будьте уважні!
+	$0,5 \bar{S}^{(\beta)} \leq \Delta_\beta \leq 1,5 \bar{S}^{(\beta)}$	Помірне погіршення	Треба відпочити!
+ або -	$ \Delta_\beta < 0,5 \bar{S}^{(\beta)}$	Персональна норма	Це ваша норма
-	$0,5 \bar{S}^{(\beta)} \leq \Delta_\beta \leq 1,5 \bar{S}^{(\beta)}$	Помірне поліпшення	Ви в гарній формі
-	$ \Delta_\beta > 1,5 \bar{S}^{(\beta)}$	Суттєве поліпшення	Ви в чудовій формі

Клінічні дослідження показали, що 95 % записів ЕКГ здорових добровольців потрапляють у референтний інтервал $\beta_T \leq 0,87$. Це дає змогу оцінювати ризик розвитку ІХС у конкретного користувача за правилом (7.9), яке в цьому випадку мають вигляд:

$$\begin{aligned} \text{Низький ризик розвитку ІХС, якщо } x_0^{(\beta)} &\leq 0,87, \\ \text{Високий ризик розвитку ІХС, якщо } x_0^{(\beta)} &> 0,87. \end{aligned} \quad (7.11)$$

Для наочності подання інформації, результати оброблення ЕКГ відображаються на спеціальних індикаторах (рис. 7.10). Індикатор функціонального стану має вигляд термометра зі шкалою, поділеною на три зони – зелену (НОРМА: $\beta < 0,7$), жовту (ЗАДОВІЛЬНО: $0,7 \leq \beta \leq 1,05$), червону (УВАГА: $\beta > 1,05$), а також супроводжуються відповідними голосовими повідомленнями.

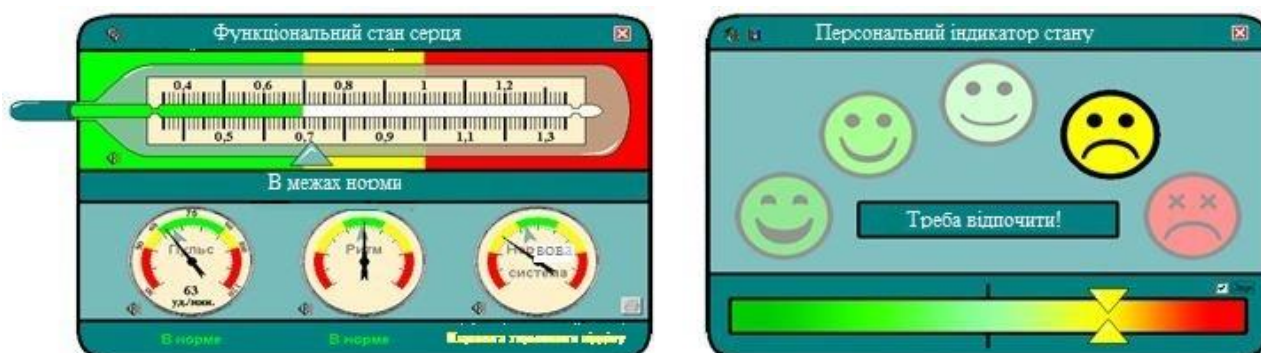


Рис. 7.10. Індикатори функціонального стану користувача

Інтегральний результат тестування надається також у вигляді розгорнутої текстової та графічної інформації (рис. 7.11).

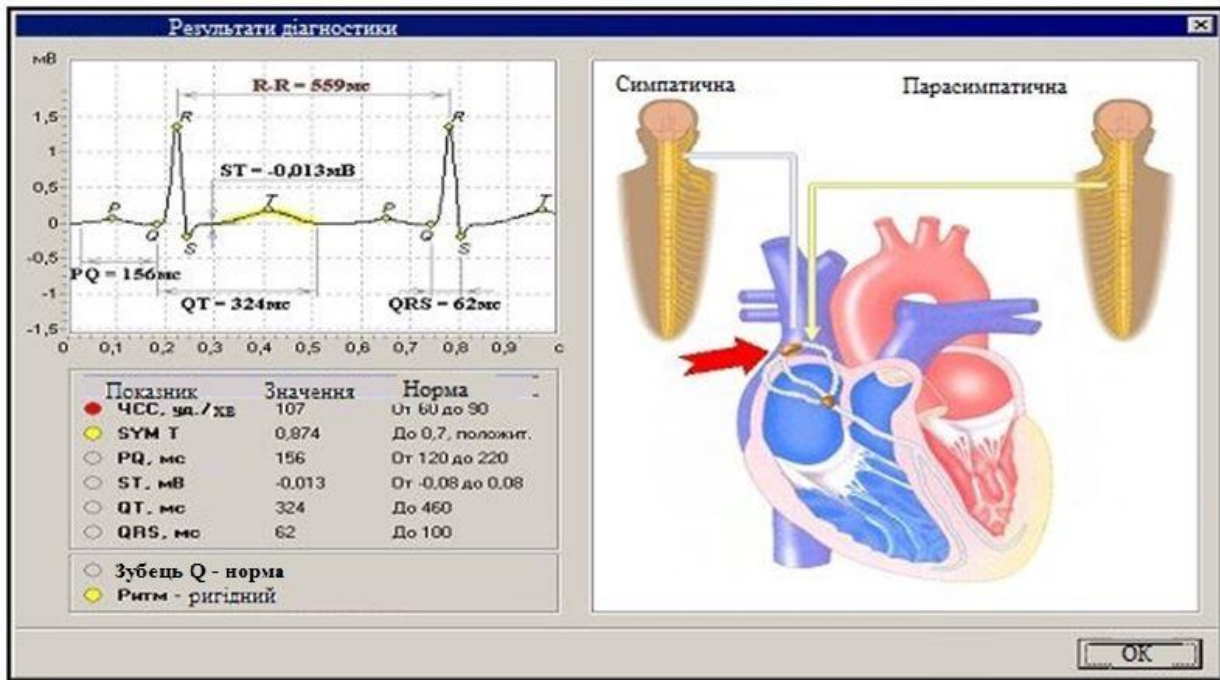


Рис. 7.11. Відображення результату тестування

Цей спосіб дає змогу на відповідних ділянках графічних зображень (мнемосхемах) серця та відділів вегетативної нервової системи відобразити виявлені відхилення від норми у вигляді миготливих елементів жовтого кольору (помірне відхилення) або червоного кольору (суттєве відхилення).

Реалізація фазаграфії на смартфоні забезпечила б не тільки зручність використання методу в польових умовах, а й віртуальний зв'язок із сімейним лікарем, якому передаються результати тестування засобами Інтернет-технологій.

Незважаючи на спрощений спосіб реєстрації ЕКГ тільки в одному відведенні ФАЗАГРАФ[®] підтвердив свою ефективність у медичних установах, спортивних організаціях, підприємствах з підвищеним техногенним ризиком, навчальних закладах (школах) та в інших сферах застосування. Нині майже 200 комплектів ФАЗАГРАФ[®] впроваджено в різних регіонах України.

Фазаграфія поклала основу новому напрямку комп'ютерних засобів превентивної медицини для домашнього застосування. За наявною інформацією, низка користувачів ФАЗАГРАФ[®], які раніше вважали себе здоровими, з його допомогою вперше дізналися про відхилення в роботі серця, які згодом було підтверджено поглибленими обстеженнями в медичних закладах.

7.3. Інтелектуальний ритмограф на смартфоні

Математичний аналіз варіабельності серцевого ритму (ВСР) дає змогу виділити цей метод, як самостійний неінвазивний метод клінічної кардіології – метод кардіоінтервалографії. Статистичні та спектральні показники ВСР відображають роботу не тільки серцево-судинної системи, а й механізмів регуляції цілісного організму – вегетативною нервовою системою.

Динамічний ряд тривалостей кардіоциклів можна оцінювати за допомогою фотоплетизмограми, яка має інформацію про кровонаповнення певної ділянки тіла. Проведені дослідження показали, що використання інтелектуальних обчислювальних алгоритмів дає змогу проводити реєстрацію фотоплетизмограми за допомогою вбудованої камери смартфона без додаткових технічних засобів.

На основі цієї ідеї створено мобільний додаток до смартфона AI-РИТМОГРАФ.

Користувач закриває камеру смартфона фалангою свого пальця, який освітлюється вбудованим ліхтариком (рис. 7.12). Формується послідовність зображень у вигляді функцій

$$\Psi_{km1}(x, y), \Psi_{km2}(x, y), \Psi_{km3}(x, y), \dots,$$

кожна з яких характеризує яскравість пікселів з координатами $x = k$, $y = m$ на площині (x, y) зображення фаланги пальця у фіксовані моменти часу $z = 1, 2, \dots$

З кожним ударом серця за рахунок зміни кровотоку в капілярах яскравість кадрів змінюється. Послідовність величин

$$g_z = \frac{1}{Q_x Q_y} \sum_{k=1}^{Q_x} \sum_{m=1}^{Q_y} \Psi_{kmz}(x, y), \quad z = 1, 2, \dots, N, \quad (7.12)$$

які характеризують середню яскравість зображень фаланги пальця, має інформацію про дискретні значення пульсової хвилі q_1, q_2, \dots, q_N (рис. 7.12). У формулі (7.12) Q_x , Q_y – кількість пікселів по горизонталі і вертикалі зображення.

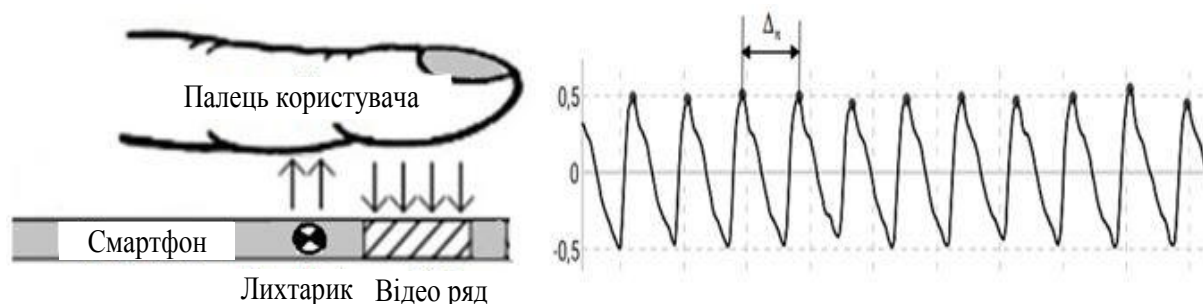


Рис. 7.12. Принцип реєстрації пульсової хвилі за допомогою вбудованої камери смартфона

Перший етап оброблення сигналу передбачає застосування процедури (3.35)–(3.37) видалення тренду, яка автоматично адаптується до послідовності q_1, q_2, \dots, q_N сигналу.

Наступний етап оброблення сигналу – визначення локальних максимумів пульсової хвилі на тлі можливих спотворень. Застосування спеціальної адаптивної процедури із заданим порогом нечутливості дає змогу визначити характерні точки пульсової хвилі, які поділяють кожен цикл на анакротичну та дикротичну фази (рис. 7.12) і побудувати вихідний динамічний ряд $\Delta_1, \Delta_2, \dots, \Delta_M$ кардіоінтервалів.

Для підвищення надійності та точності визначення показників варіабельності серцевого ритму в АІ-РИТМОГРАФ реалізовано алгоритм, який дає змогу у процесі реєстрації пульсової хвилі автоматично коригувати динамічний ряд кардіоінтервалів $\Delta_1, \Delta_2, \dots, \Delta_M$ з метою видалення одиничних артефактів.

Для цього в ковзному вікні послідовно накопичують кожен п'ятірку динамічного ряду кардіоінтервалів, які ранжують за зростанням. Поточне значення модифікованого масиву оцінюють як арифметичне середнє між другим, третім і четвертим ранжованими значеннями, тобто

$$\delta_j = \frac{d_{j-1} + d_j + d_{j+1}}{3}, \quad j = 3, \dots, M - 2. \quad (7.13)$$

Розглянута процедура поєднує переваги медіанної фільтрації та ковзного середнього. Оскільки перше та останнє значення чергової п'ятірки не бере участі у процедурі (7.13), то поодинокі викиди автоматично видаляються.

АІ-РИТМОГРАФ автоматично обчислює стандартні показники ВСР та надає результати тестування у вигляді графічних образів, прийнятих у традиційній кардіоінтервалографії (рис. 7.13).

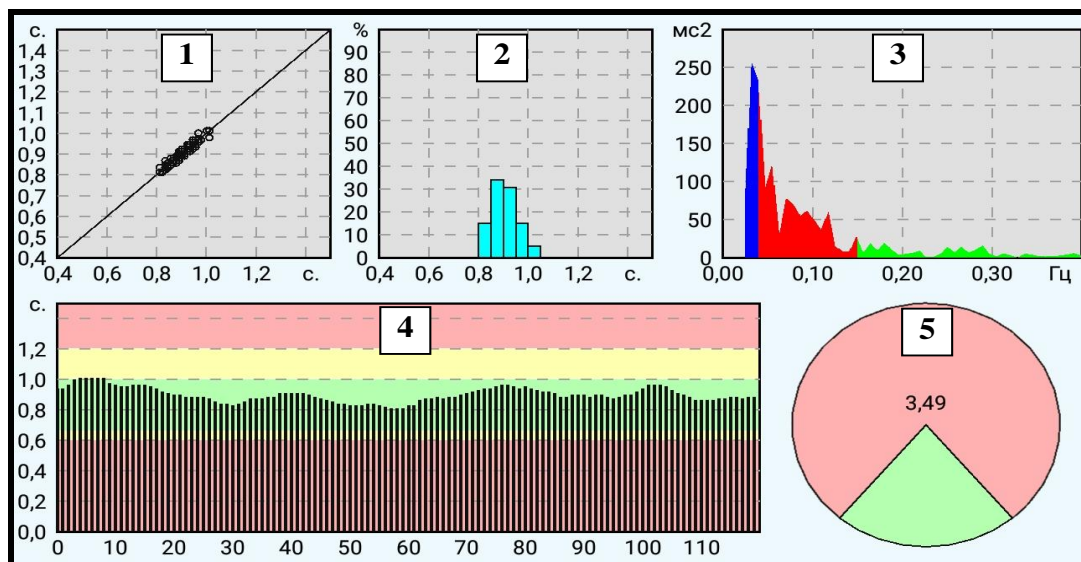


Рис. 7.13. Нормальный ритм:

1 – скатерограма, 2 – гістограма, 3 – спектрограма, 4 – ритмограма,
5 – діаграма симпато-вагусного балансу

Для самостійного оцінювання результатів користувачу надається інформація за двома основними показниками ВСР (табл. 7.2 та 7.3).

Таблиця 7.2. Інтерпретація значення показника ЧСС, уд/хв

Умова	Частота пульсу	Колір індикатора
$ЧСС < 50$	Суттєво знижена	Червоний
$50 \leq ЧСС < 60$	Помірно знижена	Жовтий
$60 \leq ЧСС \leq 90$	Нормокардія	Зелений
$90 < ЧСС \leq 100$	Помірно підвищена	Жовтий
$ЧСС > 100$	Суттєво підвищена	Червоний

Таблиця 7.3. Інтерпретація значення показника $SDNN$, мс

Умова	Варіабельність ритму	Колір індикатора
$SDNN < 30$	Суттєво знижена	Червоний
$30 \leq SDNN < 50$	Помірно знижена	Жовтий
$50 \leq SDNN \leq 70$	Нормотонія	Зелений
$70 \leq SDNN \leq 90$	Помірно підвищена	Жовтий
$SDNN > 90$	Суттєво підвищена	Червоний

AI-РИТМОГРАФ – зручний і надійний засіб превентивної медицини, що дає змогу користувачеві у домашніх та польових умовах за

допомогою смартфона отримувати важливу інформацію про поточний стан симпатичного та парасимпатичного вузлів вегетативної нервової системи та прогнозувати адаптацію на основі аналізу статистичних і спектральних показників ВСР. Для цього формуються персоналізовані рішення за правилами (7.8), (7.9) на основі обчислення індивідуальних характеристик x_0 і \bar{S} за всіма показниками ВСР.

7.4. Інтелектуальний артеріограф на смартфоні

Відомо, що визначення властивостей кровоносних судин є важливою ланкою для виявлення, профілактики та лікування захворювань серцево-судинної системи. Еластичні судини дають змогу зберігати ударний об'єм крові, зменшити навантаження на серце і забезпечити плавність протікання крові від судин великого діаметра до судин меншого діаметра. В результаті пульсуючий кровотік від серця перетворюється на безперервний потік через судинне русло, що дуже важливо для нормального функціонування організму.

Старіння організму супроводжується втратою еластичності кров'яних судин. Збільшення жорсткості артерій призводить до збільшення швидкості поширення пульсової хвилі і цей фактор у нині визнається одним із головних факторів ризику гіпертензії та виникнення ішемічної хвороби серця.

В останні роки набули поширення неінвазивні методи артеріографії, що забезпечують визначення швидкості поширення пульсової хвилі в медичних установах за допомогою спеціальної апаратури.

На основі подальшого розвитку інтелектуальних алгоритмів оброблення пальцевої фотоплетизмограми створено прилад АІ-АРТЕРІОГРАФ, який дає змогу на смартфоні в домашніх умовах інтегрально оцінювати властивості кровоносних судин. Для цього побудовано надійні обчислювальні алгоритми, за допомогою яких будується «еталонна» пульсова хвиля й автоматично визначаються характерні точки, що відповідають моментам виникнення прямої пульсової хвилі, породженої ударом серця (точка A), та зворотної пульсової хвилі (точка B), відбитої від кінцівок (рис. 7.14).

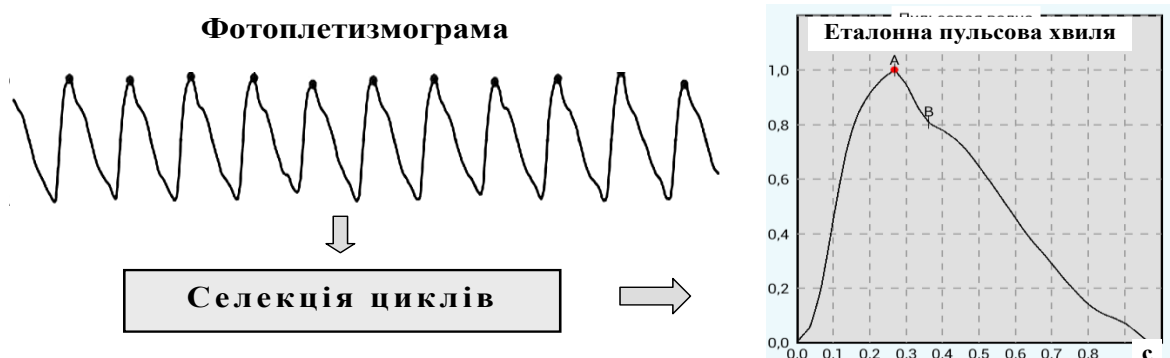


Рис. 7.14. Побудова еталонної пульсової хвилі

Для побудови еталонної пульсової хвилі проводиться автоматична селекція та видалення ненадійних циклів фотоплетизмограми з подальшим усередненням тільки надійних циклів. Для цього застосовується модифікований алгоритм селекції та усереднення циклів ЕКГ, який було розглянуто в розділі 4.

Для спрощення розрахунку відстаней між парами пульсових хвиль реалізовано процедуру модифікації фрагментів $\omega_n^{(i)}$ анакротичної ($i = I_1$) та дикротичної ($i = I_2$) фаз кожного n -го циклу фотоплетизмограми на основі операторного перетворення

$$\tilde{\omega}_n^{(i)} = a_n^{(i)} \omega_n^{(i)} \left(\frac{\theta}{b_n^{(i)}} \right), \quad n = 1, 2, \dots, \quad i \in \{I_1, I_2\}, \quad (7.14)$$

де $a_n^{(i)}$, $b_n^{(i)}$ – параметри лінійного розтягування (стиснення) за амплітудою та часом.

Для автоматичного виділення точки B на усередненій пульсовій хвилі обчислюється друга похідна сигналу на основі алгоритму визначення похідної (4.16) з наступним адаптивним згладжуванням. Ці алгоритми дають змогу надійно визначати на смартфоні точку B навіть тоді, коли її важко знайти візуально (рис. 7.15).

За еталонною пульсовою хвилею обчислюються п'ять показників, які мають діагностичну цінність, а саме:

– інтервал часу DT між точками B і A (час розповсюдження пульсової хвилі, с):

$$DT = TB - TA; \quad (7.15)$$

– відношення ΔA амплітуд сигналу в точках B і A :

$$\Delta A = \frac{AB}{AA}; \quad (7.16)$$

– відносний час ΔT розповсюдження пульсової хвилі:

$$\Delta T = \frac{DT}{T}; \quad (7.17)$$

– нахил спадання пульсової хвилі:

$$S = \frac{AA - AB}{DT}; \quad (7.18)$$

– швидкість розповсюдження пульсової хвилі (м/с):

$$V = \frac{L}{DT}, \quad (7.19)$$

де T – загальна тривалість усередненої пульсової хвилі, а L – довжина шляху, яким проходить пульсова хвиля за час DT .

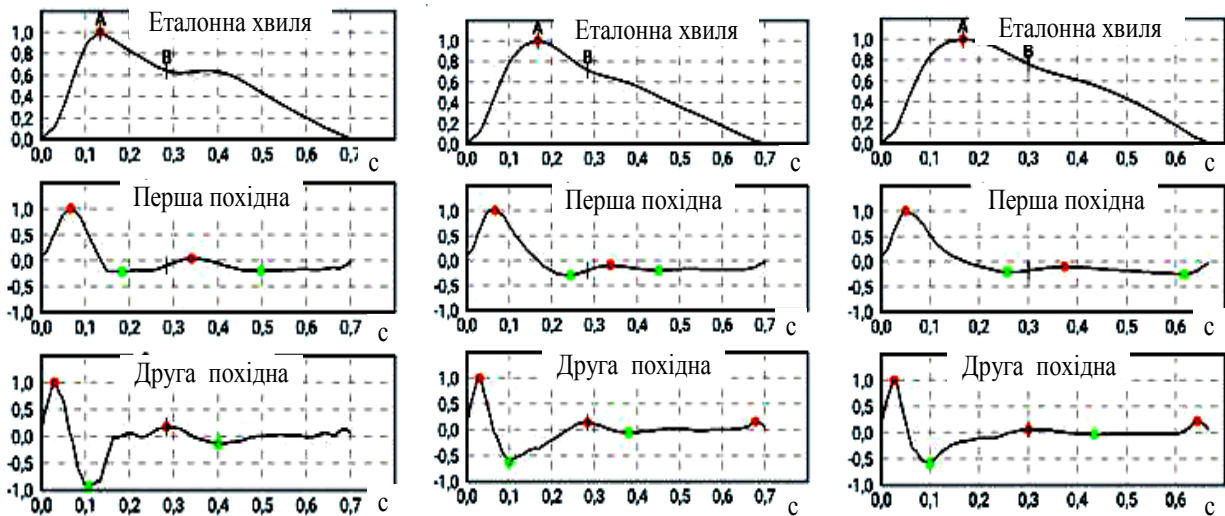


Рис. 7.15. Еталонні пульсові хвилі та їх похідні

Величину L автоматично обчислюють за співвідношенням, отриманим на основі стандартних пропорцій тіла людини з урахуванням його росту.

Обчисливши інтегральні характеристики x_0 і \bar{S} за показниками (7.15)–(7.19), можна приймати персоніфіковані рішення. Зокрема, за величиною $x_0^{(V)}$, що характеризує середні значення швидкості поширення пульсової хвилі V , пацієнт може оцінити стан кровоносних су-

дин на основі порівняння $x_0^{(V)}$ з референтними діапазонами, що відповідають його віку.

Статистичне оброблення понад 1000 фотоплетизмограм, отриманих від 30 волонтерів обох статей віком від 20 до 80 років, підтвердила, що обчислена згідно з (7.19) швидкість розповсюдження пульсової хвилі V (м/с) і вік H (років) випробуваного з коефіцієнтом кореляції $r = 0,8$ описує лінійне рівняння регресії

$$V = 5,1807 + 0,0671H, \quad (7.20)$$

що узгоджується з відомими результатами медичних досліджень.

7.5. Інтелектуальний тонометр для домашнього застосування

Артеріальна гіпертензія – одне з найбільш поширених захворювань серцево-судинної системи, яким страждають 30 % дорослого населення, а з віком поширеність хвороби збільшується і досягає 65 % в осіб старше 65 років. Якщо діагностику та лікування захворювання проводити несвоєчасно, то можуть виникнути важкі ускладнення – інфаркт міокарда та інсульт головного мозку, які часто закінчуються летальним кінцем або інвалідністю пацієнта.

Цифрові тонометри – одні з перших засобів цифрової медицини для домашнього застосування. На першому етапі становлення ринку медики висловлювали свої побоювання про можливі негативні наслідки застосування цих виробів, але нині такі тонометри використовуються майже в кожній родині.

Побутові тонометри реалізують осцилометричний метод вимірювання, який ґрунтується на реєстрації амплітуди пульсацій тиску повітря в момент проходження крові через здавлену манжетою ділянку артерії. Метод дає змогу автоматично вимірювати артеріальний тиск навіть за слабкими тонами Короткова, за наявності феномена «аускультативного провалу» та інших ефектах, що становлять труднощі процесу автоматизації вимірювання за методом Короткова.

Більшість наявних на ринку домашніх цифрових тонометрів надають користувачеві значення трьох показників: систолічного артеріального тиску SBP , діастолічного артеріального тиску DBP і частоти пульсу HR . Ці показники можуть зберігатися у внутрішній пам'яті.

Водночас інтелектуальні можливості домашніх тонометрів ще не достатньо вичерпані. Стрімкий розвиток мікроелектроніки та інтеле-

ктуальних методів оброблення сигналів дає змогу вже сьогодні реалізувати в домашніх тонометрах низку важливих додаткових функцій.

Розглянемо одну з таких можливостей – оцінювання варіабельності артеріального тиску.

Розглянемо значення систолічного артеріального тиску SBP_n , $n = 1, 2, \dots$, які спостерігалися у конкретного користувача за досить великий проміжок часу (тижні, місяці, роки), як реалізації випадкової величини P з розподілом ймовірності \mathfrak{R}_{SBP} .

Позначимо носій цього розподілу множиною

$$\Omega_{SBP} = \{P : \mathfrak{R}_{SBP} > 0\}, \quad (7.21)$$

а M_{SBP} – середнє значення SBP_n , $n = 1, 2, \dots$

Нехай $\Omega_{SBP}^{(0)} = [100, 140]$, мм рт. ст. – референтний діапазон прийнятих у медичній практиці нормальних значень SBP . Для спрощення міркувань не будемо далі розрізняти діапазони норм різних вікових і гендерних груп.

Ґрунтуючись на загальному підході до аналізу взаємного розміщення популяційної Ω_0 та персоніфікованої Ω_i областей у просторі ознак розглянемо такі варіанти взаємного розміщення Ω_{SBP} та $\Omega_{SBP}^{(0)}$ відносно осі значень систолічного артеріального тиску (рис. 7.16).

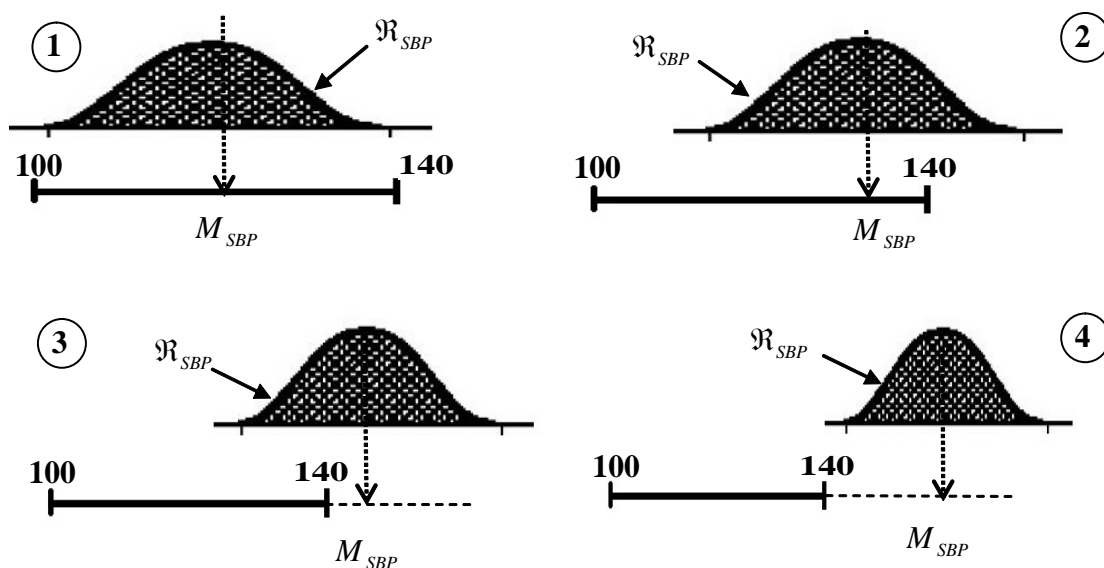


Рис. 7.16. Взаємне розміщення областей Ω_{SBP} та $\Omega_{SBP}^{(0)}$

Варіант 1. $\Omega_{SBP} \subset \Omega_{SBP}^{(0)}$, тобто область Ω_{SBP} результатів вимірювання повністю входить у референтний інтервал $\Omega_{SBP}^{(0)} = [100,140]$ нормальних значень систолічного артеріального тиску.

Варіант 2. $(\Omega_{SBP} \cap \Omega_{SBP}^{(0)}) \neq \emptyset$, $\Omega_{SBP} / (\Omega_{SBP} \cap \Omega_{SBP}^{(0)}) \neq \emptyset$, тобто область Ω_{SBP} лише частково входить в область $\Omega_{SBP}^{(0)}$, але $M_{SBP} \in \Omega_{SBP}^{(0)}$, тобто середні результати вимірювань належить референтному інтервалу $\Omega_{SBP}^{(0)} = [100,140]$ нормальних значень систолічного артеріального тиску.

Варіант 3. $(\Omega_{SBP} \cap \Omega_{SBP}^{(0)}) \neq \emptyset$, $\Omega_{SBP} / (\Omega_{SBP} \cap \Omega_{SBP}^{(0)}) \neq \emptyset$, тобто область Ω_{SBP} лише частково входить в область $\Omega_{SBP}^{(0)}$, але $M_{SBP} \notin \Omega_{SBP}^{(0)}$, тобто середні результати вимірювань не належить референтному інтервалу $\Omega_{SBP}^{(0)} = [100,140]$ нормальних значень систолічного артеріального тиску.

Варіант 4. $(\Omega_{SBP} \cap \Omega_{SBP}^{(0)}) = \emptyset$, тобто область Ω_{SBP} результатів вимірювань виходить за межі референтного інтервалу $\Omega_{SBP}^{(0)} = [100,140]$ нормальних значень систолічного артеріального тиску.

Якщо деякі результати вимірювань виходять за межі інтервалу $\Omega_{SBP}^{(0)} = [100,140]$, але середнє значення M_{SBP} входить у цей інтервал, то такого пацієнта будемо вважати умовно здоровим зі схильністю до гіпертонії.

Зауважимо, що для практичної реалізації запропонованого підходу зовсім не обов'язково обробляти весь масив вимірюваних значень SBP . Достатньо за кожним наступним вимірюванням SBP_n , $n = 1, 2, \dots$ лише коригувати мінімальне SBP_{\min} та максимальне SBP_{\max} значення отриманих результатів за схемою

$$SBP_{\min, n} = SPP_n, \quad \text{якщо} \quad SPP_n < SBP_{\min, n-1}, \quad (7.22)$$

$$SBP_{\max, n} = SPP_n, \quad \text{якщо} \quad SPP_n > SBP_{\max, n-1} \quad (7.23)$$

та уточнювати поточне середнє значення M_{SBP} за допомогою рекурентної формули, яка має вигляд

$$M_{SBP, n} = M_{SBP, n-1} + \frac{1}{n} (SBP_n - M_{SBP, n-1}), \quad (7.24)$$

поставивши початкові умови $M_{SBP, 0} = 0$ і $SBP_{\min, 0} = SBP_{\max, 0} = SBP_0$.

Аналогічно можна переоцінювати поточне значення стандартного відхилення σ_{SBP} систолічного артеріального тиску.

Після кожного чергового вимірювання можна також обчислювати поточні значення коефіцієнта варіації Пірсона

$$V_{SBP} = \frac{\sigma_{SBM}}{M_{SBM}} \cdot 100 \% \quad (7.25)$$

та індексу

$$I_{SBP} = \frac{N_{SBP}^{(E)}}{N} \cdot 100 \% , \quad (7.26)$$

який характеризує відсоток кількості вимірювань $N_{SBP}^{(E)}$, за якими систолічний артеріальний тиск перевершував поріг $SBP = 140$ мм рт. ст.






Для отримання персоніфікованих рішень будемо за кожним поточним результатом SBP_n оцінювати величину

$$\Delta_{SBP,n} = SBP_n - M_{SBP}, \quad n = 1, 2, \dots , \quad (7.27)$$

яка характеризує відхилення SBP_n від знайденого середнього значення M_{SBP} , та порівнювати $\Delta_{SBP,n}$ з поточним значенням середньоквадратичного відхилення σ_{SBP} . Це дає змогу додатково відображати на екрані тонометра якісну інформацію про поточний функціональний стан пацієнта у вигляді наочних графічних образів (смайликів).

Один з варіантів формування такої інформації за умови, що $SBP > 100$ мм рт. ст, наведено в табл. 7.4.

Таблиця 7.4. Варіант формування персоніфікованих рішень

Знак Δ_{SBP}	Умова	Повідомлення	Індикатор
+	$\Delta_{SBP} > 1,5\sigma_{SBP}$	Небезпечний стан!	
+	$0,5\sigma_{SBP} \leq \Delta_{SBP} \leq 1,5\sigma_{SBP}$	Будьте уважні!	
+ або -	$ \Delta_{SBP} < 0,5\sigma_{SBP}$	Стабільний стан	
-	$0,5\sigma_{SBP} \leq \Delta_{SBP} \leq 1,5\sigma_{SBP}$	Гарний стан	
-	$\Delta_{SBP} > 1,5\sigma_{SBP}$	Ви в чудовій формі	

Аналогічно можна надавати персоніфіковані рішення про результати вимірювання діастолічного тиску DBP_n , $n = 1, 2, \dots$.

Таким чином, незначні доробки домашнього тонометра забезпечують персоніфіковану діагностику і дають змогу оцінювати довгострокову варіабельність артеріального тиску (ВР) між відвідуваннями лікаря (**visit-to-visit variability**).

Якісне оцінювання результатів вимірювання артеріального тиску допомагає користувачеві визначати необхідність додаткового прийому призначених лікарем ліків, а оцінювання індивідуальних характеристик артеріального тиску надати лікарю більш повну інформацію про пацієнта для прийняття обґрунтованих діагностичних рішень.

7.6. Інтелектуальний стетоскоп на смартфоні

Гострі респіраторні захворювання – найпоширеніші і соціально значимі хвороби, з якими останнім часом все частіше стикається населення. Масовий характер цих захворювань передбачає розподілену систему надання медичних послуг, коли домашнє спостереження і лікування набуває важливого значення.

Така задача особливо актуальна у зв'язку з пандемією COVID-19, оскільки, з одного боку, важливо своєчасно діагностувати та розпочати лікування пацієнта, якщо є загроза розвитку вірусної пневмонії, тобто мінімізувати ймовірність пропуску цілі, а, з другого боку, запобігти необґрунтованим відвідуванням медичних установ, які несуть небезпеку контакту здорового пацієнта з можливими носіями коронавірусної інфекції, тобто, мінімізувати ймовірність хибної тривоги.

Коротко опишемо інформацію про інтелектуальну ІТ, котра дає змогу в домашніх умовах сигналізувати про можливі порушення респіраторного дихання користувача і необхідність відвідування лікаря для більш повного обстеження.

Респіраторні шуми дихання користувача реєструються за допомогою вбудованого мікрофона смартфона (рис. 7.17).

На основі алгоритму локального перетворення Фур'є (**Short-Time Fourier Transformation**) за звуковими файлами будуються фоноспіворограми

$$\Psi_1, \dots, \Psi_N,$$

кожна характеризує рівень Ψ звукового сигналу частотою $f \in F$ в момент часу $t \in T$:

$$\Psi = \Psi(f, t), \quad (7.28)$$

де $F = [f_1, f_2]$ діапазон реєстрованих частот в інтервалі спостережень $T = [t_1, t_2]$.



Рис. 7.17. Сигналізатор респіраторних порушень на смартфоні

За навчальною вибіркою спостережень (фоноспірограм) $\Psi_1, \Psi_2, \dots, \Psi_N$ респіраторного дихання, зареєстрованою в нормальному функціональному стані, будується еталон нормального дихання конкретного пацієнта. Близькість двох фоноспірограм Ψ_μ та Ψ_ν з навчальної вибірки оцінює величина

$$L_{ij} = \min_{\rho=0, \dots, \Theta} \frac{1}{K} \sum_{k=1}^K |\Psi_\mu^{(k)}(f, t) - \Psi_\nu^{(k)}(f, t - \rho)|, \quad (7.29)$$

яка характеризує середню різницю енергії звуку $\forall k \in F \times T$, де Θ – максимально припустимий зсув у часі характерних точок фоноспірограм.

За матрицею парних відстаней $L_{\mu\nu}$ між μ -ю ($\mu = 1, \dots, N$) та ν -ю ($\nu = 1, \dots, N$) фоноспірограми Ψ_μ та Ψ_ν визначають персональні харак-

теристики – еталонну фоноспірограму x_0 і середню відстань \bar{S} між фоноспірограмами навчальної вибірки конкретного користувача.

Згідно з загальним підходом до персоніфікованих рішень у цьому випадку діагностичне правило має вигляд:

$$\begin{aligned} &\text{Персональна норма, якщо } S(x_t, x_0) \leq \lambda \bar{S}, \\ &\text{Ознаки респіраторного порушення, якщо } S(x_t, x_0) > \lambda \bar{S}, \end{aligned} \quad (7.30)$$

де x_t – поточна фоноспірограма, а $\lambda \geq 1$ коефіцієнт, який визначає допустиме відхилення поточної фоноспірограми від еталонної.

Звісно правило (7.30) не дає можливість класифікувати тип порушення, а лише сигналізує пацієнту про необхідність звернення до медичного закладу для здобуття кваліфікованої медичної допомоги.

Завдання до комп'ютерного практикуму

Завдання 7.1. Необхідно розробити програму для визначення та відображення статистичних характеристик параметрів серцевого ритму у вигляді скатерограми та гістограми.

Під час виконання практикуму необхідно:

- ввести дискретні значення ритмограми (послідовності RR -інтервалів) з текстового файлу РИТМ.txt. У файлі записані дискретні відліки ритмограми, які розміщено в одну колонку (зверху до низу);
- побудувати скатерограму RR -інтервалів, яка є графічним відображенням послідовних пар кардіоінтервалів (попереднього та наступного) у двовимірній координатній площині (рис. 7.18, ліворуч);
- побудувати гістограму RR , яка відображає розподіл значень RR -інтервалів (рис. 7.18, праворуч);
- обчислити значення індексу напруги за формулою:

$$IH = \frac{AMO}{2M_0 \cdot \Delta R}, \quad (7.31)$$

де M_0 , с та AMO , % – відповідно мода і амплітуда моди ритмограми, а ΔR , с – різниця між максимальним та мінімальним значеннями RR -інтервалів.

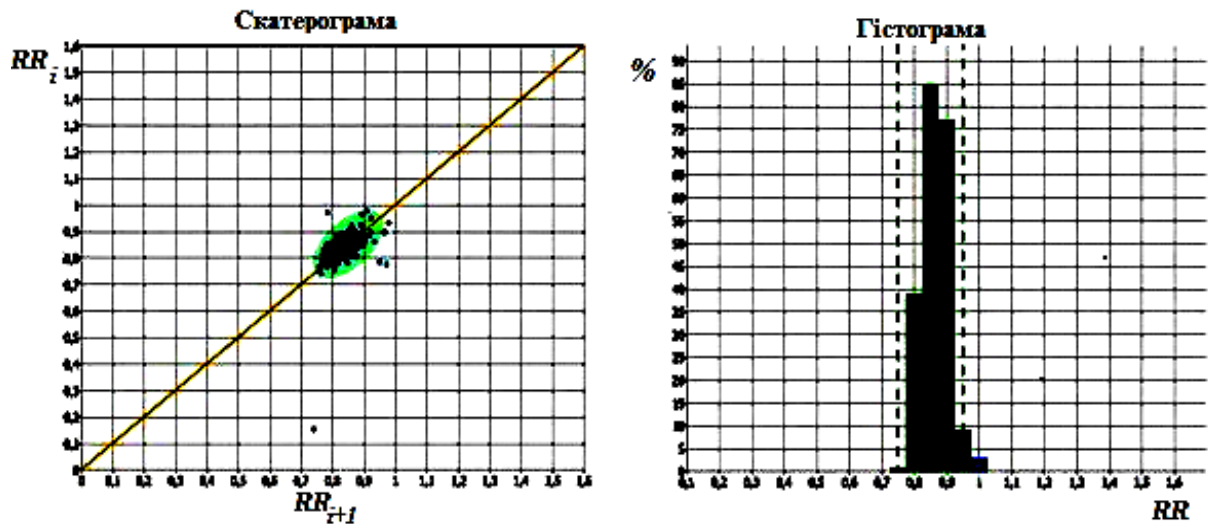


Рис. 7.18. Графічне відображення результатів завдання 7.1.

Запитання для самоконтролю

1. Сформулюйте основний принцип персоніфікованої медицини.
2. Охарактеризуйте підхід до створення інтелектуальних систем превентивної медицини.
3. Як розраховують індивідуальні інтегральні характеристики конкретної людини?
4. Наведіть приклади інтелектуальних систем превентивної медицини.

РОЗДІЛ 8

ІНТЕЛЕКТУАЛЬНІ БІОМЕТРИЧНІ СИСТЕМИ

8.1. Загальні характеристики біометрії

Інформація про штучний інтелект буде не повною, якщо не розглянути ще один важливий напрям застосування інтелектуальні систем – ідентифікація та верифікація людини.

Сучасний рівень розвитку ІТ та зростання вимог до забезпечення інформаційної безпеки призвели до того, що традиційні засоби ідентифікації (коди, ключі, карти доступу) перестають бути ефективними. З'явився новий клас систем – біометричні системи, які ґрунтуються на індивідуальних характеристиках конкретної людини

Біометричні методи є ефективним інструментом ідентифікації особистості і застосовуються для захисту від несанкціонованого доступу, контролю прибуття співробітника на роботу, пошуку злочинців, виявлення терористів тощо. Біометричними ознаками є вимірювальні фізичні характеристики особистості або особливості її поведінки, які є унікальними для конкретної людини, вони завжди є безпосередньо з суб'єктом і їх не потрібно запам'ятовувати, як пароль у традиційних системах.

Безпечність доступу до об'єкта характеризують два показника: FAR (False Acceptance Rate) – ймовірністю помилкової відмови в доступі людині, яка насправді має допуск (помилка I роду), та FRR (False Rejection Rate) – ймовірністю помилкового збігу біометричних характеристик двох людей (помилка II роду). Вибір необхідного співвідношення між показниками FAR і FRR є компромісним рішенням, прийняття якого залежить від багатьох факторів.

Крім показниками FAR і FRR до характеристик біометричних систем належать стійкість до муляжу (імітації біометричних параметрів), швидкість роботи, простота використання та вартість системи.

Біометричні ознаки поділяються на дві групи: фізіологічні та поведінкові. До першої групи належать: відбиток пальця (дактилоско-

пічні дані), цифрове зображення обличчя у 2D або 3D-проекціях, зображення райдужної оболонки або сітківки ока, вен руки, голос, форма вуха та інші. Поведінковими параметрами є динаміка підпису, клавіатурний почерк, манера ходьби тощо.

Дактилоскопія є одним із найпопулярніших методів, яка вже досить давно реалізується за допомогою комп'ютерних алгоритмів. Спеціальні сенсори роблять високоякісний цифровий знімок папілярного візерунка пальця, який потім перетворюється на код, що зберігається в базі даних.

Метод є недорогим, швидким та досить надійним. На ринку є велика кількість сканерів відбитків і спеціалізованих програм для обробки. Сучасні сканери мають розширений функціонал: можуть враховувати рельєф ліній, силу натискання і температуру, що підвищує достовірність ідентифікації. Недоліком методу є те, що на пальцях часто виникають порізи і подряпини, які можуть утруднити розпізнавання відбитка пальця.

Другий за популярністю метод – ідентифікація за геометрією обличчя. Метод ґрунтується на розпізнаванні осіб, які потрапляють у кадр під час використання систем відеоспостереження. Такі системи широко використовуються у місцях великого скупчення людей: на вокзалах, у метро, в аеропортах. Ефективність методу залежить від параметрів відеокамери, освітлення, швидкості переміщення, віддаленості об'єкта спостереження. Виникнення сучасних методів розпізнавання образів і створення 3D-зображень привело до поліпшення якісних показників цього методу, але разом з тим і до дорожнечі обладнання.

Ідентифікація по райдужній оболонці очей багатьма експертами визнається найточнішим способом встановлення особистості. Малюнок райдужної оболонки майже не змінюється протягом життя людини, для проведення процедури ідентифікації не потрібен фізичний контакт з пристроєм, оскільки зчитування проводиться на комфортній відстані від очей. Сканери захоплюють зображення, оцифровують його і порівнюють із зображеннями з бази даних.

Такі системи мають високу надійність, швидкість і точність, яка істотно залежить від якості камери. До недоліків належить дорожнеча обладнання. Проте через високу точність такі системи використо-

вуються організаціями, у сферу діяльності яких входить робота з секретними документами.

Найкращими показниками за точністю розпізнавання має метод сканування сітківки ока. Метод заснований на отриманні зображення кровоносних судин на задній стінці ока. Недоліки методу: складна процедура отримання зображення судин вимагає досить довгої нерухомості досліджуваного суб'єкта, що викликає дискомфорт. Вартість систем досить висока. Потрібен довгий час для отримання та оброблення зображення.

Метод розпізнавання за малюнком вен руки є досить точним, має високу достовірність. Це одна з найновіших технологій. Однак у методу є і недоліки: деякі захворювання можуть призвести до зміни малюнка вен. Крім того, наявні сканери дуже чутливі до сонячних променів та інших перешкод.

Вибір необхідного співвідношення між показниками FAR і FRR є компромісним рішенням, прийняття якого залежить від багатьох факторів. В табл. 8.1. наведено відомі з літератури значення FAR та FRR різних систем біометрії

Таблиця 8.1. Порівняльні характеристики біометричних методів

Біометрична ознака	Умови тестування	FRR	FAR
Відбиток пальців	Популяція, яка включала робітників ручної праці та літніх осіб	2,2 %	2,2 %
Обличчя	Поліцейська база фотографій	4,0 %	0,1 %
Голос	Мовлення різних текстів	3,0 %	1,0 %
Райдужна оболонка ока	Контрольоване освітлення	1,1 %	0,1 %

Фахівці постійно шукають нові напрямки систем біометрії, засновані на аналізі таких індивідуальних характеристик людини, які менш схильні до можливих підробок. Однією з таких характеристик є електрокардіограма, яка несе інформацію про електричну активність серця.

8.2. Фазовий портрет ЕКГ як засіб біометрії

У 1995 році на шостій Міжнародній конференції «Computer analysis of Images and Patterns» (CAIP'95), яка проходила в м. Празі, українські вчені зробили доповідь «Computer Analysis and Recognition of Cognitive Phase Space Electro-Cardio Graphic Image». В цій доповіді вперше була сформульована гіпотеза про можливість ідентифікації особистості за електрокардіограмою.

Гіпотеза ґрунтувалася на результатах експериментів, які демонстрували індивідуальні особливості фазового портрету ЕКГ. Згодом можливість ідентифікації особистості за ЕКГ-сигналами в часовій області активно досліджувались іншими авторами, які для оброблення сигналу застосовували віконне перетворення Фур'є (**Short-Time Fourier Transformation**), вейвлет перетворення (**Wavelet transform**), перетворення Радона (**Radon transform**), калманівської фільтрації (**Kalman filter**) та інші математичні методи.

Як видно з рис. 8.1, індивідуальні особливості ЕКГ наочно виявляються під час її відображення на фазовій площині з координатами $z(t)$, $\dot{z}(t)$, де $z(t)$ – сигнал про електричну активність серця, а $\dot{z}(t)$ – швидкість зміни сигналу.

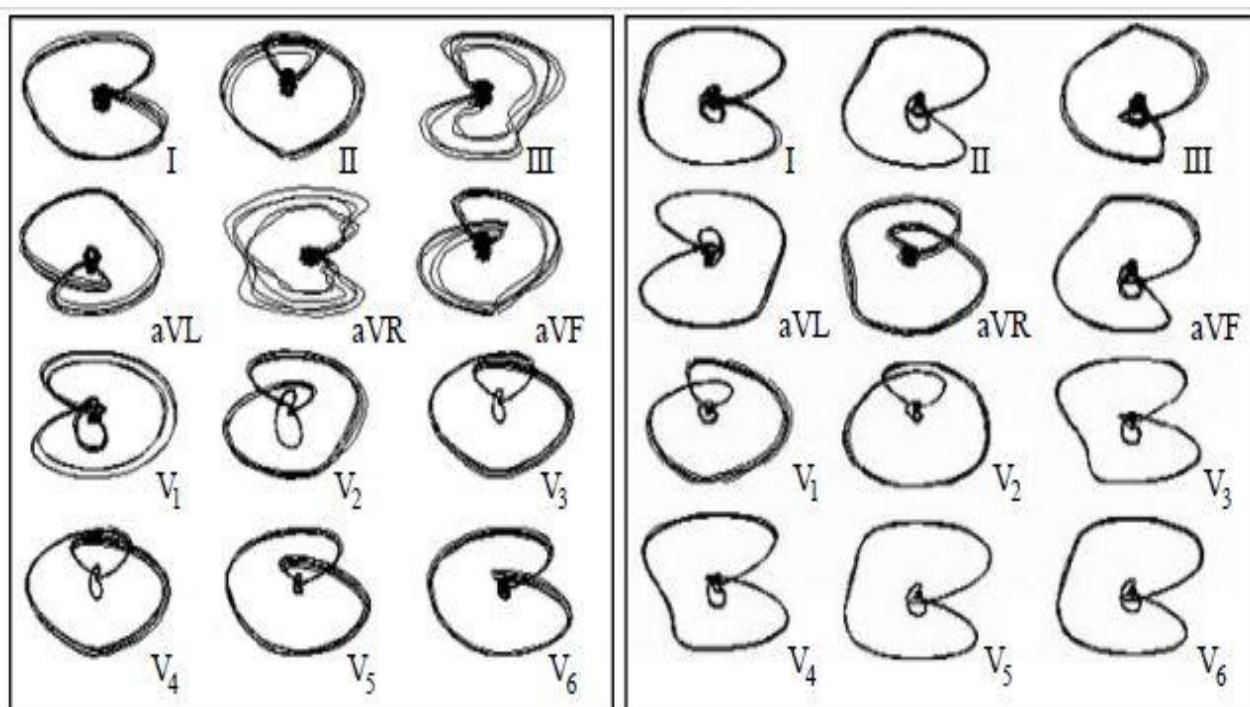


Рис. 8.1. Фазові портрети ЕКГ двох осіб (ліворуч та праворуч) у 12 традиційних відведеннях

Оскільки реєстрація сигналу одночасно в кількох відведеннях, особливо в грудних відведеннях V_1, \dots, V_6 , незручна для вирішення задач біометрії, розглянемо більш практичний спосіб, що передбачає відображення фазового ЕКГ, що реєструють лише в одному відведенні (першому стандартному) за допомогою сенсора ФАЗАГРАФ[®] з бездротовим інтерфейсом передачі даних Bluetooth (рис. 8.2).

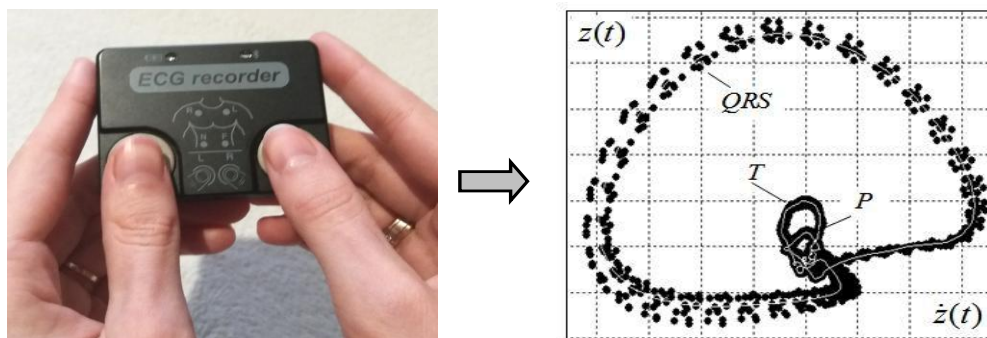


Рис. 8.2. Формування фазового портрета ЕКГ:

а – сенсор ЕКГ, б – фазовий портрет на площині $z(t), \dot{z}(t)$

Експериментальні дослідження показали, що відмінність фазових портретів у координатах $z(t), \dot{z}(t)$ більш виражена, ніж відмінність сигналів $z(t)$ у часовій області, на основі яких сформовані ці фазові портрети. Для ілюстрації на рис. 8.3 чотири різні записи ЕКГ у часовій області та у фазових координатах.



Рис. 8.3. Порівняння графічних образів чотирьох записів ЕКГ
а – у часовій області, б – на фазовій площині $z(t), \dot{z}(t)$

Будемо розрізняти дві постановки задачі біометрії.

Задача ідентифікації. Існує обмежений список $B = \{B_1, \dots, B_M\}$ осіб, що тестуються. Потрібно з використанням автоматичного оброблення фазового портрету одного з осіб $B_m \in B$ визначити його номер $m = 1, \dots, M$ у списку $B = \{B_1, \dots, B_M\}$.

Задача верифікації (аутентифікації). Людина, яка тестується, повідомляє своє прізвище (login), що визначатиме його номер $m = 1, \dots, M$ у списку $B = \{B_1, \dots, B_M\}$. Потрібно з використанням автоматичного оброблення фазового портрета ЕКГ («інтелектуального» пароля) підтвердити або спростувати названий login. Крім того, допускається, що прізвища цієї людини може не бути у списку $B = \{B_1, \dots, B_M\}$.

Багаторічні спостереження показали, що фазові портрети одноканальної ЕКГ на кшталт відбитків пальців мають індивідуальні особливості (рис. 8.4), які зберігаються протягом великого проміжку часу (рис. 8.5), якщо, звичайно, в цей період не відбулися серйозні органічні ураження серця. З огляду на ці факти було побудовано біометричні системи, що ґрунтуються на аналізі фазових портретів ЕКГ.

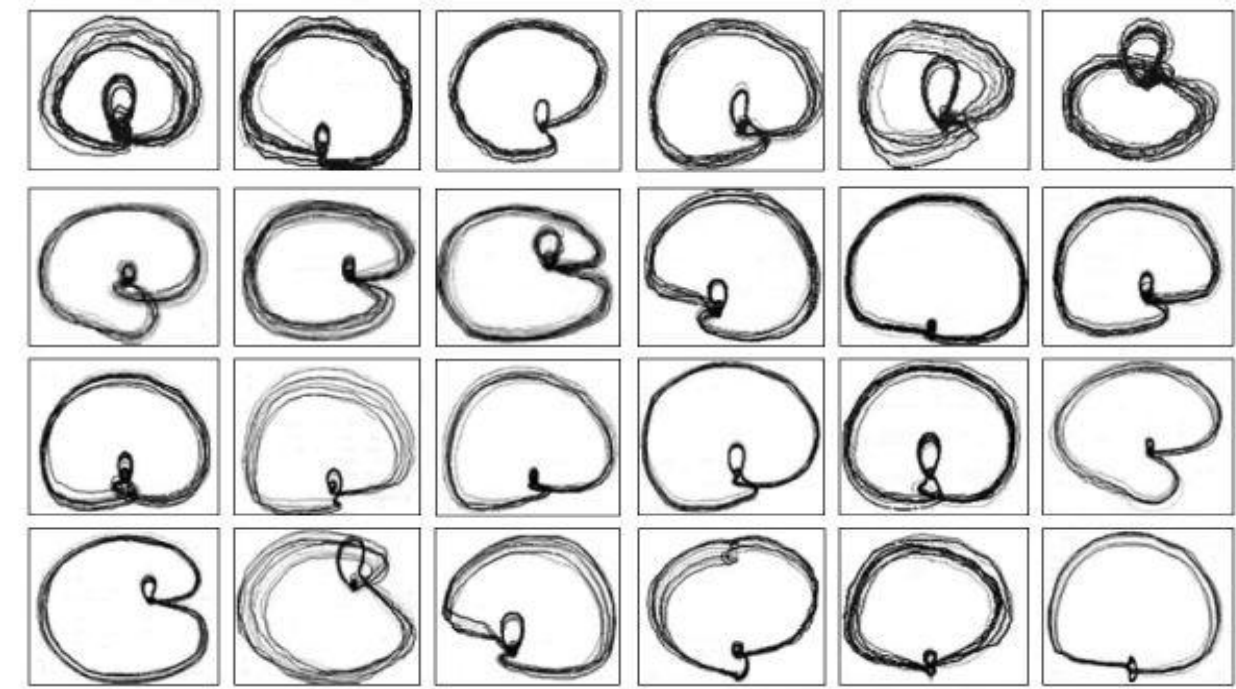


Рис. 8.4. Приклади фазових портретів одноканальної ЕКГ 24 осіб

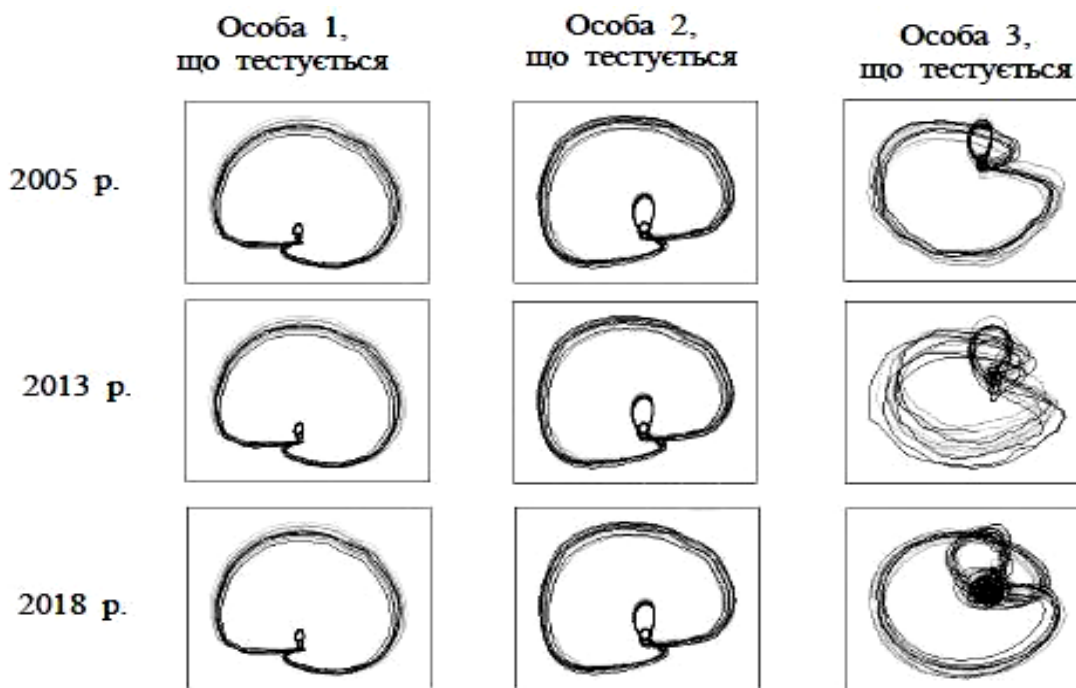


Рис. 8.5. Динаміка фазових портретів одноканальної ЕКГ трьох осіб

8.3. Ідентифікація людини за фазовим портретом одноканальної ЕКГ

Розв'язування задачі ідентифікації ґрунтується на порівнянні фазового портрета поточної ЕКГ особи V_m , що тестується, із заздалегідь побудованими еталонами обмеженої групи V_1, \dots, V_M осіб. Оскільки під час фізичних або емоційних навантажень відбуваються зміни окремих ділянок ЕКГ, зокрема, зубця T , далі будемо аналізувати найбільш стабільний фрагмент фазового портрета. Експерименти показали, що таким є фрагмент, відповідний шлунквовому комплексу QRS .

Для оцінювання близькості двох фазових портретів F_i і F_j будемо обчислювати відстані Гаусдорфа

$$L_H(F_i, F_j) = \max \left\{ \max_{q_k \in F_i} \min_{q_r \in F_j} \rho(q_k, q_r), \max_{q_r \in F_j} \min_{q_k \in F_i} \rho(q_k, q_r) \right\}, \quad (8.1)$$

де $\rho(q_k, q_r) = \|q_k - q_r\|$ – евклідова відстань між нормованими векторами $q_k = (z_k, \dot{z}_k) \in F_i^{(QRS)}$, $k = 1, \dots, K_i^{(QRS)}$, та $q_r = (z_r, \dot{z}_r) \in F_j^{(QRS)}$,

$r = 1, \dots, K_j^{(QRS)}$, що належать QRS -фрагментам i -го та j -го фазових портретів.

Для побудови еталона фазового портрета ЕКГ (ФП ЕКГ) m -го, $m = 1, \dots, M$ представника групи B_1, \dots, B_M зареєструємо $W^{(m)}$ його ЕКГ у різний час доби і сформуємо матрицю $W^{(m)} \times W^{(m)}$ Гаусдорфових відстаней (8.1) між парами QRS -фрагментів відповідних фазових портретів $F_1^{(m)}, \dots, F_W^{(m)}$. Номер рядка цієї матриці, сума елементів якої мінімальна, визначає еталонний фазовий портрет $F_0^{(m)}$, тобто

$$F_0^{(m)} = \arg \min_{1 \leq i \leq W^{(m)}} \sum_{j=1}^{W^{(m)}} L_H(F_i, F_j). \quad (8.2)$$

Рішення на користь m -го представника групи приймають лише в тому випадку, коли

$$L_H(F_t, F_0^{(m)}) = \min_{1 \leq \gamma \leq M} L_H(F_t, F_0^{(\gamma)}) \wedge L_H(F_t, F_0^{(m)}) < \lambda^{(m)}, \quad (8.3)$$

де $L_H(F_t, F_0^{(\gamma)})$ – Гаусдорфові відстані між поточним фазовим портретом F_t особи, що тестується, та всіма еталонами $F_0^{(1)}, \dots, F_0^{(M)}$, визначеними згідно з (8.2), а $\lambda^{(m)}$ – поріг, що визначається за максимальним елементом матриці Гаусдорфових відстаней, побудованих на етапі формування $F_0^{(m)}$. Поріг $\lambda^{(m)}$ забезпечує необхідний компроміс між ймовірностями пропуску цілі та хибної тривоги під час ідентифікації m -го представника групи.

Для оцінювання надійності запропонованого підходу було побудовано експериментальну базу даних з 3133 ЕКГ 167 різних осіб. У 96,6 % (3027 ЕКГ) ідентифікація осіб згідно з правилом (8.3) була правильною.

Незважаючи на високі результати, запропонований підхід має обмежені можливості. Річ у тім, що фазові портрети одноканальної ЕКГ деяких осіб мають лише незначні відмінності (рис. 8.4). Не є винятком ситуації, коли ФП ЕКГ різних людей взагалі майже тотожні, причому ймовірність таких збігів збільшується зі зростанням їхньої кількості у групі.

З викладеного випливає, що побудову надійної системи ідентифікації конкретної людини за ФП ЕКГ у чималій групі виправдано розглядати лише як далеку мету.

Водночас, як ближню мету можна сформулювати простішу, але важливу задачу: зараховувати ФП ЕКГ конкретної людини до певного класу (кількість яких можна задати не дуже великою) і проводити верифікацію особи зіставленням її поточного ФП ЕКГ з відомим номером класу його ФП ЕКГ, який, як і група крові, належить цій особі. У разі розбіжності класів з високою ймовірністю можна стверджувати про негативний результат верифікації і приймати відповідні рішення.

8.4. Верифікація людини за фазовим портретом одноканальної ЕКГ

Для розроблення інтелектуальної системи верифікації визначимо характерні типи ФП ЕКГ за наявною навчальною вибіркою, що містить ЕКГ різних осіб, за якими побудовані відповідні фазові портрети

$$F = \{F_1, \dots, F_Q\}. \quad (8.4)$$

Визначимо можливість альтернативного підходу до оцінювання близькості фазових портретів $F_i \in F$ та $F_j \in F$, який ґрунтується на аналізі їхніх зображень. Для цього обчислюватимемо абсолютну величину різниці між функціями $\Psi_i(x, y)$ і $\Psi_j(x, y)$, які характеризують кількість чорних крапок у полі зображень фазових портретів $F_i \in F$ і $\Psi_j(x, y)$ і належать квадратним коміркам зі сторонами ε у координатах x і y фазової площини (рис. 8.6):

$$L_{ij} = \sum_{x,y} |\Psi_i(x, y) - \Psi_j(x, y)|. \quad (8.5)$$

Перед тим, як обчислювати відстані (8.5), проведемо нормування фазових координат за формулами

$$z^*(t_k) = \frac{z(t_k) - \min_{1 \leq k \leq K} z(t_k)}{\max_{1 \leq k \leq K} z(t_k) - \min_{1 \leq k \leq K} z(t_k)}, \quad k = 1, \dots, K,$$

$$\dot{z}^*(t_k) = \frac{\dot{z}(t_k) - \min_{1 \leq k \leq K} \dot{z}^*(t_k)}{\max_{1 \leq k \leq K} \dot{z}^*(t_k) - \min_{1 \leq k \leq K} \dot{z}^*(t_k)}, \quad k = 1, \dots, K,$$

що забезпечить виконання умов $z^*(t_k) \in [0,1]$ та $\dot{z}^*(t_k) \in [0,1]$.

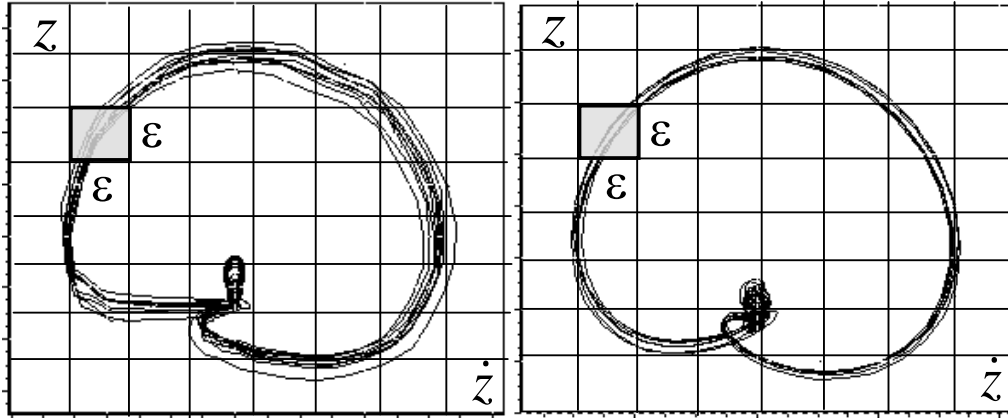


Рис. 8.6. Метод оцінювання відстаней між зображеннями ФП ЕКГ

Сформуємо квадратну матрицю відстаней (8.5)

$$\Lambda = \begin{pmatrix} L_{11} & L_{12} & \dots & L_{1Q} \\ L_{21} & L_{22} & \dots & L_{2Q} \\ \dots & \dots & \dots & \dots \\ L_{Q1} & L_{Q2} & \dots & L_{QQ} \end{pmatrix} \quad (8.6)$$

між всіма парами зображень фазових портретів навчальної вибірки, за допомогою якої проведено кластеризацію ФП ЕКГ.

Першу пару кластерів ФП ЕКГ визначає елемент матриці (8.6), що відповідає максимуму відстаней L_{ij} . Для визначення решти кластерів упорядкуємо за зростанням елементи рядка матриці (8.6), що відповідають першому кластеру.

Наявність «стрибків» на послідовності \mathfrak{R}_μ , $\mu = 1, \dots, Q$, упорядкованих відстаней (рис. 8.7) визначає межі між кластерами ФП ЕКГ. Такі стрибки легко можна відшукати за допомогою нескладної обчислювальної процедури.

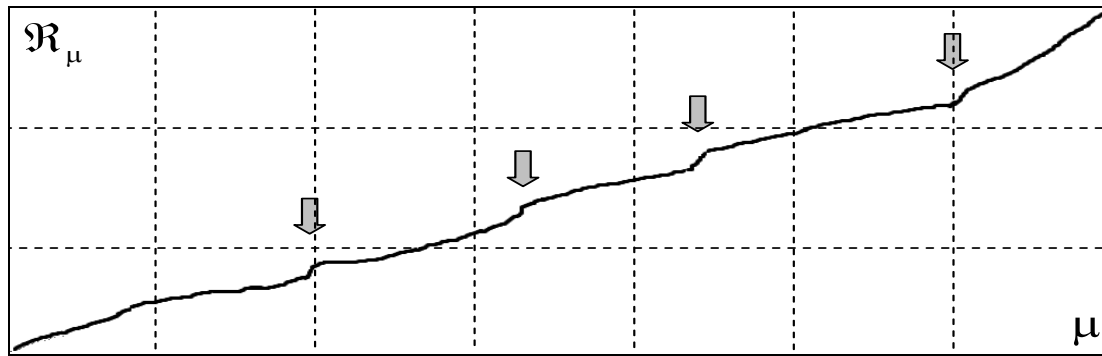


Рис. 8.7. Графік упорядкованих відстаней між кластерами ФП ЕКГ

Експериментально встановлено, що за критерієм максимуму відношення міжкласової та внутрішньокласової відстаней оптимальний розмір комірок має бути $\varepsilon = 0,125H$, де H – розмір поля зображень ФП ЕКГ. У такому разі поле зображень накривають 64 квадратні комірки.

За аналізом 300 ЕКГ 115 різних осіб визначено $J = 5$ характерних кластерів ФП ЕКГ (рис. 8.8).

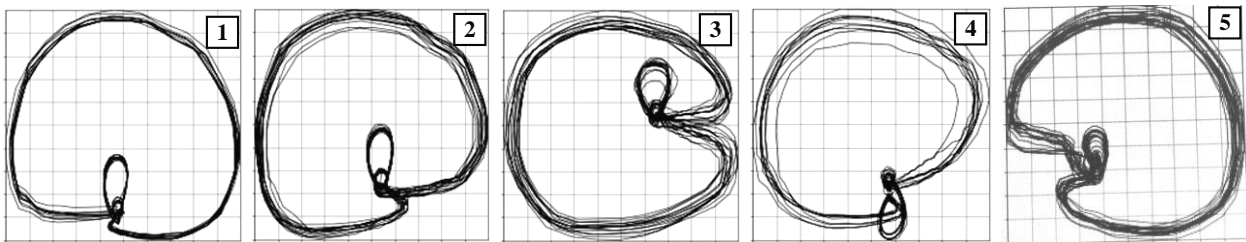


Рис. 8.8. Характерні кластери фазових портретів ЕКГ

Виокремимо з множини (8.4) підмножину фазових портретів, що відповідають ν -му кластеру ($\nu = 1, \dots, 5$), і визначимо еталон цього кластера у такий спосіб:

$$\mathfrak{Z}_0^{(\nu)} = \arg \min_{1 \leq i \leq Q_\nu} \sum_{j=1}^{Q_\nu} L_{ij}^{(\nu)}, \quad (8.7)$$

де $L_{ij}^{(\nu)}$ – відстані (8.5) між парами ФП ЕКГ ν -го кластера у навчальній вибірці, а Q_ν – кількість ФП ЕКГ ν -го кластера.

Верифікація особи проводиться за еталонами $\mathfrak{Z}_0^{(1)}, \dots, \mathfrak{Z}_0^{(5)}$, побудованими згідно з (8.7). Таку інформацію (логін особи, що тестується) можна отримати під час сканування мікропроцесорного чипа біо-

метричного паспорта, кредитної картки або іншого персонального документа.

Система верифікації (рис. 8.9) автоматично визначає номер еталону $\mathfrak{Z}_0^{(t)}$, який має мінімальну відстань (8.7) від зображення фазового портрета F_t поточної ЕКГ, тобто

$$\mathfrak{Z}_0^{(t)} = \arg \min_{1 \leq v \leq 5} L(F_t, \mathfrak{Z}_0^{(v)}), \quad (8.8)$$

де

$$L(F_t, \mathfrak{Z}_0^{(v)}) = \sum_{x,y} |\Psi^{(t)}(x,y) - \Psi^{(v)}(x,y)|, \quad v = 1, \dots, 5, \quad (8.9)$$

– відстань між зображенням $\Psi^{(t)}(x,y)$ фазового портрета поточної ЕКГ особи, що тестується, і зображеннями п'яти еталонів $\mathfrak{Z}_0^{(1)}, \dots, \mathfrak{Z}_0^{(5)}$, що зберігаються в базі даних системи доступу.

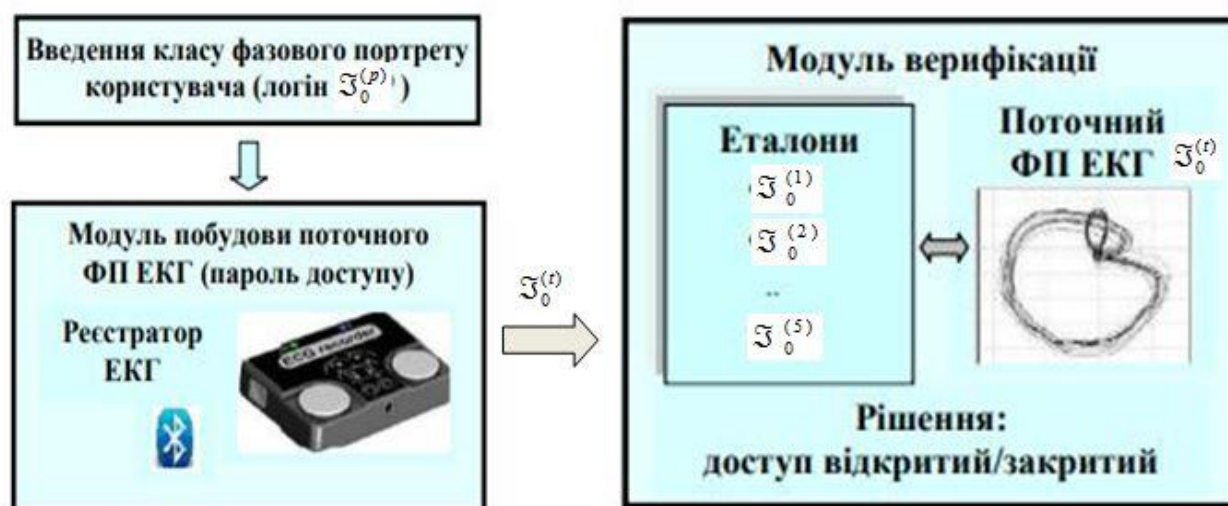


Рис. 8.9. Інтелектуальна система верифікації людини

Робимо висновок, що верифікація успішна лише тоді, коли визначений згідно з (8.9) еталон $\mathfrak{Z}_0^{(t)}$ збігається з уведеним логіном $\mathfrak{Z}_0^{(p)}$, тобто

$$\mathfrak{Z}_0^{(t)} = \mathfrak{Z}_0^{(p)}. \quad (8.10)$$

В іншому випадку верифікація вважається негативною.

Під час тестування запропонованого підходу на екзаменаційній вибірці, що містить 204 записи ЕКГ 62 різних осіб, виявилася лише одна хибна верифікація (0,5 %).

8.5. Перспективи подальшого розвитку

Крім розглянутих методів оцінювання близькості ФП ЕКГ, які ґрунтуються на формулах (8.1) і (8.5), можна побудувати інші процедури, що мають властивості метрик. Тож виникає природне запитання: яка з метрик ефективніша?

Розглянемо формальну схему обґрунтованого вибору міри близькості ФП ЕКГ за навчальною вибіркою, що містить G фазових портретів F_1, \dots, F_G . Вважатимемо, що кожен i -й фазовий портрет належить одному з різних $M \geq 2$ осіб з групи $B = \{B_1, \dots, B_M\}$, причому для кожної особи є декілька фазових портретів, тобто $G \gg M$.

Нехай маємо $N \geq 2$ різних мір близькості S_1, \dots, S_N . Кожна n -на міра S_n ($1 \leq n \leq N$) породжує власне рішення R_n , за яким проаналізований фазовий портрет $F_t \in \{F_1, \dots, F_G\}$ належить одному з представників групи $B = \{B_1, \dots, B_M\}$.

Розглянемо функцію

$$\Omega_t^{(n)} = \begin{cases} 0, & \text{якщо рішення правильне,} \\ 1, & \text{якщо рішення помилкове,} \end{cases} \quad (8.11)$$

яка визначає втрати від хибного рішення, що приймається за фазовим портретом F_t на основі міри $S_n, 1 \leq n \leq N$. Функція (8.11) дає змогу оцінити якість кожного n вирішувального правила у такий спосіб:

$$\Omega^{(n)} = \sum_{t=1}^G \Omega_t^{(n)},$$

та вважати найкращою міру $S_{n_0}, 1 \leq n_0 \leq N$, яка задовольняє умову

$$n_0 = \arg \min_{1 \leq n \leq N} \Omega^{(n)}. \quad (8.12)$$

Співвідношення (8.12) надає можливість обрати найкращу з мір близькості ФП ЕКГ.

Доцільно також дослідити можливість підвищення достовірності рішень, які приймаються, на основі переходу до ФП ЕКГ у тривимірному просторі з координатами $z(t), \dot{z}(t), \ddot{z}(t)$. Перші експерименти показали, що навіть якщо двовимірні ФП ЕКГ різних людей майже збігаються (рис. 8.10, а), то під час переходу у тривимірний простір

вдається виявити деякі відмінності нібито тотожних фазових портретів для певних кутів проектування (рис. 8.10, б).

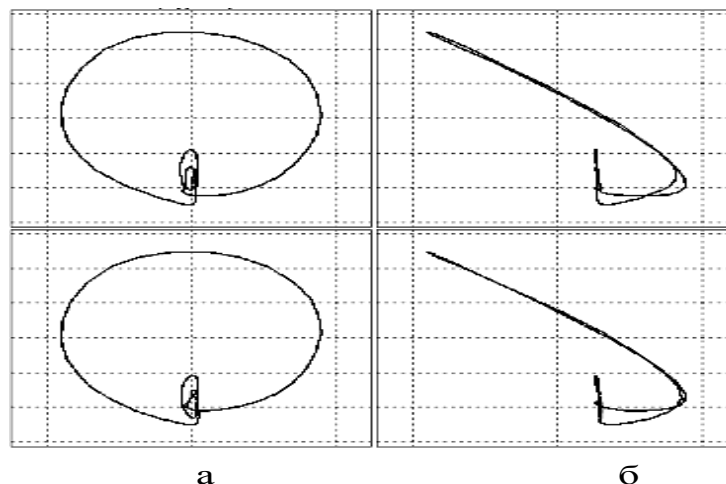


Рис. 8.10. Проекції на площину тривимірних фазових портретів двох осіб для різних кутів проектування

Результати досліджень цих задач на репрезентативній вибірці спостережень дозволять розширити можливості біометричних систем для розв'язання практичних задач у різних сферах застосування.

Інтелектуальні біометричні системи, що ґрунтуються на аналізі фазового портрета одноканальної ЕКГ, можуть бути застосовані як для самостійного використання, так і у складі існуючих систем. Наприклад, якщо в мікропроцесорний чип біометричного паспорта ввести інформацію про клас ФП ЕКГ власника, то під час перетину кордону можна здійснювати його верифікацію не тільки за відбитком пальця, як це роблять зараз, але і використовувати додаткову інформацію, що підвищить достовірність рішень, що приймаються.

Запитання для самоконтролю

1. У чому полягає відмінність біометричних систем?
2. Дайте загальну характеристику відомих систем біометричної ідентифікації.
3. Якими показниками характеризують безпечність доступу до об'єктів за допомогою біометричних систем?
4. Охарактеризуйте фазовий портрет ЕКГ як засіб біометрики.

ПІСЛЯМОВА

На основі поданого матеріалу можна сформулювати такий важливий висновок.

У межах алгоритмічного підходу побудова інтелектуальних ІТ та систем передбачає активну участь розробника технології, що створює ефективні обчислювальні алгоритми та комп'ютерні процедури, технологічний ланцюг яких забезпечує перехід від початкових даних до кінцевого продукту (див. рисунок).



РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Шаховська Н. Б. Системи штучного інтелекту : навч. посіб. / Н. Б. Шаховська, Р. М. Камінський, О. Б. Вовк – Львів : Львівська політехніка, 2018. – 392 с.
2. Нікольський Ю. В. Системи штучного інтелекту : навч. посіб. / Ю. В. Нікольський, В. В. Пасічник, Ю. М. Щербина – Львів : Магнолія, 2021. – 280 с .
3. Савченко А. С. Методи та системи штучного інтелекту : навч. посіб. для студ. напряму підготов. 6.050101 «Комп'ютерні науки» / А. С. Савченко, О. О. Синельников – Київ : НАУ, 2017. – 190 с
4. Троцько В. В. Методи штучного інтелекту : навч.-метод. і практ. посіб. / В. В. Троцько – Київ : Університет «КРОК», 2020. – 86 с.
5. Басюк Т. М. Машинне навчання : навч. посіб. / Т. М. Басюк, В. В. Литвин, Л. М. Захарія, Н. Е. Кунанець – Львів : Новий Світ-2000, 2019. – 315 с.
6. Zgurovsky M. Z. The Fundamentals of Computational Intelligence: System Approach / M. Z. Zgurovsky, Y. P. Zaichenko – Springer International Publishing, 2017. – 395 p.
7. Мороз О. Г. Перебірні алгоритми індуктивного моделювання на основі генетичних операторів / О. Г. Мороз, В. С. Степашко – Київ : Освіта України, 2021. – 216 с.
8. Файнзільберг Л. С. Теорія прийняття рішень : підруч. для студ. спец. «Комп'ютерні науки та інформаційні технології», спеціалізації «Інформаційні технології в біології та медицині» / Л. С. Файнзільберг, О. А. Жуковська, В. С. Якимчук – Київ : Освіта України, 2018. – 246 с.
9. Файнзільберг Л. С. Фазовий портрет електрокардіограми як засіб біометрії // Кібернетика та системний аналіз. – 2022. – Т. 58. – № 3. – С. 183–192.
10. Файнзільберг Л. С. 25-річний досвід створення та впровадження інтелектуальних ІТ оброблення біомедичних сигналів складної форми // Control systems and computers. – 2022. – N. 1. – P. 44–63.

ПРЕДМЕТНИЙ ПОКАЖЧИК

А

Адаптація 30, 113, 124
 Альтернація серця електрична 76
 Амплітудно-частотна характеристика (АЧХ) 99
 Атрактора розмірність 162

Б

Біометрична система 291

В

Вибірка навчальна 184, 224
 Властивості інтелекту 31
 Взаємозамінність 31
 Відстань

- Гаусдорфа 150
- Демарау-Левенштейна 179
- Евклідова 149
- Левенштейна 150
- Манхетенська 149
- Махаланобіса 149
- Хемінга 150

Г

Генетичний алгоритм 252
 Гіпотеза компактності 231
 Гнучкість 31

Д

Діагностичне правило 270, 274, 285, 288
 Діагностичний тест 299

Е

Ентропія

- ковзна 172
- оцінки 158
- перестановка 152
- Шеннона 157

 Ефект Гіббса 103

- ослаблення 106
- причина 105

 Ефективність бінарного класифікатора 202

З

Завада

- адитивна 43, 94
- випадкова 119
- внутрішня 43
- імпульсна 132
- зовнішня 43
- неадитивна 43
- частотна 98

 Згладжування

- адаптивне 125
- ковзне середнє 121
- експоненційне 120

- I**
- Метрика (відстань) 149
- Гаусдорфа 150
 - Демарау-Левенштейна 179
 - Левенштейна 177
- Інтелект
- властивості 31
 - штучний 17
- Й**
- Ймовірність
- апостеріорна 198
 - апріорна 198
- К**
- Класифікація
- безпомилкова 197
 - за відстанню 232
 - лінійна 225
 - циклів ЕКГ 148
- Кодограма 185
- Комунікабельність 31
- Критерій ефективності комп'ютерної процедури 54
- М**
- Медицина
- персоніфікована 264
 - цифрова 38
- Модель
- генеративна 65
 - інтерполяційна 58
 - ІТ оброблення сигналів 42
 - носія даних МНД 49
 - оброблення даних МОД 49
 - породження штучних сигналів 54
- Н**
- Навчальна вибірка 184, 224
- Навчання машинне 223, 228
- О**
- Оболонка опукла 227
- Оцінка
- похідної 143
 - середнього циклу 154, 156
 - частоти появи патерну 188
- П**
- Патерн кодограми 186
- Персептрон
- одношаровий 233
 - обмеження 239
 - навчання за алгоритмом Козинця 244
 - навчання за алгоритмом Розенблата 241
- Помилка
- пропуску цілі 201

– хибної тривоги 201
 Правило
 – вирішувальне 203
 – класифікації ЕКГ 80, 189,
 221
 Преваленс хвороби 207

Р

Розподіл
 – ознаки 194
 – умовний 201,
 Ризик
 – апостеріорний 207
 – апріорний 207
 – середній 200, 207
 – нейронної мережі 250, 251

С

Середнє квадратичне
 відхилення 164, 174
 Сигнал
 – складної форми 40,
 – тестовий 169
 – циклічний 136
 – штучний 72
 Структура
 – ІТ 45
 – серцево-судинної системи
 56
 Система інструментальна 51
 Синтез прикладних ІТ 45
 Специфічність 203
 Спотворення сигналу 59

Стратегія
 – Байєса 197
 – Вальда 202
 – Неймана-Пірсона 202

Т

Теорема
 – Байєса 198
 – Новікова 243
 – Вальда 202
 – Цибенко 250
 Тест «Китайська кімната» 28
 Технологія 30
 – інтелектуальна 31
 – інформаційна 31
 – оброблення сигналів 40
 Тюрінга
 – машина 18
 – тест 26
 Траєкторія
 – опорна 151, 152
 – фазова 149

У

Узагальнення 40

Ф

Функція
 – втрат 227
 – Гауса 71
 – індикаторна 177
 – Хевісайда 159

Фільтр	– ЕКГ 140, 293, 296
– адаптивний випадкових завад 125	
– адаптивний частотних завад 112	Х
– верхніх частот 99	Хаотичність сигналу 157
– дрейфу ізоелектричної лінії 130	Херста показник 164, 170
– каскадний 34	Ц
– нерекурсивний 100	
– нижніх частот 99	Цикл
– смуговий 99	– домінантний (опорний) 144
– ранговий (медіанний) 132	– нетиповий 152
– рекурсивний 100	– серцевий 57
– режекторний 99	– усереднений 137
Фур'є перетворення	
– дискретне (ДПФ) 101	
– швидке 118	
Фазова	Ч
– координата 139	
– площа 140	Чутливість 203
– траєкторія 149	
Фазовий портрет	
– сигналу 139	

ФАЙНЗІЛЬБЕРГ Леонід Соломонович



Доктор технічних наук, професор.

Головний науковий співробітник Міжнародного науково-навчального центру інформаційних технологій та систем НАН та МОН України.

Професор кафедри біомедичної кібернетики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Викладає навчальні дисципліни – «Методи та системи штучного інтелекту» та «Теорія вибору альтернатив». Керує роботами аспірантів та дипломними роботами бакалаврів та магістрів.

Фахівець у галузі комп'ютерних методів та засобів оброблення сигналів складної форми. Автор понад 300 наукових праць, зокрема 9 книг та 50 винаходів, захищених патентами України, США, Японії, Англії, ФРН та інших країн.

Розробив стохастичні моделі породження циклічних сигналів в умовах збурень та ефективний метод відновлення корисного сигналу за сукупністю спотворених реалізацій у фазовому просторі з використанням метрики Гаусдорфа.

Запропонував адаптивні процедури заглушення гармонічних та випадкових збурень. Сформулював математичні умови, що гарантують ефективність діагностичних ознак з погляду зменшення апріорного ризику.

Запропонував оригінальні моделі колективних рішень в умовах протиріч. Розробив нові методи інтерпретації циклічних сигналів на основі фазового портрета ковзної ентропії та відстані Левенштейна.

Запропонував оригінальний метод біометричної ідентифікації за фазовим портретом одноканальної ЕКГ. Розробив загальний підхід для створення інтелектуальних засобів превентивної медицини для домашнього використання.

Розробив наукові засади побудови комп'ютерних систем ФАЗАГРАФ, ІКАР, АІ-РИТМОГРАФ, MAGWIN, ВУГЛЕРОД, ГРАДУС, ЧУГУН, ТЕРМОГРАФ, які впроваджені в Україні, Угорщині та Німеччині.

Член редколегій журналів «Cybernetics and Computer Engineering», «Control systems and computers», «Клінічна інформатика та телемедицина», «Клінічна та профілактична медицина».

Нагороджений Грамотою Верховної Ради України та Відзнакою НАН України «За підготовку наукової зміни». Має звання «Кращий винахідник НАН України». Переможець Всеукраїнського конкурсу «Винахід року – 2015».

ORCID 0000-0002-3092-0794

Scopus Author ID: 6508231615 (44 статі, h-index 4)

Google Scholar (h index 19, h10-index 35)

Персональний сайт: <http://fainzilberg.irtc.org.ua/>

FAINZILBERG Leonid Solomonovich



Doctor of technical sciences, professor. Chief Researcher of the International Scientific and Educational Center for Information Technologies and Systems of the National Academy of Sciences and the Ministry of Education and Science of Ukraine.

Professor of the Department of Biomedical Cybernetics of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Academic disciplines – «Methods and systems of artificial intelligence» and «Theory of choice of alternatives». Supervises the work of graduate students and theses of bachelors and masters

Specialist in the field of computer methods and means of complex shape signals processing.. Author of more than 300 scientific papers, including 9 monographs and 50 inventions protected by patents of Ukraine, USA, Japan, England, Germany and other countries.

Developed stochastic models for the generation of cyclic signals under perturbations and an effective method for recovering a useful signal from a set of distorted realizations in the phase space using the Hausdorff metric.

He proposed adaptive procedures for suppressing harmonic and random disturbances. Formulated mathematical conditions that guarantee the effectiveness of diagnostic features in terms of reducing a priori risk. He proposed original models of collective decisions under contradictions.

Developed new methods for interpreting cyclic signals based on the phase portrait of moving entropy and the Levenshtein distance. He proposed an original method of biometric identification based on the phase portrait of a

single-channel ECG. Developed a general approach to the creation of intelligent preventive medicine tools for home use.

Developed the scientific basis for building computer systems FA-ZAGRAF, IKAR, AI-RITHMOGRAPH, MAGWIN, CARBON, GRADUS, IRON, THERMOGRAPH, which are implemented in Ukraine, Hungary and Germany.

Member of the editorial boards of the journals «Cybernetics and Computer Engineering», «Control systems and computers», «Clinical informatics and telemedicine», «Clinical and preventive medicine».

He was awarded the Diploma of the Verkhovna Rada of Ukraine and the Badge of Distinction of the National Academy of Sciences of Ukraine «For the preparation of a scientific shift». He has the title of «The Best Inventor of the National Academy of Sciences of Ukraine».

Winner of the All-Ukrainian competition «Invention of the Year – 2015».

ORCID 0000-0002-3092-0794

Scopus Author ID: 6508231615 (**44** articles, h-index **4**)

Google Scholar (h index **19**, h10- index **35**)

Personal website: <http://fainzilberg.irtc.org.ua/>

FAINZILBERG Leonid Solomonovich



Doktor der technischen Wissenschaften, Professor. Leitender Forscher des Internationalen Wissenschafts- und Bildungszentrums für Informationstechnologien und Systeme der Nationalen Akademie der Wissenschaften und des Ministeriums für Bildung und Wissenschaft der Ukraine.

Professor der Abteilung für biomedizinische Kybernetik der Nationalen Technischen Universität der Ukraine «Ihor Sikorsky Kyiv Polytechnic Institute».

Akademische Disziplinen – «Methoden und Systeme der künstlichen Intelligenz» und «Theorie der Wahl von Alternativen». Er betreut die Arbeiten von Diplomanden und Diplomarbeiten von Bachelor- und Masterarbeiten.

Er ist Spezialist auf dem Gebiet von Computermethoden und Werkzeugen zur Verarbeitung komplexer Signale. Autor von mehr als 300 wissenschaftlichen Arbeiten, darunter 9 Monographien und 50 Erfindungen, die durch Patente der Ukraine, der USA, Japans, Englands, Deutschlands und anderer Länder geschützt sind.

Er entwickelte stochastische Modelle der Erzeugung zyklischer Signale unter gestörten Bedingungen und eine effektive Methode zur Wiederherstellung des Nutzsignals aus einer Menge verzerrter Realisierungen im Phasenraum unter Verwendung der Hausdorff Metrik.

Er schlug adaptive Verfahren zur Unterdrückung harmonischer und zufälliger Störungen vor. Formulierte mathematische Bedingungen, die die Wirksamkeit diagnostischer Zeichen im Hinblick auf die Reduzierung des a priori-Risikos garantieren. Er schlug originelle Modelle kollektiver Entscheidungen unter Bedingungen von Widersprüchen vor.

Er entwickelte neue Methoden zur Interpretation zyklischer Signale basierend auf dem Phasenporträt der gleitenden Entropie und der Levenshtein-Distanz.

Er schlug eine originelle Methode zur biometrischen Identifizierung vor, die auf dem Phasenporträt eines Einkanal-EKGs basiert. Er entwickelte einen allgemeinen Ansatz zur Schaffung von intellektuellen Mitteln der Präventivmedizin für den Heimgebrauch.

Er entwickelte die wissenschaftliche Grundlage für den Bau der Computersysteme FASEGRAPH, IKAR, AI-RHYTHMOGRAPH, MAGWIN, CARBON, GRADUS, CHUGUN, THERMOGRAPH, die in der Ukraine, Ungarn und Deutschland eingesetzt werden.

Mitglied im Editorial Board der Zeitschriften «Cybernetics and Computer Engineering», «Control systems and computers», «Clinical informatics and telemedicine», «Clinical and Preventive Medicine».

Ausgezeichnet mit dem Zertifikat der Werchowna Rada der Ukraine und dem Ehrenzeichen der Nationalen Akademie der Wissenschaften der Ukraine «Für die Vorbereitung einer wissenschaftlichen Schicht».

Hat den Titel «Bester Erfinder der Nationalen Akademie der Wissenschaften der Ukraine». Gewinner des allukrainischen Wettbewerbs «Erfindung des Jahres – 2015».

ORCID 0000-0002-3092-0794

Scopus Autoren-ID: 6508231615 (**44** Artikel, h-Index **4**)

Google Scholar (h-Index **19**, h10-Index **35**)

Persönliche Website: <http://fainzilberg.irtc.org.ua/>

Навчальне видання

Л. С. Файнзільберг

**МЕТОДИ ТА СИСТЕМИ
ШТУЧНОГО ІНТЕЛЕКТУ**

Підп. до друку 17.07.2023. Формат 60×84/16.

Папір офсетний. Друк цифровий.

Ум. друк. арк. 18,48.

Наклад 300 прим.

Видавець і виготовлювач ТОВ «7БЦ»
03067, м. Київ, вул. Олекси Тихого, 84
e-mail: 7bc@ukr.net, тел: (044) 592-00-80

Свідоцтво суб'єкта видавничої справи
ДК №5329 від 11.04.2017

